



Vlaanderen
verbeelding werkt

Best Practices voor de Archivering van Sociale Media in Vlaanderen en Brussel

Fase 2 - D6

Eindrapport - Een duurzaam model voor de captatie en archivering van sociale media

Thomas Peeters, Nastasia Vanderperren, Ellen van Keer, Rony Vissers,
Katrien Weyns

31 augustus 2022

Inhoud

Inhoud	2
Inleiding.....	1
1. Contacten	2
Met contentpartners.....	2
Met gelijkaardige projecten	2
2. Vooronderzoek juridische aspecten.....	3
Gebruiksvoorwaarden van de platformen	3
Auteursrecht.....	4
Portretrecht / recht op afbeelding.....	4
Gegevensbescherming	4
Oplossingsrichtingen	5
Beleid en wettelijk kader	5
Organisatorische en technische maatregelen.....	5
Inhoudelijke afbakening	6
3. Vooronderzoek tools	1
4. Pilotprojecten.....	1
4.1. Verhouding tot de nodenbevraging	2
4.2. Analyse van datasets	3
4.2.1. Pagefreezer.....	3
Gebruiksvriendelijkheid.....	3
Captatie	4
4.2.2. Webpreserver.....	4
Gebruiksvriendelijkheid.....	4
Captatie	5
4.2.3. Archive-It	5
Gebruiksvriendelijkheid.....	5
Captatie	6
4.2.4. Twitter API.....	6
Gebruiksvriendelijkheid.....	6
Captatie	6
5. Naar een duurzaam samenwerkingsverband.....	8
6. Voorbereiding op de derde fase.....	9
Bevraging van onderzoekers	9
Uittekenen pilotprojecten en afbakenen scope	9

7. Algemene conclusie.....	10
Bijlage 1 – Rapport nodenbevraging content partners.....	12
Bijlage 2 – Sjabloon van een Evaluatieformulier.....	15
Bijlage 3 – Data Bevraging Onderzoekers.....	19
Bijlage 4 – Data Bevraging Gebruikerservaring.....	24

Inleiding

Dit is de zesde deliverable (D6) van het project *Best Practices voor de Archivering van Sociale Media in Vlaanderen en Brussel* gecoördineerd door KADOC-KU Leuven en meemoo. Dit rapport is het eindresultaat van projectfase 2 en omvat een overzicht van alle werkpakketten van het afgelopen werkjaar. In dit document is verder sprake van volgende termen:

<i>Socialmediaplatform</i>	Vaak afgekort tot “platform” gaat het in deze context om de sociale netwerksites zelf, zoals Facebook, of websites met door gebruikers gegenereerde content, zoals YouTube.
<i>Socialmedia-account</i>	Één specifieke gebruikerstoegang tot een socialmediaplatform. Een account heeft een hoofd-URL, bijvoorbeeld https://twitter.com/PACKEDvzw .
<i>Socialmediapost</i>	Publicatie van content op een account op een specifiek tijdstip. Een post neemt de vorm aan van een URL die resulteert in een HTML-pagina en kan embedded media, zoals een video of afbeelding, bevatten.
<i>Socialmediastream</i>	Één of meerdere posts en accountinformatie die binnen een bepaald datumbereik gearchiveerd worden van één socialmedia-account.
<i>Profielpagina</i>	Bevat alle gegevens over het socialmedia-account in kwestie. Het gaat dan bijvoorbeeld om de datum waarop het account werd aangemaakt, een profielfoto, aantal volgers en dergelijke meer.
<i>Comments</i>	Uitgeschreven reacties van andere gebruikers op socialmediaposts van het account.
<i>Likes</i>	Korte reacties die andere gebruikers maken op een bepaalde socialmediapost in een vooraf vastgelegde vorm. Vaak in de vorm van een duim, handgeklap, hartje of dergelijke meer.
<i>Repost</i>	Een socialmediapost die door een andere gebruiker opnieuw wordt geplaatst op het eigen account. Komt het vaakst voor op Twitter onder de naam “Retweet”.
<i>Story</i>	Reeks van posts die voor een beperkte duur, bijvoorbeeld 24u, gepubliceerd zijn en daarna automatisch weer verdwijnen.
<i>API</i>	Voluit “application programming interface”. Een verzameling definities waarmee een computerprogramma kan communiceren. Het API vormt zo de toegang tot een ander programma of website.

1. Contacten

Met contentpartners

Wat betreft de contacten met onze content partners is de situatie heel gelijkaardig aan die van de eerste fase. Het project werd opnieuw getrokken door KADOC en meemoo maar kreeg ook in de tweede fase advies, inzicht en medewerking van een *tiental content partners*. Die partners droegen bij via een nieuwe nodenbevraging en gingen tijdens de pilootprojecten actief aan de slag met tools. De analyse van die nodenbevraging vormde een waardevolle bron voor het evalueren van de commerciële tools. Tot slot leverden de klankbordmomenten steeds leerrijke inzichten in mogelijke samenwerkingsverbanden in de toekomst.

De klankbordgroep kwam doorheen de tweede fase zes keer samen. De aanleiding daarvoor scharnierde steeds op belangrijke mijlpalen in het project. Denk aan de opstart van fase twee, het opstarten en afronden van de pilootprojecten en een bijeenkomst met resultaten uit de analyses. De voorstelling van dit eindrapport was meteen ook het sluitstuk op deze fase. Tussentijds hielden we de content partners op de hoogte via nieuwsbrieven met updates wanneer die zich voordeden.

Met gelijkaardige projecten

Het project *Best Practices voor de Archivering van Sociale Media in Vlaanderen en Brussel* was niet het enige lopende initiatief dat onderzoek deed naar de captatie en bewaring van sociale media. Naast ons eigen project liep nog steeds het project [BESOCIAL](#) (KBR, ARA, UGent, UNamur en UCL)

Met BESOCIAL gingen we in deze tweede fase verder op het elan van de eerste. Doorheen het afgelopen jaar bleven we in wederzijds contact over vorderingen en inzichten uit elkaars projecten. BESOCIAL nam deel aan de klankbordgroepen en er was tussentijds overleg over het archiveringsluik. Verder bleven we via nieuwsbrieven en mailupdates op de hoogte van elkaars vooruitgang en resultaten.

Het Instituut voor Mediastudies van KU Leuven (IMS) dat ook het gebruik van sociale media verkende, nam opnieuw deel aan dit project. Vooral op vlak van captatie en ontsluiting naar onderzoek toe waren hun inzichten erg waardevol. Zo gaven ze bijkomend advies in welke vorm hashtag streams interessant kunnen zijn voor onderzoek. Inzichten uit hun projecten, zowel juridisch als praktisch werden vervolgens meegenomen in de opbouw van de pilootprojecten en het verder onderzoek naar een juridisch kader voor capteren.

Tot slot stelden we het project voor op een [ontmoetingsmoment](#) tussen Belgische en Nederlandse collega's. Aan Nederlandse zijde was het project [Sociale Media Archivering](#) van het Rijksprogramma voor Duurzaam Digitale Informatiehuishouding terwijl België vertegenwoordigd werd door zowel KADOC, meemoo als onze collega's van BESOCIAL. Waar we vanuit dit project vooral werken aan het opbouwen en aanvullen van onze erfgoedcollecties zijn de Nederlandse collega's eerder gericht op het verzamelen van bewijsmateriaal ter verantwoording van overheidsbeslissingen. Ondanks die verschillende invalshoeken liepen veel van de ervaringen gelijk en blijft het zoeken naar duurzame oplossingen. Een opvolgbijeenkomst wordt gepland in het najaar.

2. Vooronderzoek juridische aspecten

Een belangrijke vaststelling in Fase 1 van dit project was dat het archiveren van sociale media door privaatrechtelijke culturele archieven in Vlaanderen en Brussel vanuit juridisch oogpunt aanzienlijke obstakels stelt en dat het juridische luik een complexere uitdaging vormde dan aanvankelijk was gedacht.

De meest voor de hand liggende piste om sociale media op te nemen als (onderdeel van) een ‘klassieke’ archiefoverdracht via daartoe specifiek aangepaste schenkings- of bewaargevingsovereenkomsten met de accounteigenaars als archiefvormers bleek toch problemen te stellen. Bovendien kunnen op deze manier niet alle in beoogde types van captatie, inhoud en valorisatie van de datasets worden afgedekt. Vanaf Fase 2 willen we bijvoorbeeld ook content van meerdere accounts op basis van hashtags archiveren zonder dat voorafgaandelijk (archiverings)overeenkomsten met alle betrokken accounteigenaars kunnen worden afgesloten. Bovendien zijn er nog andere partijen in het spel waarmee mogelijks overeenkomsten of toestemmingen dienen te worden geregeld, vb. een verwerkerovereenkomst met IT-infrastructuur leveranciers, een contract met de platform eigenaars etc.

Vanuit de behoefte om een beter zicht te krijgen op de bestaande juridische context en noden werd een apart juridisch onderzoeksluik opgenomen in Fase 2 van het project. Als eerste stap werd de inventaris gemaakt van de juridische hindernissen en concrete oplossingen die we momenteel zien en die in Fase 3 nader onderzocht zullen worden, in samenspraak met externe juridische experts. In fase 2 waren er in dat kader bijkomende contacten met juristen van BESOCIAL en werden intern aan de KU Leuven relevante experts samengebracht om van gedachten te wisselen en concrete stappen vast te leggen.

Gebruiksvoorwaarden van de platformen

Een (onverwacht) groot obstakel is dat de platformen vaak zeer beperkende voorwaarden opleggen m.b.t. het gebruik en grootschalige en integrale archivering van sociale media content omdat zij daar zelf een businessmodel rond opgebouwd hebben. Bijvoorbeeld de [Twitter Terms of Service](#) (TOS) verbieden *data scraping* zonder voorafgaandelijke toestemming en laten *crawlen* enkel toe onder voorwaarden. Ook via de API wordt geen onbeperkte toegang gegeven. De *academic research track* heeft de minste beperkingen, maar de vereisten om hiervoor in aanmerking te komen zijn bijzonder hoog. Onze aanvraag werd geweigerd.

De meest platformen staan wel toe om de streams van eigen accounts te downloaden maar dit is vaak ook niet de integrale content. Bovendien verbieden bepaalde platformen het delen van data met derden. Voor hashtag-archivering (d.i. thematisch) is deze werkwijze evenmin geschikt. Voor grootschalige archivering van sociale media wordt vaak een beroep gedaan op diensten van commerciële bedrijven maar het is ons niet helemaal duidelijk hoe die hun werking juridisch (kunnen) verantwoorden. Op dit moment is er ook (nog) geen sluitende wettelijke uitzonderingsregeling waardoor de partners in dit project deze contractuele voorwaarden zouden kunnen ondervangen. Bovendien blijven er ook technische beperkingen bestaan.

Hoe we deze hindernis juridisch gezien kunnen opvangen zal in samenspraak met (externe en academische) juridische experts in Fase 3 nader bekeken worden. Een bijkomende moeilijkheid is dat de gebruiksvoorwaarden van de platformen (moedwillig) voortdurend veranderen en dit moet gemonitord worden.

Auteursrecht

Om auteursrechtelijk beschermd materiaal te reproduceren of publiek mee te delen dient voorafgaandelijke toestemming aan de auteur(s) of -rechthebbende(n) te worden gevraagd. Een moeilijkheid is dat, zeker bij online content, de rechtenstatus en rechthebbende(n) vaak niet geëxpliciteerd en dus onzeker zijn. Bovendien bevatten social media streams vaak gemengde content die ook uit diverse bronnen kan komen met meerdere rechthebbenden en uiteenlopende rechtenstatussen tot gevolg. Ook wanneer materiaal onder een vrije licentie ter beschikking wordt gesteld, kunnen er uiteenlopende gebruiksvoorwaarden aan gekoppeld zijn. Zeker bij *bulkarchivering* is het onbegonnen werk om dit allemaal gedetailleerd in kaart te brengen en toestemmingen van alle rechthebbenden voorafgaandelijk te regelen.

In het auteursrecht zijn wel een aantal wettelijke uitzonderingen bepaald die relevant zijn voor de partners en beoogde activiteiten in dit project. Voor de captatie van sociale media kunnen we ons bijvoorbeeld beroepen op de zogeheten preservatie-exceptie die o.m. publiekelijk toegankelijke archieven zonder winstgevend doel toelaat om een beperkt aantal kopieën van beschermde werken te maken voor de bewaring van het cultureel en wetenschappelijk erfgoed (zonder voorafgaandelijke toestemming dus). Op dit punt stellen zich m.a.w. geen al te grote problemen. De voorwaarden en modaliteiten voor ontsluiting liggen echter heel wat ingewikkelder en zullen nader bekeken worden in Fase 3.

Bovendien kijken we uit naar nieuwe mogelijkheden en bijkomende uitzonderingen voor tekst- en datamining en niet-in-de-handel-verkrijgbare werken die in nationale wetgeving zullen voorzien worden n.a.v. de omzetting van de EU [DSM-richtlijn](#). Er wordt verwacht dat dit in de loop van Fase 3 vorm zal krijgen. Dit zal dan ook van nabij opgevolgd worden.

Portretrecht / recht op afbeelding

Het portretrecht bestaat in het Belgisch rechtssysteem niet als een aparte rechtsfiguur. Bij afbeeldingen van herkenbare personen zijn twee verschillende wettelijke kaders van toepassing. Enerzijds bepaalt de auteurswetgeving dat een portret niet zonder toestemming gereproduceerd of verspreid mag worden tot 20 jaar na het overlijden van de geportretteerde. Toestemming kan evenwel impliciet zijn, zoals bij publieke figuren of gebeurtenissen.

Anderzijds is de gegevensbescherming van toepassing omdat een afbeelding een persoonsgegeven is. Deze bescherming eindigt bij het overlijden van de betrokkene.

Een vraag is in hoeverre de foto's en filmpjes van mensen op sociale media als gewone dan wel als 'bijzondere' en dus extra beschermde persoonsgegevens moeten worden behandeld omdat er soms informatie over o.a. politieke voorkeuren uit te halen is. Bij *bulkarchivering* is het immers onbegonnen werk om alle content met herkenbare personen en gevoelige informatie te filteren. Wanneer we nieuwe technologieën zoals gezichtsherkenning gaan toepassen om bijvoorbeeld de metadatering van beeldmateriaal te vergemakkelijken komen we onvermijdelijk in de extra beschermde zone van bijzondere gegevens terecht omdat we dan biometrische data verwerken. Er is wel een wettelijke uitzondering voorzien voor de verwerking van gevoelige gegevens voor verwerkingen met het oog op archivering in het algemeen belang. Daarbij zullen niettemin bijzondere veiligheidsmaatregelen in acht moeten worden genomen.

Gegevensbescherming

Persoonsgegevens zijn een zeer ruim begrip binnen de gegevensbescherming. Alle gegevens die terug te leiden zijn tot unieke personen die comments of posts schrijven of daarin voorkomen zoals een naam, woonplaats, e-mailadres, IP-adres of geboortedatum vallen hieronder. Om als organisatie

dergelijke gegevens te verwerken is een specifieke wettelijke basis nodig. De partners in het project kunnen zich hierbij beroepen op de uitoefening van een taak van algemeen belang en bovendien een aantal afwijkingen voor verwerkingen “met het oog op archivering in het algemeen belang” toepassen wanneer dit opgenomen is in hun wettelijke opdracht. Een moeilijkheid is dat privaatrechtelijke archiefinstellingen vaak maar een summiere officiële opdrachtomschrijving hebben die een beperkte rechtszekerheid biedt.

De meest essentiële afwijkingen voor de captatie en preservatie van sociale media voor archivering in het algemeen belang zijn rechtstreeks in de Europese wetgeving (AVG) zelf ingebouwd. Uitzonderingen op de basisprincipes van doelbinding en van opslagbeperking laten toe dat historisch waardevolle persoonsgegevens bij het vervallen van hun primair nut toch blijvend worden bewaard. Bovendien is de uitoefening van het recht op vergetelheid dat een betrokkene in staat stelt om wissing van al zijn gegevens te vragen, niet van toepassing. Daarenboven geldt de directe informatieplicht niet maar moet de informatie wel publiek wordt gemaakt. De nationale implementatiewet staat aanvullende afwijkingen toe maar stelt ook bijkomende voorwaarden. Zo moet het algemeen belang van de archieven voorafgaandelijk verantwoord worden. Bij overheidsarchieven ligt dit vast in gevalideerde selectielijsten. Bij social media accounts van individuele personen en private organisaties door privaatrechtelijke archieven kan enkel teruggegrepen worden naar (schenkings- of bewaargevings)overeenkomsten met de archiefvormers. Bij archivering o.b.v. hashtags zoals bv. #goktips #climate ligt het nog moeilijker. Daarbij wordt vooral naar het collectieplan van de archiverende instelling gekeken maar dit biedt maar een beperkte rechtszekerheid. Verder focust de nationale wetgeving vooral op de modaliteiten van ontsluiting, die in Fase 3 worden bekeken. Daarbij zal ook gekeken worden in hoeverre uitzonderingsregimes voor historisch, wetenschappelijk en statistisch onderzoek en voor journalistieke doeleinden extra bewegingsruimte kunnen creëren.

Oplossingsrichtingen

Beleid en wettelijk kader

Voor de gebruiksvoorwaarden van de socialemediaplatformen die een onverwacht groot struikelblok vormen zal in samenspraak met externe en academische juridische experts verder naar oplossingen worden gezocht. De eerste verkennende gesprekken hebben reeds plaatsgevonden.

Vanuit de vaststelling dat het bestaande wettelijke kader voor privaatrechtelijke archiefinstellingen op dit moment onvoldoende rechtszekerheid biedt om relevante wettelijke uitzonderingen toe te passen, en met name het uitzonderingsregime voor verwerkingen met het oog op archivering in de gegevensbescherming werden reeds contacten gelegd met het departement Cultuur, Jeugd en Media met de vraag om dit deze opdracht expliciet op te nemen in het kader van het (nieuwe) cultureelerfgoeddecreet.

De partners zelf zullen ‘archiveren in het algemeen belang’ en ‘sociale media’ in hun eigen missie en opdracht opnemen zodra de mogelijkheid zich aanbiedt. In Fase 3 zal een model voor een addendum of een clausule voor (bestaande) overeenkomsten met archiefvormers ontwikkeld en ter beschikking gesteld worden.

Organisatorische en technische maatregelen

Aangezien het praktisch onmogelijk is om de nodige toestemmingen van alle rechthebbenden en mogelijke betrokkenen te regelen zal ook in grote mate met risicobeperkende maatregelen worden gewerkt. De belangrijkste zijn: 1/ een copyright disclaimer, 2/ een opt-out- en take-down-procedure en 3/ een archief-privacystatement. Hiervoor zullen in Fase 3 modeldocumenten ontwikkeld worden, zoals een checklist voor algemenere maatregelen (vb. aanstellen DPO, uitrollen

gegevensbeschermingsbeleid, uitvoeren DPIA, ...) en concrete handvaten voor specifieke risicoanalyses.

Inhoudelijke afbakening

In dit project zullen we enkel publieke accounts en publieke gegevens (bereikbaar zonder login of zonder het account te moeten volgen) capteren, zoals we bij vele andere instellingen om begrijpelijke redenen zien. Bij publieke berichten kan men veronderstellen dat de nodig (impliciete) toestemmingen werd gegeven / geregeld en er zijn ruime uitzondering voor publieke informatie voorzien. Er zal een onderscheid gemaakt worden tussen captatie en ontsluiting. Het capteren van de integrale inhoud is een primordiale doelstelling voor dit project. In hoeverre de inhoud ook (aan verschillende doelpublieken) ter beschikking kan worden gesteld zal onderzocht worden in Fase 3.

3. Vooronderzoek tools

Het vooronderzoek naar de captatietools van commerciële dienstverleners werd al gevoerd in een vorige fase van dit project. In deze tweede fase werd uit die *long list* een gerichte lijst aan potentiële kandidaten gecontacteerd met oog op een concreet pilootproject.

In sommige gevallen weigerde de dienstverlener een pilootproject omwille van een grote post-corona drukte op de werkvloer, dat was het geval bij Archiefweb. In andere gevallen bleek de service van de dienstverlener toch niet overeen te stemmen met de wensen en noden van onze content partners. Hanzo, dat eerder gericht is op strategisch capteren van digitaal bewijsmateriaal, en Intradyn, dat achter de schermen capteerde met de tool van Pagefreezer, waren daarvan de voorbeelden. Tot slot antwoordden twee dienstverleners gewoon niet op vragen omtrent een pilootproject. Zo kregen we van Mirrorweb en Digital Footprints geen reactie meer.

Tool?	Weerhouden voor pilootprojecten?
<i>Archive-It</i>	Ja
<i>Digital Footprints</i>	Nee
<i>Hanzo</i>	Nee
<i>Intradyn</i>	Nee
<i>Mirrorweb</i>	Nee
<i>Pagefreezer</i>	Ja
<i>Webpreserver</i>	Ja
<i>Facebook for Developers (API-Key)</i>	Nee
<i>Twitter for Developers (API-Key)</i>	Ja

4. Pilootprojecten

Na het vooronderzoek en de selectie van drie commerciële tools gingen de pilootprojecten van start in april 2022. De pilootprojecten rond Pagefreezer en Webpreserver liepen gedurende één maand terwijl het pilootproject met Archive-It twee weken duurden.

Onderstaande tabel geeft weer welke pilootprojecten succesvol uitgevoerd werden, welke platformen centraal stonden en welke tools getest werd.

<i>Geteste Tool</i>	<i>Content Partner</i>	<i>Archiefvormer</i>	<i>Platform</i>	<i>Project Uitgevoerd</i>		
<i>Pagefreezer Webpreserver</i>	<i>ADV N</i>	<i>N-VA</i>	<i>Facebook</i>	<i>Ja</i>		
			<i>Twitter</i>	<i>Ja</i>		
		<i>Bart De Wever</i>	<i>Facebook</i>	<i>Ja</i>		
			<i>Instagram</i>	<i>Ja</i>		
			<i>Twitter</i>	<i>Ja</i>		
<i>Pagefreezer Webpreserver</i>	<i>Amsab-ISG</i>	<i>Kunstencentrum Vooruit</i>	<i>Facebook</i>	<i>Ja</i>		
			<i>Youtube</i>	<i>Nee</i>		
		<i>Diverse hashtags</i>	<i>Twitter</i>	<i>Ja</i>		
<i>Pagefreezer Webpreserver</i>	<i>AMVB</i>	<i>Kaaitheater</i>	<i>Facebook</i>	<i>Ja</i>		
			<i>Instagram</i>	<i>Ja</i>		
			<i>Twitter</i>	<i>Ja</i>		
<i>Pagefreezer Webpreserver</i>	<i>CAVA</i>	<i>Humanistisch Verbond</i>	<i>Facebook</i>	<i>Ja</i>		
			<i>Instagram</i>	<i>Ja</i>		
		<i>CAVA</i>	<i>Instagram</i>	<i>Ja</i>		
			<i>Instagram</i>	<i>Ja</i>		
<i>Pagefreezer Webpreserver Archive-It</i>	<i>KADOC</i>	<i>CD&V</i>	<i>Facebook</i>	<i>Ja</i>		
			<i>Instagram</i>	<i>Ja</i>		
		<i>Diverse hashtags rond (sport)gokken</i>	<i>Twitter</i>	<i>Ja</i>		
			<i>Youtube</i>	<i>Nee</i>		
			<i>Facebook</i>	<i>Ja</i>		
		<i>Pagefreezer Webpreserver Archive-It</i>	<i>Meemoo</i>	<i>PACKED vzw</i>	<i>Twitter</i>	<i>Ja</i>
					<i>Youtube</i>	<i>Ja</i>
				<i>Art In Flanders</i>	<i>Facebook</i>	<i>Ja</i>
					<i>Instagram</i>	<i>Ja</i>
					<i>Instagram</i>	<i>Ja</i>
<i>Archive-It</i>	<i>Letterenhuis</i>	<i>Tom Naegels</i>	<i>Facebook</i>	<i>Nee</i>		
			<i>Instagram</i>	<i>Nee</i>		
		<i>Gaea Schoeters</i>	<i>Facebook</i>	<i>Nee</i>		
			<i>Instagram</i>	<i>Nee</i>		
<i>Archive-It</i>	<i>Liberas</i>	<i>Bart Tommelein</i>	<i>Facebook</i>	<i>Ja</i>		
			<i>Instagram</i>	<i>Ja</i>		
		<i>Twitter</i>	<i>Ja</i>			
		<i>TikTok</i>	<i>Ja</i>			
		<i>Willemsfonds</i>	<i>Facebook</i>	<i>Ja</i>		
			<i>Instagram</i>	<i>Ja</i>		
			<i>Twitter</i>	<i>Ja</i>		

Archive-It	MHKA	Liberas	Instagram	Ja
			Twitter	Ja
		CKV	Facebook	Ja
Archive-It	VAi		Instagram	Nee
			Twitter	Nee
		Bovenbouw Architectuur	Instagram	Ja
		Vlaams Architectuurinstituut	Twitter	Ja
		Diverse hashtags	Instagram	Ja

Het merendeel van de geplande pilootprojecten kwam dus na opstart ook weer tot een succesvol einde. Van zowat elke contentpartner konden we rekenen op enthousiaste medewerking wat betreft analyse van de inhoud van datasets en de gebruiksvriendelijkheid van de captatietools.

Een handvol pilootprojecten werd niet afgerond omwille van enkele administratieve redenen. Het pilootproject met de Youtubekanalen van respectievelijk Kunstencentrum Vooruit en CD&V werd niet hernomen omdat er in een vorige fase een voldoende werkbare tool werd gevonden voor dit platform. Zowel Amsab als KADOC focusten daarom in deze fase op tools voor de andere platformen. Voor de socialmedia-accounts van PACKED vzw en Art in Flanders gingen alleen de captatierondes in Webpreserver en Archive-It door. Door onduidelijkheden in verband met de beheerder van de accounts in kwestie kon het capteren door Pagefreezer niet tijdig goedgekeurd worden. Tot slot waren er nog de accounts van CKV waarvan enkel de captatie van het Facebookaccount via Archive-It voltooid werd wegens tijdgebrek.

Tot slot vormde de tools zelf nog een struikelblok. De pilootprojecten met een Facebook- of Instagramaccount als centraal focuspunt botsten tegen technische problemen. Bij beide dienstverleners, en dus bij alle drie de tools, heersten er op het moment van de pilootprojecten problemen met het capteren van Meta-producten. Captaties van zowel Facebook als Instagram bevatten daardoor enkel inlogschermen en geen inhoudelijke data. De pilootprojecten werden dus wel opgezet maar konden geen werkbare datasets capteren.

4.1. Verhouding tot de nodenbevraging.

Voorafgaand aan de pilootprojecten van fase 2 bevroegen we de content partners over hoe zij een optimale commerciële tool of dienstverlener zagen. Die nieuwe bevraging bouwde verder op wat we al aftoetsten in de eerste fase. Dat wil dus zeggen dat de algemene inzichten over de functionaliteit van een tool die we al in fase 1 afbakenden voor deze tweede fase van kracht bleven. Kortom, een tool capteert bij voorkeur nog steeds zo volledig mogelijk, neemt nog steeds alle ingebedde media op, doet dat nog steeds zo geautomatiseerd mogelijk, levert nog steeds genoeg metadata op en is nog steeds zo breed inzetbaar mogelijk. Het enige verschil in deze fase is dat de nodenbevraging dieper inging op de gebruiksvriendelijkheid van de tools en service van dienstverleners, polste naar de noodzakelijkheden van captatie van hashtags en keek naar mogelijkheden tot verdere automatisering van het capteren en updaten van datasets.

Wat betreft de service van de dienstverleners verwachtten de content partners waar voor hun geld. Het kostenplaatje ten opzichte van gratis tools moest ook betekenen dat er gemoedsrust voor in de plaats komt. Een zekere garantie dat de tool blijft werken op de verschillende platformen, degelijke graad van automatisering en een responsieve helpdesk zijn daarbij de belangrijkste aandachtspunten. Een belangrijk aandachtspunt voor deze pilootprojecten was dan ook die mogelijkheden tot verdere

automatisering. Het feit dat er een commerciële dienstverlener schuilgaat achter de tool wekt bij een aantal content partners het vertrouwen om in een hoge graad van automatisering te kunnen werken en om op gezet tijdstippen automatisch updates te maken aan eerdere gecapteerde datasets. Die nood aan geautomatiseerd update toonde zich ook als het over het capteren van hashtags ging. Vooral met het oog op event based captaties vinden een aantal content partners het belangrijk om op gezette tijdstippen automatisch te kunnen capteren. Verder waren de wensen en noden voor captatie van hashtags erg gelijkaardig aan die van de captatie van posts in het algemeen. Een post wordt bij voorkeur zo volledig mogelijk gecapteerd, wat wil zeggen met inbegrip van embedded media, likes, comments en visuele context.

Een uitgebreider verslag van de nodenbevraging kan geraadpleegd worden in het *Rapport Nodenbevraging Content Partners* in bijlage 1. Daarin wordt naast deze algemene indruk voor elk aspect van captatie een aantal concretere richtlijnen op een rij gezet. De belangrijkste zaken om hier te onthouden zijn echter de nood aan geautomatiseerd capteren, zowel in bulk als voor updates, en een dienstverlener die zorgt voor een vlot draaiende captatietool.

4.2. Analyse van datasets

Net zoals dat het geval was tijdens de pilootprojecten van de eerste fase gingen ook deze keer de contentpartners aan de slag met verschillende tools. Door beperkingen op de omvang van de testgroep per dienstverlener werden de contentpartners ditmaal opgesplitst in twee groepen. ADVN, CAVA, Amsab-ISG en AMVB testten de tools Pagefreezer en Webpreserver terwijl VAI, Liberas, Letterenhuis, MuHKA en IMS de tool Archive-It aan tests onderworpen. KADOC en meemoo wierpen een blik op alle tools tegelijkertijd.

Elk van de contentpartners maakte daarbij gebruik van het schema voor analyse van datasets dat we eerder al publiceerden als de eerste deliverable bij deze fase van het project. Dit schema laat toe om een gerichte manier de ruwe inhoud van gecapteerde datasets te evalueren. Het evaluatieformulier is een in te vullen checklist in Excel. Door aan te vinken welke elementen wel en niet terug te vinden zijn in een gecapteerde dataset is het mogelijk een duidelijk overzicht op te bouwen van wat elke tool precies kan. Bovendien maakt het formulier het mogelijk om na individuele evaluatie ook een vergelijking te maken tussen de captaties van verschillende tools. Het eindresultaat is een overzicht waarin je kan focussen op de totale omvang en volledigheid van een gecapteerde dataset, of kan focussen op specifieke criteria zoals aanwezigheid van media of exacte inhoud van een gecapteerde post. Het modelformulier is als tweede bijlage bijgevoegd aan dit rapport.

Hieronder volgt een verslag per geteste tool met enerzijds een samenvatting van de gecapteerde datasets en hun inhoud en anderzijds een korte review van de gebruiksvriendelijkheid van elke tool op basis van de criteria uit de nodenbevraging. Meer gedetailleerde data uit de captatierondes kan geraadpleegd worden in de vierde bijlage.

4.2.1. Pagefreezer

Gebruiksvriendelijkheid

Het capteren verloopt binnen Pagefreezer via het dashboard op de website van de dienstverlener. Het proces loopt geautomatiseerd op basis van een lijst accounts. De werking van Pagefreezer is uitermate gebruiksvriendelijk. Het bedrijf neemt de technische kant volledig op zich. Als gebruiker stel je een lijst van socialemedia-accounts op die Pagefreezer vervolgens achter de schermen capteert. Je installeert zelf niets en doet geen eigen crawls of scrolls. De enige interactie met de tool is een vraag om toestemming tot captatie binnen het sociale mediaplatform in kwestie. Het resultaat, de gecapteerde

dataset, kan je raadplegen in het online dashboard. In datzelfde dashboard krijgt de dataset ook geautomatiseerd updates.

Er is echter één knelpunt. Enkel socialemedia-accounts die je zelf beheert kunnen gecapteerd worden. Hoewel de vertegenwoordiger eerst verklaarde dat het mogelijk zou zijn om accounts van derden te capteren mits schriftelijke toestemming van de beheerder en het verlenen van toestemming via het platform, bleek dit uiteindelijk toch niet te lukken. Dit maakt Pagefreezer meteen zo goed als onbruikbaar voor een archiefinstelling. Het hoofdoel is nu eenmaal het capteren van accounts van personen en organisaties (dus derden) als aanvulling op de papieren collecties.

Captatie

Pagefreezer capteert enkel accounts in eigen beheer. Daarvan capteert de tool de publieke posts. Wat na captatie in het dashboard ter raadpleging verschijnt, lijkt zeer volledig te zijn. De datasets die je daar oproept zijn volledig dynamisch, omvatten metadata over de captatie en bevatten de hele tijdslijn van het account. Voor Facebook bevat de gecapteerde dataset ook functionele links naar het profiel, foto's en links in de zijbalk. De embedded media is in het dashboard volledig functioneel.

Veelbelovend, maar bij de export loopt het mis. Pagefreezer laat toe te exporteren naar PDF, WARC, EML en MHTML. De export naar PDF bevat niet de volledige look-and-feel van het socialemediaplatform: gecapteerde posts worden op een statische manier gepresenteerd zoals ze eerder zichtbaar waren op het platform. De dataset bevat afbeeldingen, onderschrijft, comments en aantal likes. Bovendien heeft elke post metadata, zoals publicatie- en captatiedatum en een signatuur. Gezien de aard van het formaat is de embedded media echter enkel statisch bij export naar PDF, waardoor we video en audio verliezen. Een export naar WARC zou dat probleem oplossen. Maar dan moet die export wel lukken. Ondanks verschillende configuraties, van een beperkte date range naar de volledige tijdslijn, lukte het niet om een werkende WARC te exporteren. Na export kunnen we dus spreken over onvolledige datasets.

Pagefreezer heeft één groot pluspunt. Het bedrijf is een officiële partner van Meta en mag dus de API's van zowel Facebook als Instagram rechtstreek aanspreken voor captatie. Dat wil zeggen dat de tool kan capteren zonder de *terms and conditions* van de platformen te overtreden. Vanuit juridisch oogpunt is dat een belangrijk punt om in het achterhoofd te houden.

4.2.2. Webpreserver

Gebruiksvriendelijkheid

Komend uit dezelfde stal als Pagefreezer kunnen we ook Webpreserver onderbrengen in het segment gebruiksvriendelijke tools. Je voegt de tool toe als extensie op Google Chrome, logt in via de gegevens die je kreeg van de dienstverlener en je kan meteen beginnen capteren. Dit kan bovendien op twee intuïtieve manieren: je capteert de webpagina waarop je je bevindt via de optie *Quick Capture*, of je capteert een account in bulk via de *Bulk Settings*.

Voor beide manieren bestaat er bovendien een automatische piloot. De optie *Quick Capture* start met capteren na één klik op de knop, de optie *Bulk Capture* doet hetzelfde na configuratie van de te capteren set links. Met een minimum aan tijd start je dus een captatiesessie op. Die autopiloot durft echter haperen, zowel in de *Quick* als *Bulk Capture*. De platformen Facebook en Instagram lijken hiervoor het meest vatbaar te zijn. Pas nadat je manueel de pagina beweegt, start de tool opnieuw met automatisch capteren. Het is bovendien niet altijd duidelijk of de tool nog capteert of tot stilstand is gekomen. Er is geen statusbalk of waarschuwing. Het geautomatiseerde karakter van deze tool valt dus met een korrel zout te nemen: om een onverwachte stop te remediëren vereist de tool regelmatig menselijke controle en tussenkomst.

De tool maakt het ook mogelijk om zeer gericht te capteren dankzij het capteren met de *Date Range* of door specifieke links te capteren. Dankzij die optie *Date Range* in *Quick Capture* en de optie *Collect The Links* in *Bulk Capture* is het dus ook mogelijk om gericht elementen toe te voegen aan bestaande datasets. Dit proces verloopt echter niet geautomatiseerd: je moet zelf de data afbakenen of de te capteren links uitsluiten.

Tot slot werkt ook Web Preserver in overeenstemming met Twitter en Meta's *terms & conditions*. Het moederbedrijf Pagefreezer is zoals gezegd officiële partner en die toestemming vanwege de platformen geldt volgens de vertegenwoordiger dan ook voor deze tool.

Captatie

De datasets die Webpreserver capteert, bevatten de publieke posts van het account met inbegrip van eventuele comments. Tijdelijke stories, privé-accounts of interne berichten vallen buiten de mogelijkheden. Webpreserver capteert bovendien met behoud van look-and-feel. Gecapteerde posts worden dus gepresenteerd in de visuele context van het platform in kwestie. Op het vlak van metadatering vat de tool onder meer de titel, captatiedatum, enige trefwoorden indien toegevoegd, account, platform en URL-seed met tot slot nog een beschrijving en de checksum. De tool neemt ook embedded media op maar levert na export deze media als statisch en niet-functioneel bestand af in een PDF en WARC.

Webpreserver levert uiteindelijk onvolledige captaties. Datasets gecapteerd uit Facebook bevatten geen uitgekapt comments. Datasets uit Instagram tonen de comments enkel als je via de optie *collect the links* individuele posts capteerde. Op die manier krijg je voor elke post een afzonderlijk WARC-bestand met daarin de visuele context van het platform, afbeelding, ondertekening en likes en comments. De captatie van alle platformen is na export in een statische vorm gegoten. De embedded media zijn niet langer functioneel of ontbreken volledig. Gezien posts op sociale media, mede onder invloed van de populariteit van TikTok, hoe langer hoe meer bewegend beeld bevatten is dat toch een probleem.

Tot slot is de werking van Webpreserver voor Twitter anders dan voor Facebook en Instagram. Voor Twitter is er geen *Bulk Capture* waardoor je beperkt bent tot de *Quick Capture*. Het is dan weer wel mogelijk om in bulk de interne berichten van een Twitter-account in eigen beheer te capteren.

4.2.3. Archive-It

Gebruiksvriendelijkheid

Archive-It werkt net zoals Pagefreezer met een eigen dashboard van waaruit de captaties gestuurd worden. Na inloggen kan je meteen de eerste crawls instellen, zonder verdere installatie. In tegenstelling tot Pagefreezer verloopt het captatieproces dus niet volledig geautomatiseerd. Voordat Archive-It z'n werk doet leg je een collectie aan waarin je aangeeft welke seeds gecapteerd moeten worden met welke afbakening van data. Op basis van die informatie capteert Archive-It automatisch op eenmalig of herhalende basis naargelang de instellingen.

In het dashboard is het vervolgens mogelijk om enkele statistieken over de captatie te bekijken. Een gecapteerde dataset is enkel raadpleegbaar na export of via de *Waybackmachine*. Het exporteren verloopt tot slot rechtstreeks via de API. Er is geen downloadknop en je vraagt gecapteerde datasets op via een URL. Ofwel vraag je datasets op per collectie, ofwel vraag je datasets op per crawl. Automatisch en in bulk downloaden is mogelijk mits een bijkomend script, zonder bijkomend script moet handmatig elke WARC aangeklikt worden voor download.

De nood aan een correcte scope en het exporteren via de API zorgt er wel voor dat Archive-It een zekere leercurve heeft. Het vergt wat inlezen en uitproberen voor je de werking volledig in de vingers hebt en voor je het gewenste resultaat bereikt.

Captatie

De inhoud van de gecapteerde dataset wordt in het geval van Archive-It bepaald door het correct instellen van de *scoping* (afbakening en configuratie) van de crawl. De basismogelijkheden van de tool, zonder verdere specificatie van de *scope*, capteren de eigen berichten van het account in kwestie. Comments op berichten horen daar niet bij maar embedded media en eventuele weblinks worden wel mee opgenomen, net zoals de look-and-feel. Mits een bredere *scoping* is er nog meer mogelijk. Bijvoorbeeld het ophalen van data uit andere accounts en websites die gelinkt worden in een post. Wat betreft metadata kan je in de export zelf de nodige metadatering toevoegen waar nodig.

De invloed van de platformen op de werking van de tool vormde een probleem. Op het moment van de *trial* ondervond Archive-It [problemen](#) met de captatie van platformen onder de paraplu van Meta, namelijk Instagram en Facebook. Captaties van beide platformen leverden de facto lege WARC-bestanden op, de tool liep steeds vast op de inlogpagina. Aanpassingen aan de *scoping* van de captatie of zelfs het toevoegen van login-gegevens verhielpen niets. Voor andere socialemediaplatformen bleken er geen problemen te zijn, maar ook Archive-It ondervindt dus moeilijkheden bij het capteren.

4.2.4. Twitter API

Gebruiksvriendelijkheid

Bij het capteren via de Twitter API moet een zekere drempel worden overwonnen. In de eerste plaats omdat het aanvragen van toegang tot de API niet vanzelfsprekend is. Enkel onderzoekers met een zeer concrete onderzoeksvraag krijgen na een uitgebreide aanvraagprocedure toegang tot de *Academic Track*. Die *Academic Track* is qua mogelijkheden het hoogst haalbare en laat toe om tot tien miljoen tweets per maand te capteren zonder beperking in de tijdspanne. Het niveau daar net onder, de *Elevated Track*, is beperkt tot het capteren van twee miljoen tweets en een periode van maximum zeven dagen terug in de tijd. Afhankelijk van de publicatiehoeveelheid van accounteigenaars en de hoeveelheid van geïssueerde accounts biedt deze laatste optie de mogelijkheid voor archiefinstellingen om beperkte datasets op te bouwen.

Na een goedgekeurde toegang moet er zelf een query gecodeerd worden om de data uit de API op te vragen. Enige technische kennis is dus een minimumvereiste om succesvol te capteren.

Captatie

De volledigheid van de gecapteerde datasets staat of valt bij het soort API key waarvoor je de toelating wel of niet krijgt. Als je niet beschikt over de *Academic Track* zal je nooit helemaal tot bij het begin van een account kunnen teruggaan voor captatie. Voor instellingen die enkel een periodieke aanvulling van een beperkt aantal bestaande datasets willen uitvoeren is dit een werkbare optie maar voor wie helemaal tot in het begin van een account wil capteren wordt het moeilijk.

De Twitter API aanspreken, leverde tijdens de pilootprojecten datasets op in het JSON-formaat. Dat betekent meteen ook dat enige visuele context of embedded media wegvallen. Enkel een URL naar de embedded media wordt gecapteerd, niet het beeld zelf. Inhoudelijk is de captatie best volledig. Je krijgt het eigenlijke bericht samen met een heleboel metadata zoals de Twitter ID's, auteur, bron van de post, datering en meer.

Kortom, er is heel wat mogelijk wat betreft capteren via het Twitter *developer platform* maar alles opgeteld zijn er gebruiksvriendelijkere alternatieven die zeer vergelijkbare resultaten opleveren.



Twarc is daar een goed voorbeeld van. Door het rechtstreeks aanspreken van de Twitter API voldoet je gecapteerde dataset evenwel aan Twitters *terms and conditions*.

5. Naar een duurzaam samenwerkingsverband

Een laatste onderdeel van de nodenbevraging bij deze fase wierp een blik op de verdere toekomst. Ook na het aflopen van dit project zal het landschap van platformen, tools, wetgeving en data verder blijven evolueren. Met dat in het achterhoofd bevroegen we de content partners over mogelijkheden om in de toekomst de vinger aan de pols te houden.

Uit de bevraging bleek dat alle content partners blijvend willen inzetten op een captatiebeleid. De partners zien de voordelen van een samenwerkingsverband tussen verschillende instellingen. Meer dan vijftig procent vond een mogelijke samenwerking een echte meerwaarde terwijl de overige veertig procent het op z'n minst al *nice to have* vond. Zo'n samenwerkingsverband kan echter verschillende kanten op, van een gedeelde infrastructuur of licentie om kosten te drukken, over de CEST-pagina's met resources en handleidingen up-to-datetot een *zelfhulpgroep* waarin vragen en issues behandeld kunnen worden.

Tijdens de laatste klankbordgroep kwam het vooruitzicht van een samenwerkingsverband opnieuw op tafel. De algemene teneur was dat de betalende tools nog te veel bleken te lijden onder bugs en storingen om hun hoge kostenplaatje te verantwoorden. Gedeelde infrastructuur die werkt op basis van de API's bleek eerder al geen alomvattende en te verkiezen oplossing voor de deelnemende erfgoedorganisaties. Om echter zelf, ieder individueel, aan de slag te gaan met gratis tools is het dan weer te moeilijk om geschikte technische profielen aan te trekken. Een praktijknetwerk die nieuwe ontwikkelingen opvolgt en ervaringen uitwisselt op gezette tijdstippen kwam daarom als meest wenselijke optie uit de bus.

Zo'n praktijknetwerk zou steunen op twee contactpijlers. Enerzijds een halfjaarlijkse bijeenkomst om brede tendensen en vooruitzichten te bespreken. Anderzijds een online kennisuitwisselingsplatform om *in real time* en duurzaam problemen en oplossingen te verwerken. CEST vormt in dat kader een belangrijk instrument. Bedoeling is om zowel op lange termijn gestructureerd ontwikkelingen op te volgen, als op de korte termijn ad hoc problemen aan te pakken. Er wordt gezocht naar samenwerking met bestaande initiatieven en platformen, zoals het Nederlandse Kennisnetwerk Informatie en Archief, VVBAD en DAV naast de nu aanwezige partners. Ook bij Waalse collega-instellingen wordt gepolst of er interesse is om aan te sluiten. Intussen zet KBR in op een nieuwe project Belgicaweb waarin ze de ontsluiting van webcollecties verder zullen ontwikkelen.

6. Voorbereiding op de derde fase

Een deel van deze fase bestond opnieuw uit voorbereiding voor een volgende fase. In dit geval keken we al verder naar een derde fase die zal focussen op het onderzoek naar mogelijkheden voor het ontsluiten van de gecapteerde content. Uitdieping van het juridische kader en een bevraging van het potentiële doelpubliek stond daarbij centraal.

Bevraging van onderzoekers

Naast onze contentpartners bevroegen we ook een aantal onderzoekers ter voorbereiding van een derde fase. In die derde fase zullen we namelijk dieper ingaan op het ontsluiten van gecapteerde datasets. Voor archieven vormen onderzoekers een belangrijke doelgroep en daarom bevroegen we hen alvast om te polsen in hoeverre data uit sociale media een interessante bron vormen. We mikten op academische onderzoekers uit verschillende disciplines, van *Digital Humanities* tot Politieke Wetenschappen of *Cybersecurity* tot Media Studies.

Opvallend is dat alle respondenten in hun onderzoek al gebruik maakten van gecapteerde data. Dat benadrukt voor ons project eens te meer het belang van captatie en ontsluiting van datasets op een structurele basis, voor dit doelpubliek is sociale media als bron duidelijk al lang niks nieuws meer. Wat betreft platformen ligt de nadruk ook hier vooral op de populaire platformen. Facebook, Instagram, Twitter en Youtube zijn voor elke respondent van belang terwijl LinkedIn, Vimeo of TikTok voorlopig eerder *nice to have* zijn. Uit de bevraging blijkt wel dat berichtendiensten of fora als Reddit, Telegram en Discord ook in toenemende mate van belang zijn.

Wat verder meteen opvalt in de resultaten is dat we dezelfde tendens zien bij onderzoekers als bij onze contentpartners. Het motto is ook hier “Hoe meer data, hoe beter”. Uit de reacties blijkt dat een gecapteerd bericht zo volledig moet zijn. Zowel de boodschap als de ingebedde media als reacties en comments zijn essentieel en dat zowel in posts op de tijdlijn als in de stories. In termen van afbakening wil men ook zo weinig mogelijk beperkingen. Een gecapteerde dataset omvat liefst een volledig account of volledige *hashtag feed*. Tot slot is ook de bijgevoegde metadata best zo exhaustief mogelijk. Alleen zaken als geolocatie, harvesttool en archiefvormer blijken van ondergeschikt belang aan metadata over de inhoud van de post of het type van de ingebedde media.

Aan dat “hoe meer, hoe beter” motto zit echter één belangrijke kanttekening. Publiek beschikbare data mag dan wel *as is* beschikbaar zijn na captatie, als het aankomt op afgeschermd informatie zoals privéberichten, privé-accounts of besloten groepen wensen de bevraagde onderzoekers die data geanonimiseerd te gebruiken of zelfs helemaal niet beschikbaar te maken uit privacy-overwegingen.

We bevroegen de onderzoekers tot slot over de mogelijkheden van hergebruik van datasets en hoe zij zo’n hergebruik zagen. Daarbij is het duidelijk dat een zo breed mogelijke inzetbaarheid belangrijk is. Datasets zijn best meteen na captatie beschikbaar voor download. Verder kijkend zou het mogelijk moeten zijn om in de eerste plaats gecapteerde data te gebruiken als bron maar daarnaast ook beschikbaar zijn voor presentatie in publicaties, lezingen of online expo’s.

Volledige resultaten van de bevraging zijn beschikbaar in bijlage 3 achteraan dit rapport.

Uittekenen pilootprojecten en afbakenen scope

In aanloop naar de derde fase werd ook nagedacht over mogelijke pilootprojecten voor de ontsluiting van gecapteerde datasets, waarbij het aantal ideeën gereduceerd werd door de wettelijke mogelijkheden waarbinnen gewerkt moet worden. Wat hieronder volgt zijn enkele mogelijke pistes die we verder uitdiepen in de pilootprojecten van de komende fase.

Een eerste, voor de hand liggende, piste is het ontsluiten van datasets voor wetenschappelijk onderzoek. Het kan dan gaan om exhaustieve captaties van volledige accounts, om langdurige captaties van bepaalde hashtagstreams of om eerder event based captatie van een hashtag over een beperkte periode. Een voorbeeld van zo'n captatie is het pilootproject rond hashtagcaptatie dat deze fase liep, meer bepaald capteerden we enkele datasets op basis van hashtags omtrent sportgokken en online weddenschappen voor IMS.

Naast zo'n datasets voor onderzoek vormt een aanbod voor het onderwijs een tweede mogelijke piste. Datasets voor onderwijs zouden eerder opgebouwd worden rond een bepaald maatschappelijk thema relevant aan de eindtermen. Denk aan een dataset uit de accounts van politici in verkiezingstijd, een dataset op basis van hashtags rond grote gebeurtenissen als de klimaatmarsen of de oorlog in Oekraïne. De bedoeling is om een lessenpakket aan te bieden waarin leerlingen zelf aan de slag kunnen met gecapteerde datasets.

Enkele content partners hadden tot slot ook het voorstel om te capteren vanuit een tentoonstellingsoogpunt. De centrale vraag is daar "Hoe geven kunstenaars, architecten en ontwerpers hun werk een plaats op sociale media". Een gecapteerd sociale media-account zou op die manier kunnen fungeren als een soort digitale retrospectieve, een online aanvulling op het fysieke werk.

7. Algemene conclusie

De tweede fase van het project draaide globaal gezien rond vier grote luiken. We besteedden tijd aan juridisch onderzoek, aan voorbereidingen op de derde fase en aan het uittekenen van een samenwerkingsverband in de nabije toekomst. Het grootste deel van de tijd werd logischerwijs gereserveerd voor een tweede ronde aan pilootprojecten, dit keer met commerciële tools.

In deze tweede fase sloten we een deel van het onderzoek van de eerste fase over duurzame captatiemogelijkheden af. De pilootprojecten toonden aan dat ook de commerciële tools geen totaaloplossing opleverden. Een belangrijke eerste vaststelling is dat veel van de commerciële spelers een andere kijk op captatie hebben. De bedrijven zijn er vaak op ingesteld om strategisch en gericht bewijsmateriaal te capteren. Denk aan specifieke posts of specifieke e-mails. Exhaustieve captatie van hele accounts is vaak moeilijk. In tweede instantie bleek de kost van zo'n betalende tools vooropgestelde budgetten van de content partners ver te overschrijden. Bovendien blijven de tools tegen dezelfde problemen aanlopen als hun gratis tegenhangers. Met andere woorden, de meerwaarde van een betalende tool bleef voorlopig nog uit. Dat wil meteen ook zeggen dat we nog steeds geen one size fits all tool hebben gevonden. Op basis van de pilootprojecten uit zowel de eerste als de tweede fase konden we een aantal werkbare [adviezen](#) formuleren. Een vergelijking van de verschillende methodes kan geraadpleegd worden in de vijfde deliverable, beschikbaar op [CEST](#).

Aansluitend op het inzicht dat een *be all end all* oplossing zich voorlopig nog niet aandient, dringt de nood aan een samenwerkingsverband zich op. Zo'n samenwerkingsverband zou het werk van de vinger aan de pols houden, verdelen over meerdere partners en kan de snel evoluerende wereld van tools en platformen overzichtelijk houden. Zowel op technisch vlak als voor het juridische aspect is dit een meerwaarde. De content partners van dit project zijn alvast vragende partij voor verdere samenwerking naar de toekomst toe. Over de taalgrens heen vormt het AAFB een mogelijke partner maar daarnaast liggen er ook mogelijkheden buiten de landsgrenzen. Organisaties als KIA en IIPC springen zo al direct in het oog.

Als derde luik vormde verder onderzoek naar het juridisch kader voor captatie een belangrijk facet van deze fase. Samenkomsten tussen de juridische experts van meemoo, de juridische dienst van KU Leuven, de Leuvense rechtenfaculteit en de BESIOICIAL-experts toonden aan dat het een complex domein betreft waar geen eenduidige antwoorden te vinden zijn. Er komen verschillende rechtsdomeinen samen bij de captatie en archivering van sociale media en dat betekent dan ook dat het moeilijk is om een oplossing voor elk obstakel uit te tekenen. Ook hier is nauwere samenwerking dus een nuttige strategie naar de toekomst toe. Vele handen maken licht werk maar in groep is het ook realistischer om te wegen op toekomstige wetgeving.

Tot slot is er de voorbereiding op de derde fase. We maakten namelijk al een kleine voorzet op wat we in 2023 verder zullen uitdiepen. Centraal stonden daar een eerste bevraging van wetenschappelijk onderzoekers, de eerste kantlijnen voor pilootprojecten en verder onderzoek naar het juridisch kader. Op die basis beginnen we in september aan de derde fase en laatste fase van dit project.

Bijlage 1 – Rapport nodenbevraging content partners

Op dinsdag 16 november 2021 vulden acht van onze content partners de survey in. Op basis van die acht inzendingen volgen onderstaande inzichten over de noden en verwachtingen voor het capteren van hashtags in de tweede fase van het project.

Pilootprojecten fase 2

Belangrijk om weten is dat alle acht content partners graag hun pilootproject hernamen. Ze wensten daarbij ook zelf aan de slag te gaan met de commerciële tools. Omdat dit minder vanzelfsprekend is dan bij de gratis beschikbare tools moesten content partners verdeeld worden over de testrondes van de verschillende beschikbare tools. Dit omwille van beperkingen op het aantal gebruikers of omwille van een te grote supportlast bij de aanbieder.

De accounts die in fase 1 gecapteerd werden als deel van een pilootprojecten werden (bijna allemaal) ook in fase twee opnieuw gecapteerd om vergelijking tussen de datasets uit beide pilootprojecten toe te laten.

Hashtags capteren

Een content partner gaf nog een mogelijke functionaliteit mee als aanvulling op al het bovenstaande: moest een tool kunnen aangeven hoe vaak een hashtag precies gebruikt of gedeeld werd, dan is dat *nice to have*. Bijkomende statistieken over de hashtag zijn dus een mooi meegenomen functionaliteit. Tot slot kwam ook de nood aan een updatemogelijkheid bovendrijven. Uit fase 1 bleek dat tools een account konden capteren maar dat een actief account updaten daarentegen moeilijk bleef. Als archivaris moest je of wel alles opnieuw capteren ofwel manueel een tijdsbereik capteren. Een tool met de functie om op gezette tijdstippen automatisch captaties aan te vullen bleek dus nog steeds te ontbreken.

Wat betreft voorstellen van een te capteren hashtag leken de meeste content partners zich aan te sluiten bij het voorstel om iets rond de actuele klimaatmarsen te doen. Zo kwam uit de voorstellen #wakeupforclimate, “de hashtag gelanceerd rond de betoging van 10 oktober, trefwoorden zoals 'klimaatverandering' en 'klimaatmars', #klimaatmarsen, #actnow naar voren. Alternatieven die uit de bus kwamen waren onder andere een hashtag rond het thema “migratie”, “#kooplokaal”, “#winkelhier”. Moest het mogelijk zijn retroactief te capteren dan zag één content partner nog een mooi project in de hashtag #samentegencorona.

Commerciële tools

Gevraagd naar de captatiemogelijkheden van een tool lagen de verwachtingen van de content partners in dezelfde lijn als die van fase 1, namelijk “hoe meer er gecapteerd kan worden, hoe beter dat is”. Op alle mogelijkheden werd affirmatief gereageerd, minstens “mooi meegenomen” dus.



Uit de bevraging bleek wel dat de content partners als belangrijkste functionaliteiten de mogelijkheid om berichten van meerdere accounts en berichten met al hun metadata te capteren aanduiden. De eventuele functionaliteit om geografisch af te bakenen kreeg net zoals de rest wel een *nice to have* mee maar werd door het slechts vijftientig procent als noodzakelijk gezien. Ondanks dat alles van de functionaliteiten minstens mooi meegenomen is, tekenden er zich toch nuances af.

Wat betreft de scope aan platformen die een tool moet kunnen capteren lag de nadruk, net zoals in fase 1, opnieuw op de "grote vier". Facebook, Twitter, Instagram en YouTube waren voor alle content partners op z'n minst *nice to have* en voor het merendeel een basisvereiste. Zeker Instagram was voor iedereen een platform dat moest gecapteerd kunnen worden. In de achtergrond van die vier waren de meningen verdeelder. Wat betreft LinkedIn, Whatsapp, TikTok of Vimeo neigde de meerderheid naar onnodig of enkel mooi meegenomen. De focus lag dus best op de eerstgenoemde platformen en in tweede instantie op de "mindere goden".

Een enkele content partner legde de scope van een tool echter nog veel breder en zou ook Discord, Reddit, Flickr, Soundcloud, Tumblr en Telegram willen capteren.

Ondersteuning

Gezien we in de tweede fase met betalende, commerciële tools werkten is het ook belangrijk dat we zicht kregen op hoe de content partners de samenwerking met de dienstverlener zien. Daaruit blijkt dat de meerderheid geen noodzaak zag in een dienstverlener die de captatie uitvoert. Vijftig procent vond het hoogstens mooi meegenomen, net geen veertig procent vond het zelfs onwenselijk dat de captatie bij het bedrijf lag in plaats van in eigen handen.

Belangrijker vonden de content partners de ondersteuning die werd aangeboden. Een negentig procent wilde op z'n minst een handleiding, vijftig procent zag de meerwaarde van een helpdesk in. De rest van de partners vonden beide aspecten op z'n minst *nice to have*.

Als het aankomt op het kostenplaatje was er minder duidelijkheid. De content partners gaven aan dat veel zou afhangen van de mogelijkheden van de tools zelf en de manier waarop we de handen in de toekomst in elkaar kunnen slaan. Veel content partners moesten bovendien intern ook nog aftoetsen welke middelen er eventueel zijn, een van hen wenste voor de commerciële tools ook in de toekomst geen middelen vrij te maken. Een bovengrens tekende zich echter af rond het bedrag van circa 2000-euro op jaarbasis voor een commerciële tool. Tools die daar ver boven liggen komen dus in se niet in aanmerking voor langdurig gebruik.

In ruil voor dat kostenplaatje wensten de content partners wel een zeker niveau van betrouwbaarheid en bestaanszekerheid van de tool. Een officiële samenwerking of erkenning met de platformen, garantie dat accounts niet geblokkeerd raken en volledigheid van datasets zijn zeker voor een betalende tool nog eens zo belangrijk.

De Toekomst

Tot slot wilden alle bevroagde content partners blijvend inzetten op een captatiebeleid. Ze zagen daarbij ook de voordelen in een samenwerkingsverband tussen verschillende instellingen. Meer dan vijftig procent vond het een echte meerwaarde, de overige veertig procent vond een samenwerking *nice to have*. Die samenwerking mag dan afhankelijk van de uitkomst van het project uitlopen in gedeelde kosten, expertisedeling of een gezamenlijke infrastructuur.

Als aanvulling op het samenwerkingsverband zagen enkele content partners voornamelijk hun heil in samenwerking om de kosten te drukken, om de CEST-pagina's met resources en handleiding up-to-date te houden en als laatste in een *zelfhulpgroep* waarin vragen en issues behandeld kunnen worden.

Bijlage 2 – Sjabloon van een Evaluatieformulier

Seed: <i>Naam Account of Hashtag</i> URL	Pagefreezer	Webpreserver	Archive-It	Legende: Mogelijk met tool Niet Mogelijk met tool
				Opmerkingen
Analyse van de captatie				
1 Beperkingen dataset				
1.1	Beperking in tijd?			
1.2	Beperkt in aantal berichten?			
1.3	Onbeperkte captatie?			
2 Captatie van specifiek account				
2.1	Berichten van het account zelf?			
	corpus			
	datum			
	aantal comments/shares/likes			
2.3	Tijdelijke Stories?			
	corpus			
	datum			
	views			
2.4	Comments van derden op berichten?			
	corpus			
	datum			
	account			
	likes/dislikes			
2.5	Vermeldingen (tags)?			
	vermelding door			
	datum			

3 <i>Captatie van hashtags of trefwoorden</i>	
3.1 Berichten van openbare accounts?	
corpus	
datum	
aantal comments/shares/likes	
3.2 Berichten van privé-accounts?	
corpus	
datum	
aantal comments/shares/likes	
3.3 Tijdelijke Stories?	
corpus	
datum	
views	
3.4 Comments op de berichten?	
corpus	
datum	
account	
likes/dislikes	
4 <i>Ingebedde media</i>	
4.1 Webpagina's in links capteren?	
URL	
screenshot	
datum	
4.2 Ingebedde media capteren?	
video	
audio	
GIF	
Statische afbeelding	
5 <i>Interne communicatie</i>	
5.1 Interne berichten?	
corpus	
datum	
afzender	

ontvanger

6 *Omvang captatie*

- 6.1 Inhoudelijke data als tekst?
- 6.2 Look-and-feel aanwezig?
- 6.3 Profielpagina aanwezig?
 - gebruikersnaam
 - avatar
 - business info
 - contact info
 - more info (about)
 - story
 - milestones
- 6.4 Functionele weblinks aanwezig?
 - externe links
 - facebook events (interne link)

7 *Exportformaat*

WARC
JSON
CSV
PDF
EML

8 *Metadatering*

Titel
Datering dataset
Datering captatie
Archiefvormer
Bereik en inhoud
Selectie
Taal
Trefwoorden
Beschrijvingsniveau
Harvesttool
Configuratie tool

Harvester Account	
Platform URL seed Globale logfile	
9 <i>Extra metadata?</i>	
10 <i>Bijkomende elementen?</i>	

Bijlage 3 – Data Bevraging Onderzoekers

ID	Start time	Completion time	Email	Voor welke instelling en faculteit/departement werkt u?	Gebruikt u al gecapteerde data uit sociale media voor uw onderzoek/praktijk?
1	1/24/22 14:10:30	1/24/22 14:27:33	anonymous	Instituut voor Mediastud	Ja
2	1/24/22 17:42:04	1/24/22 17:49:42	anonymous	KU Leuven / CS Digital; P	Ja
3	1/25/22 8:55:32	1/25/22 9:08:39	anonymous	KU Leuven Centre for GI	Ja
4	1/25/22 12:58:02	1/25/22 13:27:02	anonymous	KU Leuven	Ja

Facebook	Instagram	Twitter	YouTube	TikTok	LinkedIn
Ja	Ja	Ja	Ja	Ja	Nice To Have
Ja	Nice To Have	Ja	Ja	Nice To Have	Ja
Ja	Ja	Ja	Ja	Ja	Nice To Have
Ja	Ja	Ja	Ja	Ja	Ja

WhatsApp	Vimeo	Ontbreekt er volgens jou nog een belangrijk platform aan bovenstaande lijst?	Ik wens captaties uit specifieke en openbare accounts.	Ik wens captaties uit private, afgeschermd accounts.	Ik wens captaties uit facebook pages en groups.
Ja	Nice To Have	Telegram; Discord; Facek	Ja	Ja	Ja
Nee	Ja	Niet-Europese/Amerikaa	Ja	Nee	Ja
Nee	Nee	Telegram (publieke kana	Ja	Nee	Nee
Ja	Nice To Have	Reddit, Telegram, Pinter	Ja	Ja	Ja

Ik wens captaties uit hashtags en trefwoorden.	Moeten enkel de berichten van een specifiek account opgenomen worden?	Moeten reacties van een specifiek account op berichten van derden gecaptiseerd worden?	Moeten reacties van derden op de berichten van een specifiek account gecaptiseerd worden?	Moeten berichten waarin een specifiek account vermeld wordt, de @'s dus, opgenomen worden?	Ik wens alleen tekstuele data van een post te kunnen raadplegen.
Ja	Ja	Nice To Have	Nice To Have	Nice To Have	Nice To Have
Ja	Nee	Nee	Ja	Ja	Nee
Ja	Ja	Ja	Ja	Nice To Have	Ja
Ja	Nee	Ja	Ja	Ja	Nee
Ik wens alleen metadata van een post te kunnen raadplegen.	Ik wens alleen de ingebedde media van een post te kunnen raadplegen.	Ik wens alleen de comments, likes en reacties op een post te kunnen raadplegen.	Ik wens de gehele post (en dus tekst, media, metadata en reacties) te kunnen raadplegen.	Moeten webpagina's, gelinkt in een post, gecaptiseerd worden?	Moeten interne berichten ("DM's" dus) binnen de platformen gecaptiseerd worden?
Nice To Have	Nice To Have	Nice To Have	Ja	Ja	Ja
Nee	Nee	Nee	Ja	Nice To Have	Nee
Ja	Ja	Ja	Ja	Ja	Nee
Nee	Nee	Nee	Ja	Ja	Nice To Have
Moeten de tijdelijke "stories" (specifiek voor Facebook, Instagram en Twitter) gecaptiseerd worden?	Moet de profielpagina van het account gecaptiseerd worden?	Is het voldoende om enkel ruwe data te harvesten?	Wens je de look and feel van een account te behouden?	Moet het mogelijk zijn om te kunnen zoeken en scrollen door de originele look-and-feel?	Moet ingebedde media nog functioneel zijn (dus afgespeeld en bekeken kunnen worden)?
Nice To Have	Ja	Ja	Nice To Have	Nice To Have	Ja
Ja	Ja	Nee	Ja	Nice To Have	Nice To Have
Nee	Nee	Ja	Nice To Have	Ja	Ja
Ja	Ja	Ja	Nee	Nee	Nice To Have

Moeten gelinkte webpagina's nog functioneel zijn (is een screenshot onvoldoende)?	Moet de dataset beperkt worden in de tijd?	Moet de dataset beperkt worden in het aantal gecapteerde berichten?	Moet de dataset beperkt worden op basis van gebruikte hashtags en/of trefwoorden?	Moet de dataset volledig onbeperkt zijn?	PDF
Ja	Nee	Nee	Nee	Ja	Nice To Have
Nee	Ja	Nee	Nee	Nee	Ja
Ja	Nee	Nee	Ja	Nice To Have	Ja
Nice To Have	Nee	Nee	Nee	Ja	Nee

WARC	JSON	TXT	CSV	Ontbreekt er volgens jou nog een belangrijk formaat aan bovenstaande lijst?	Titel
Nice To Have	Ja	Nice To Have	Nee		Ja
Nice To Have	Ja	Nice To Have	Ja		Ja
Nice To Have	Nice To Have	Ja	Ja		Ja
Nice To Have	Ja	Nee	Ja	Database	Ja

Datering dataset	Datering opname	Archiefvormer	Bereik en inhoud	Selectie	Taal
Ja	Ja	Ja	Ja	Ja	Ja
Ja	Ja	Ja	Ja	Nee	Ja
Ja	Ja	Nee	Ja	Ja	Nee
Ja	Ja	Ja	Ja	Ja	Ja

Trefwoorden	Harvesttool	Account	Platform	URL van de seed	Geolocatie
Ja	Ja	Ja	Ja	Ja	Ja
Ja	Nee	Ja	Ja	Ja	Nee

Ja	Ja	Ja	Ja	Ja	Nee
Ja	Ja	Ja	Ja	Ja	Ja

Ontbreekt er volgens jou nog belangrijke metadata aan bovenstaande lijst?	Datasets zijn meteen na captatie beschikbaar	Datasets zijn volledig beschikbaar voor download	Datasets zijn volledig beschikbaar voor download, mits anonimisering van persoonsgegevens	Datasets zijn beschikbaar voor download, zonder media en enkel tekst	Datasets zijn volledig beschikbaar voor raadpleging in de leeszaal
	Nice To Have	Nice To Have	Ja	Nee	Ja
	Nice To Have	Nice To Have	Nice To Have	Nee	Nee
Type bron (video; post; aJa		Nee	Ja	Nee	Nee
Inhoud van gelinkte web	Nice To Have	Ja	Ja	Ja	Nee

Het moet mogelijk zijn om datasets te gebruiken als bronmateriaal	Het moet mogelijk zijn om gebruikte datasets te presenteren in lezingen	Het moet mogelijk zijn om gebruikte datasets te presenteren in publicaties	Het moet mogelijk zijn om gebruikte datasets online te presenteren
Ja	Ja	Ja	Ja
Ja	Ja	Nice To Have	Ja
Ja	Ja	Ja	Ja
Ja	Nice To Have	Nice To Have	Nice To Have

Zie je nog andere manieren waarop je het gecapteerde materiaal wil raadplegen of gebruiken?

Raadplegen: via toegang tot een database, bij voorkeur zonder restricties op locatie, maar er kan eventueel wel verboden worden dat de gegevens lokaal worden opgeslagen.

Gebruiken: het is me niet helemaal duidelijk wat 'de gebruikte datasets presenteren' inhoudt - ik zou voornamelijke afgeleide analyses (in aggregaat) doen, en niet zozeer de dataset op zich.

Als er nog bijkomende voorstellen of opmerkingen zijn heeft u hier de ruimte om die mee te geven.

De captatie van DM's, individuele berichten of berichten in gesloten groepen lijken mij niet ethisch verantwoord en ik zou dit niet opnemen in de datasets. Ook dienen de meeste data geanonimiseerd te worden als het niet gaat over publieke instanties of organisaties.

Gerelateerd, een voorstel voor een standaard voor metadata die voor online advertenties beschikbaar moet zijn: <https://knightcolumbia.org/content/a-standard-for-universal-digital-ad-transparency>

Bijlage 4 – Data Bevraging Gebruikerservaring

Start time	Completion time	Voor welke content partner werkt u?	Welke tool werd getest?	De datastream bevatte alle berichten die ik wilde capteren.	De datastream bevatte alle comments die ik wilde capteren.
3/1/22 17:00:00	3/1/22 17:13:42	meemoo	archive-it	Niet Akkoord	Helemaal Niet Akkoord
3/7/22 7:57:51	3/7/22 8:06:42	ADV N	WebPreserver	Niet Akkoord	Niet Akkoord
3/7/22 10:40:06	3/7/22 10:53:35	Amsab-ISG	WebPreserver	Akkoord	Neutraal
3/8/22 16:09:57	3/8/22 16:20:06	M HKA	Pagefreezer	Helemaal Akkoord	Akkoord
3/18/22 16:18:27	3/18/22 16:24:16	meemoo	webpreserver	Niet Akkoord	Niet Akkoord
3/21/22 17:21:43	3/21/22 18:12:02	CAVA	Webpreserver	Helemaal Akkoord	Helemaal Akkoord
3/24/22 13:56:21	3/24/22 14:16:10	AMVB	WebPreserver	Akkoord	Niet Akkoord
3/25/22 9:13:59	3/25/22 9:19:57	Vlaams Architectuurinstituut	Archive-it	Helemaal Niet Akkoord	Helemaal Niet Akkoord

De datastream bevatte alle trefwoorden die ik wilde capteren.	De datastream bevatte alle ingebedde media die ik wilde capteren.	De datastream bevatte een representatieve weergave van de look & feel van het platform.	De datastream kon geëxporteerd worden in een gewenste vorm of formaat.	De datastream kan makkelijk en snel geëxporteerd worden	De datastream bevatte alle metadata die ik verwachtte.
Neutraal	Akkoord	Helemaal Akkoord	Helemaal Akkoord	Niet Akkoord	Akkoord
Niet Akkoord	Niet Akkoord	Neutraal	Akkoord	Akkoord	Niet Akkoord
Neutraal	Niet Akkoord	Helemaal Akkoord	Helemaal Akkoord	Helemaal Akkoord	Helemaal Akkoord
Akkoord	Akkoord	Akkoord	Niet Akkoord	Niet Akkoord	Akkoord
Neutraal	Helemaal Niet Akkoord	Neutraal	Akkoord	Akkoord	Akkoord
Neutraal	Niet Akkoord	Akkoord	Akkoord	Niet Akkoord	Akkoord
Neutraal	Akkoord	Akkoord	Neutraal	Akkoord	Akkoord
Helemaal Niet Akkoord	Helemaal Niet Akkoord	Neutraal	Neutraal	Helemaal Akkoord	Helemaal Akkoord

De datastream kan geupdate worden na verloop van tijd.	Indien de dataset niet alle elementen bevatte die u verwachtte, welke elementen ontbraken er dan precies?
Helemaal Akkoord	in geval van Facebook en Instagram ontbrak alle inhoud quasi. Bij Instagram zag ik enkel een login. ook bij Twitter werd niet de hele stream gecaptureerd (soms stopte het in 2019 of 2018) en de discussie ontbrak steeds. Het is ook moeilijk om de scope te bepalen. Als je bv. de tweets of foto's wil kunnen kopen, krijg je ook veel data mee die niet in scope zit.
Akkoord	Kon geen volledige Facebook- of Instagrampagina capteren, comments werden niet uitgekapt, het bevatte ook niet alle media (video's capteren lukt niet)
Neutraal	Er waren enkele problemen met de 'collect the links' feature, de autoscroll stopte soms en WARC files werkten niet goed.
Neutraal	Kon geen export maken in PDF noch in .warc
Niet Akkoord	video's en gif waren statische afbeeldingen geworden. er waren veel "witte stukken" in de datastream waar de posts in ontbraken. ik vond eerlijk gezegd dat de datastream eruit zag alsof er iedere pagina een screenshot gemaakt werd die vervolgens aan elkaar geplakt werden. Ik vond het er knullig uitzien. bij de bulk capture miste ik de mogelijkheid om één bestand te hebben met alle posts in. soms faalde ook de export van de WARC en kreeg ik een bestand van 0 bytes.
Akkoord	Externe links ontbraken. Een csv-overzicht van alle pagina's die binnen een bepaalde collectie gearhiveerd werden zou een meerwaarde zijn bij export van de dataset
Niet Akkoord	<p>Veel hangt af van het formaat waarin geëxporteerd word. Bij WebPreserver kan je kiezen o.a. PDF, OCR-PDF, MHTML, WARC en JPEG. Filmpjes kan je als aparte MP4's exporteren. PDF en OCR-PDF geeft getrouw de look en feel weer, maar mist alle functionele en interactieve elementen. Er kan niet op links of comments geklikt worden, en de filmpjes zitten er ook niet in. in de OCR-PDF kan je wel op tekst zoeken. JPEG is hetzelfde als PDF, maar in plaats van één bestand, worden de afzonderlijke snapshots als aparte JPEG's geëxporteerd.</p> <p>MHTML geeft wel een meer functionele versie van de pagina, maar deze kan enkel bekeken worden in Microsoft Edge. Daarenboven worden veel afbeeldingen (met name uit retweets) niet meegenomen.</p> <p>Ik heb de WARC-bestanden nog niet kunnen uitlezen. Vreemd genoeg geeft Windows Explorer dat deze bestanden 0 kB groot zijn. Ik weet niet in hoeverre ik hier belang aan moet hechten.</p> <p>Filmpjes worden als afzonderlijke MP4's geëxporteerd.</p> <p>Enkel bij de quick capture worden alle posts van een account in één bestand gebundeld. Bij de buk-capture worden de Facebook- en Instagramposts steeds in afzonderlijke bestanden ondergebracht.</p>
Helemaal Akkoord	<p>Archive-It raakte niet voorbij het Instagram login scherm. Look & feel van dit login scherm waren dan weer perfect ;-)</p> <p>We hebben ook een test gedaan met een gewone website. Daarvoor vond ik Archive-it indrukwekkend goed.</p>

Deze tool was intuïtief in gebruik.	Deze tool vereiste teveel menselijke interactie.	Deze tool was te omslachtig in gebruik.	Deze tool was te tijdsintensief.	Deze tool was voldoende snel in het capteren.	Deze tool leverde het verwachte resultaat.
Akkoord	Niet Akkoord	Niet Akkoord	Niet Akkoord	Akkoord	Akkoord
Akkoord	Akkoord	Niet Akkoord	Neutraal	Niet Akkoord	Niet Akkoord
Helemaal Akkoord	Akkoord	Neutraal	Akkoord	Niet Akkoord	Akkoord
Niet Akkoord	Niet Akkoord	Neutraal	Neutraal	Akkoord	Akkoord
Akkoord	Niet Akkoord	Niet Akkoord	Niet Akkoord	Neutraal	Niet Akkoord
Helemaal Akkoord	Helemaal Niet Akkoord	Niet Akkoord	Akkoord	Niet Akkoord	Helemaal Akkoord
Helemaal Akkoord	Neutraal	Niet Akkoord	Neutraal	Akkoord	Neutraal
Helemaal Akkoord	Helemaal Niet Akkoord	Helemaal Niet Akkoord	Helemaal Niet Akkoord	Helemaal Akkoord	Niet Akkoord

Deze tool maakt het mogelijk om op gezette tijdstippen captaties te updaten.	Ik zou deze tool systematisch gebruiken om data te capteren.	Ik zou deze tool sporadisch gebruiken als aanvulling op andere tools.	Ik kon de handleiding van de dienstverlener voor de gebruikte tool snel vinden.	Ik vond dat de handleiding van de dienstverlener duidelijke en begrijpelijke taal hanteerde.	Ik stootte op onverwachte foutmeldingen die niet in de handleiding beschreven stonden.
Helemaal Akkoord	Niet Akkoord	Niet Akkoord	Niet Akkoord	Akkoord	Niet Akkoord
Neutraal	Niet Akkoord	Niet Akkoord	Akkoord	Akkoord	Akkoord
Neutraal	Niet Akkoord	Akkoord	Helemaal Akkoord	Helemaal Akkoord	Akkoord
Neutraal	Neutraal	Neutraal	Niet Akkoord	Niet Akkoord	Neutraal
Akkoord	Niet Akkoord	Neutraal	Akkoord	Akkoord	Niet Akkoord
Akkoord	Akkoord	Neutraal	Helemaal Akkoord	Helemaal Akkoord	Akkoord
Niet Akkoord	Akkoord	Akkoord	Akkoord	Akkoord	Niet Akkoord
Helemaal Akkoord	Neutraal	Neutraal	Helemaal Akkoord	Helemaal Akkoord	Helemaal Akkoord

Ik vond dat de handleiding voldeed aan mijn verwachtingen.	Ik moest de hulp inroepen van anderen om de tool te installeren.	Ik moest de hulp inroepen van anderen om de tool te gebruiken.	Het supportstelsysteem van de dienstverlener was makkelijk bereikbaar.	Het supportstelsysteem van de dienstverlener bood snel een oplossing.	Het supportstelsysteem van de dienstverlener was voldoende transparant en begrijpbaar.
Akkoord	Helemaal Niet Akkoord	Helemaal Niet Akkoord	Neutraal	Neutraal	Neutraal
Neutraal	Niet Akkoord	Akkoord	Akkoord	Akkoord	Akkoord
Akkoord	Helemaal Niet Akkoord	Akkoord	Neutraal	Neutraal	Neutraal
Niet Akkoord	Niet Akkoord	Neutraal	Niet Akkoord	Neutraal	Niet Akkoord
Akkoord	Helemaal Niet Akkoord	Helemaal Niet Akkoord	Akkoord	Neutraal	Neutraal
Akkoord	Helemaal Niet Akkoord	Helemaal Niet Akkoord	Helemaal Akkoord	Helemaal Akkoord	Helemaal Akkoord
Akkoord	Helemaal Niet Akkoord	Helemaal Niet Akkoord	Neutraal	Neutraal	Neutraal
Niet Akkoord	Helemaal Niet Akkoord	Helemaal Niet Akkoord	Neutraal	Neutraal	Neutraal

Heeft u nog bijkomende opmerkingen of vaststellingen bij deze tool?

ik heb het supportstelsysteem niet gebruikt.
moeilijk om scope vast te leggen.
moest ik deze tool gebruiken, dan zou ik ze voor Twitter gebruiken.
Ik vond het dashboard soms ook heel traag laden.

De tool vergt toch nog teveel menselijke aanwezigheid aangezien er problemen kunnen zijn met de autopilot.

ik heb het supportstelsysteem niet gebruikt.

We waren eerder teleurgesteld in de WARC-bestanden die geëxporteerd werden omdat deze niet de look en feel volledig meenamen. De printscreens in jpg en de ge-OCR'de pdf compenseren dit wel, maar 't is jammer dat we dan meerdere bestandstypes moeten bijhouden voor het geheel gearchiveerd te hebben.

De resultaten van deze tool zijn sterk afhankelijk van het sociale medium dat gecapteerd wordt. Bij Twitter waren er helemaal geen problemen, en kan alles makkelijk gecapteerd worden.

Facebook was al wat moeilijker, maar met de 'bulk-capture' optie kon wel de gevraagde dataset gecapteerd worden. De quick-capture liep bij momenten vast. De ervaring leert dat bij het capteren van afgebakende periodes (i.p.v. de volledige account) het programma beter werkt, en veel minder vast loopt.

Voor Instagram is deze tool minder aangewezen: de bulk-capture verloopt zeer traag en leek regelmatig vast te lopen (of het capteren ging zo traag dat er niets leek te gebeuren). Ik moest bij het capteren

Supertool voor websites.

