

## Module 6 Praktijkvoorbeelden

In deze module worden diverse praktijkvoorbeelden toegelicht. Verschillende organisaties binnen en buiten Nederland delen hun ervaringen en geleerde lessen die zijn opgedaan bij het archiveren van sociale media.

- [KADOC-KU Leuven en Meemoo](#)
- [Nationaal Archief](#)
- [Regionaal Archief Alkmaar](#)
- [Stadsarchief Amsterdam](#)

### <Uitklapmenu> Leeswijzer praktijkvoorbeelden

De structuur van de praktijkvoorbeelden is als volgt:

- Samenvatting: over het project of onderzoek dat verricht is naar sociale media archivering.
- Aanpak: geeft een algemeen overzicht hoe onderzoeken en/of projecten zijn verlopen.
- Uitdagingen en complexiteit, waarbij aandacht is voor juridische uitdagingen en toegepaste technieken.
- Lessons learned, waarbij indien relevant, ook wordt er gerefereerd naar verdere literatuur en onderzoek.

### <Uitklapmenu> Algemene lessons learned

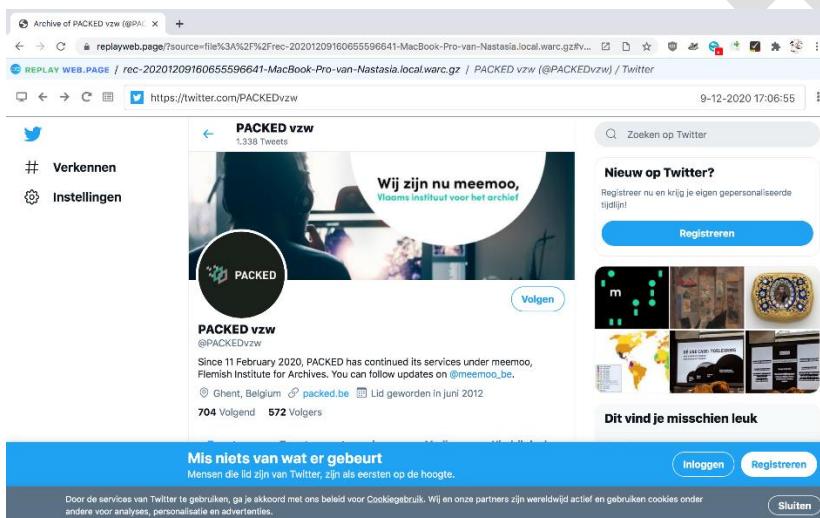
Elk praktijkvoorbeeld deelt de belangrijkste geleerde lessen die zijn opgedaan. Hieronder worden de belangrijkste benoemd.

- Vanwege het experimentele karakter van sociale media archivering wordt aangeraden om ervaringen die zijn opgedaan te delen in het netwerk, voor kennisuitwisseling.
- Vanwege het veranderende karakter van zowel de sociale media platformen zelf, als de tools waarmee het binnengehaald kan worden, is het van belang dat bij het gebruik van het account al plannen worden gemaakt voor hoe en wanneer het gearchiveerd wordt.
- Ten slotte valt op dat het testen en gebruiken van tools lastig is op overheidslaptops.
- Het advies is daarom om in gesprek te gaan met de ICT afdeling over de mogelijkheden om dat toch te kunnen bewerkstelligen.

Kijk voor meer specifieke lessons learned bij elk praktijkvoorbeeld.

## <Subpagina> KADOC-KU Leuven en Meemoo

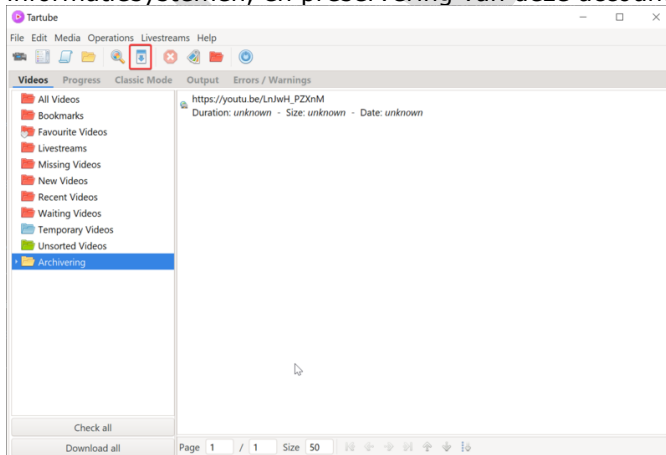
In de periode 2020 tot 2023 deed [KADOC-KU Leuven](#) in samenwerking met [Meemoo](#) en dertien andere partners onderzoek naar het archiveren van sociale media. Aan de basis van dit onderzoek stonden twee ontwikkelingen. Ten eerste de verschuiving van klassieke papieren uitgaven naar digitale publicaties zoals nieuwsbrieven en sociale media. Ten tweede was het de verwerving van nieuwe archieven zoals die van migrantengroepen van belang dat instellingen zoals KADOC-KU Leuven ook in staat zijn om digitale kanalen zoals sociale media te archiveren. Gezien de beperkte ervaring en grote uitdagingen op zowel technisch als juridisch vlak werd besloten om het project te starten met een beperkt aantal platformen (Facebook, Instagram, Twitter en YouTube) en accounts. KADOC startte met het archiveren van hun eigen accounts en die van de politieke partij CD&V. Meemoo gebruikte de sociale media van hun voorloper als testcase. Daarnaast hebben ook de andere projectpartners met de medewerking van bereidwillige archiefvormers het archiveringsproces getest. Bij de selectie van testcases is gelet op diversiteit zodat de resultaten van het project relevant zijn voor een zo groot mogelijke groep gebruikers. Hieronder worden de aanpak, uitdagingen en de geleerde lessen tijdens dit onderzoek nader toegelicht.



*De gearchiveerde twitter van PACKED, een voorloper van Meemoo.*

## <Uitklapmenu> Aanpak

Het project was opgedeeld in drie fasen. In de eerste fase (1/9/2020 - 31/8/2021) lag de focus op het ontwikkelen van best practices voor het archiveren van individuele accounts door middel van gratis verkrijgbare opensourcetools. Dit werd gedaan in een aantal pilots waarin aandacht werd besteed aan het binnenhalen van accounts, het opnemen van de binnengehaalde accounts in de informatiesystemen, en preservering van deze accounts.



*Tartube, één van de tools waarmee YouTube content is gearchiveerd.*

In de tweede fase (1/9/2021 - 31/8/2022) lag de focus op het binnenhalen van (individuele) accounts en sociale mediaberichten op basis van een trefwoord (hashtags). Dit werd gedaan met behulp van API's en commerciële tools, zoals Pagefreezer en Archive-It. Zowel gebruiksvriendelijkheid als kwaliteit van het informatieobject werden steeds in overweging genomen. Tevens werd tijdens deze fase gekeken naar het juridische deel van het archiveren van sociale media. Hier werd geconcludeerd dat de juridische context zowel per land als casus sterk kan verschillen en dat maatwerk nodig kan zijn. Meer informatie over het juridische aspect van het archiveren van sociale media in Nederland kan je vinden in [de module juridisch kader](#).

In de derde en laatste fase (1/9/2022 - 31/8/2023) werd de focus verlegd naar beschikbaarstelling. Hierin werd onderzoek gedaan naar hoe de toegang tot, en het hergebruik van, gearchiveerde sociale media kan worden gerealiseerd. De nadruk hierin lag op het onderzoeken van de verschillende tools die hiervoor gebruikt kunnen worden. Ook aan het juridische deel werd in de derde fase wederom aandacht besteed. Daaruit bleek dat terbeschikkingstelling enkel in de archiefleeszaal kan onder voorwaarde dat privacy gerespecteerd wordt. Veel archiefinstellingen beslisten om enkel op gemotiveerde vraag van onderzoekers gearchiveerde berichten ter beschikking te stellen. Het project werd afgesloten met een [studiedag](#) waarin de bevindingen van de drie fases werden gedeeld.

## <Uitklapmenu> Uitdagingen en complexiteit

### <Titel> Juridisch

Het archiveren van sociale media is onderhevig aan juridische vraagstukken waar geen eenduidige antwoorden op zijn, gezien ze verbonden zijn aan het juridische statuut van de archiverende partij. Zo zit je met vraagstukken rondom de gebruikersvoorwaarden, auteursrecht, gegevensbescherming en het portretrecht. Oplossingen voor deze vraagstukken zijn van inhoudelijke, organisatorische en juridische aard.

Focus inhoudelijk op bijvoorbeeld het archiveren van enkel publiek toegankelijke accounts en publiek toegankelijke gegevens om risico's die de eerder genoemde vraagstukken met zich meebrengen zoveel mogelijk te verkleinen. Bij het verwerken van persoonsgegevens (wat zich snel voordoet in het geval van sociale media archivering) is het belangrijk om de nodige organisatorische maatregelen te treffen in de vorm van het bijhouden van een verwerkingsregister, procedures voorzien bij inbreuken en tevens een functionaris aanwijzen voor gegevensbescherming. Tot slot kunnen archiefinstellingen het cultureel erfgoedbelang opnemen in de missie van de organisatie of de collectiebeleidsplannen, om zo het juridisch kader voor sociale media archivering te versterken.

Lees in het [juridische eindrapport](#) meer over de bredere uiteenzetting rondom de juridische vraagstukken.

### <Titel> Techniek

Gedurende het project waren er veel ontwikkelingen achter de schermen bij de sociale media platformen in kwestie. Denk bijvoorbeeld aan de inperkingen/veranderingen met de API van Twitter. Deze ontwikkelingen hadden effect op de werking van de tools die gebruikt werden om de sociale media te archiveren. Het was daarom goed om op de hoogte te blijven van deze ontwikkelingen en te blijven testen met de tools of deze nog goed werkten. Uit tests bleek ook dat er geen duidelijk verschil was tussen open-source en commerciële tools. Beide leveren dezelfde typen resultaten en zijn geen totaaloplossing voor alle platformen. Waar tools zoals Webpreserver en Archive-It wel uit de bus kwamen als werkbare opties, werd er in de praktijk wel gekozen voor open-source tools omdat de commerciële tools ruimschoots boven het budget van de projectpartners lagen.

## <Uitklapmenu> Lessons learned

- In de Europese regelgeving zijn er uitzonderingsmaatregelen van kracht voor het behoud van cultureel erfgoed en het archiveren van persoonsgevoelige gegevens (AVG, is in Vlaanderen en internationaal bekend als de GDPR) voor erfgoedinstellingen en wetenschappelijk onderzoek, waardoor er een basisrecht is om sociale mediaberichten te

archiveren. Wacht dus niet tot het te laat is en de berichten verdwenen, maar ga aan de slag.

- Tijdens het archiveren van sociale media moet je ingelogd zijn om de materie achter de inlogpagina te kunnen zien en dus te kunnen binnenhalen. Het advies is om daarvoor niet je persoonlijke sociale media account te gebruiken. Het risico bestaat dan namelijk dat onderdelen van het account waarmee je bent ingelogd worden binnengehaald, of zelfs je wachtwoord. Daarnaast kan de toegang tot je eigen account geblokkeerd worden, omdat je inbreuk pleegt op de voorwaarden van het platform als je accounts binnenhaalt.
- Test regelmatig de workflow. De techniek achter de platformen verandert voortdurend en hierdoor kan je manier van werken al snel verouderd raken.
- Wees niet terughoudend en deel je ervaringen. Dit kan bijvoorbeeld door deelname aan een praktijknetwerk. Meer samenwerking binnen de sector kan leiden tot een sterkere lobby richting de (Europese) politiek en dus wetgeving.
- Er is geen tool die alle platformen kan archiveren. Er moet dus per platform gekeken worden naar de beste oplossing.

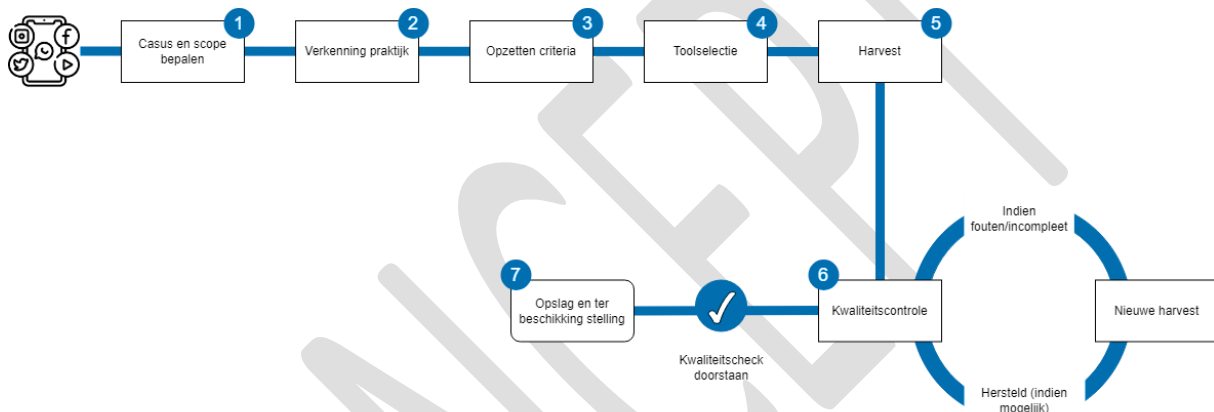
Wil je meer lezen over het project, het praktijknetwerk en het evaluatierapport? Dit is te lezen op de website van [KADOC-KU Leuven](#).

CONCEPT

## <Subpagina> Nationaal Archief

Eind 2022, begin 2023 heeft het Nationaal Archief een onderzoek uitgevoerd naar de mogelijkheden tot het archiveren van het publieksaccount op Twitter. Dit account werd in 2009 geopend en werd gebruikt voor het delen van nieuws relevant voor het algemene publiek. In 2015 kwam daar een [zakelijk account](#) bij. Om verwarring te voorkomen is in september 2021 ervoor gekozen om alle communicatie voortaan via het zakelijke account te laten verlopen. Het publieksaccount is vervolgens afgeschermd, maar niet opgeheven. Voordat het account daadwerkelijk kan worden opgeheven dient het eerst gearchiveerd te worden. Mede vanwege de noodzaak tot opheffing en het feit dat sociale media archivering een steeds urgenter onderwerp werd binnen de rijksoverheid, besloot het Nationaal Archief om een onderzoek te starten naar de verschillende technieken en tools. Het oorspronkelijke publieksaccount werd hiervoor als casus gebruikt. Hieronder worden de aanpak, uitdagingen en de geleerde lessen tijdens dit onderzoek nader toegelicht.

## <Uitklapmenu> Aanpak

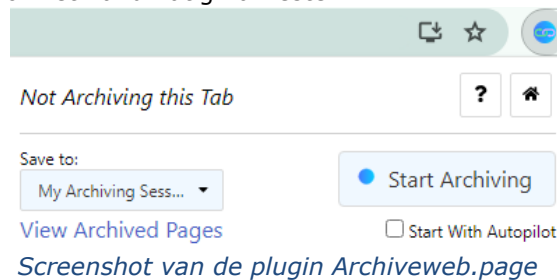


Het hele onderzoek had een doorlooptijd van een half jaar. Het onderzoek was opgedeeld in vier fasen: inventarisatie, analyse, evaluatie en terugkoppeling. Een interdisciplinair team van deskundigen op het gebied van recordkeeping, preservation en digitaal verwerven kwam voor dit project samen.

Als eerste is de casus en de scope bepaald. In overleg met de betrokken afdelingen is ervoor gekozen om het oude publieksaccount als casus te nemen voor dit onderzoeksproject. De scope omvat alle berichten van het account, vanaf april 2009 tot september 2021. Vervolgens zijn de bestaande informatiebronnen over sociale media archivering in kaart gebracht en geraadpleegd. De meest gebruikte informatiebronnen waren van het Vlaamse archiefinstituut [Meemoo](#) en het [Netwerk Digitaal Erfgoed](#).

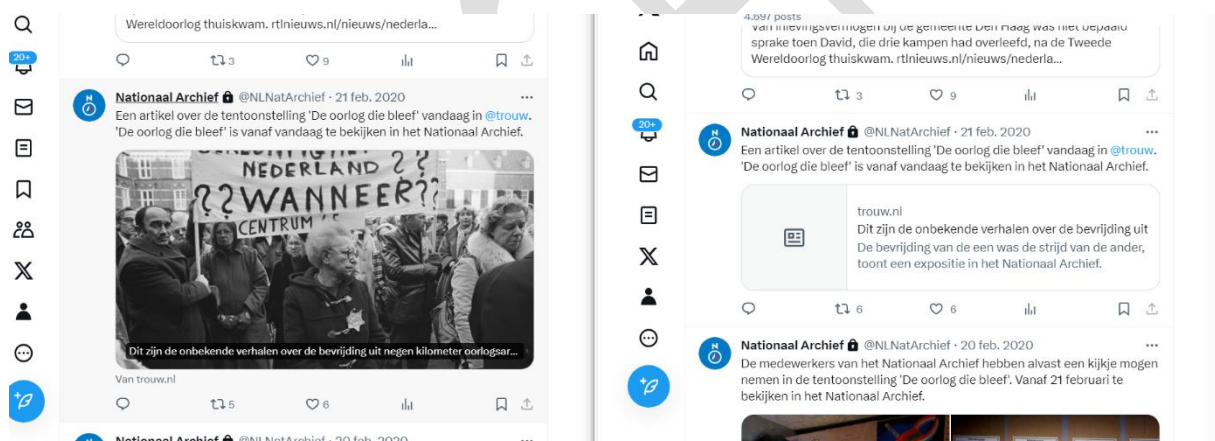
Na het in kaart brengen van alle mogelijke methodes en gebruikte technieken is het projectteam weer in gesprek gegaan met de stakeholders om de eisen en wensen op te halen met betrekking tot selectie van een tool. De eisen en wensen kwamen neer op twee zaken. Ten eerste moest de opslag van de bestanden volgens de EU-regelgeving gaan om aan de wetgeving (zie [juridisch kader](#)) te voldoen. Ten tweede was het van belang dat de look-and-feel van het platform meekwam. Dit houdt in dat het design van Twitter behouden werd, denk hierbij aan lay-out en lettertypes, maar ook het gedrag van dynamische elementen zoals bepaalde knoppen.

Met behulp van de eisen kon het projectteam al snel een shortlist opstellen van twee tools: ArchiveWeb.page en Browsertrix. Beide zijn open-source tools die vallen onder het project Webrecorder. ArchiveWeb.page is een extensie bij de Chrome browser en zorgt ervoor dat je je eigen scherm opneemt. Je harvest dus wat je zelf open hebt staan en aanklikt. Het is eenvoudig in gebruik, maar wel tijdsintensief door het handmatig harvesten.



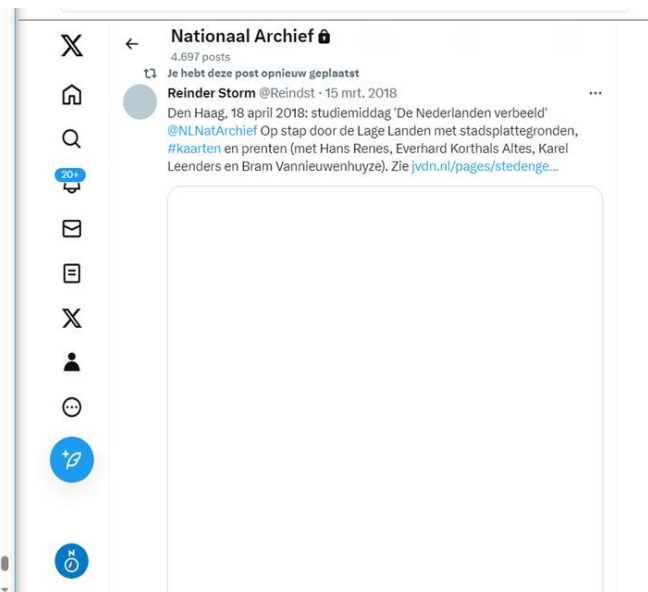
Screenshot van de plugin Archiveweb.page

Het resultaat van Archiveweb.page na het binnenhalen is een WACZ (Web Archive Collection Zipped) bestand. In dit bestandstype, in feite dus een zip-container, zitten vervolgens weer de WARC-bestanden samen met uitgebreide metadata. Het WARC bestandsformaat is een voorkeursformaat voor webarchieven en bevat tevens de look-and-feel van Twitter. Ook de andere tool op de shortlist, Browsertrix, leverde een WACZ op. Fijn aan deze tool was dat het minder handmatig was. Echter was het nog niet perfect werkend op het moment van testen en waren er bugs aanwezig die het proces compliceerden. De in 2022 ingevoerde nieuwe beperkingen van Twitter leidden tot niet-complete harvests. Omdat dit project ging over één account kon dit handmatig en is er daarom uiteindelijk gekozen voor Webrecorder.



Links is de tweet te zien zoals deze op live X te vinden was, rechts is het resultaat te zien van de gegenereerde WACZ van de tool Browsertrix. Door de beperkingen van het sociale media platform is het account niet volledig en mist hier dus de afbeelding.

Na de keuze voor de tool Webrecorder is een collega hiermee aan de slag gegaan. Het oude account van het Nationaal Archief is voor zover mogelijk geharvest. Helaas bleek na evaluatie dat enkel de tweets uit het jaar 2019 geharvest konden worden vanwege de beperkingen tot inzage (maximaal 2.000 tweets per dag). Dit zou omzeild kunnen worden door tijdelijk een betaald account te hebben voor de periode van de harvest. De kwaliteit van het gerealiseerde deel van het Twitter-harvest was goed, maar niet perfect. Onder andere de embedded video en sommige emoji's zorgden voor artefacten in de weergave. Dat betekent dus dat de content er niet is, maar wel een signalering dat laat zien dat het er hoort te zijn.



*Links een screenshot van een post van live Twitter. Rechts een screenshot van diezelfde post, die is binnengehaald. Rechts is te zien dat een afbeelding ontbreekt.*

Na de harvest is er teruggekoppeld aan de stakeholders. Tijdens deze terugkoppeling is de harvest besproken en de uitdagingen die langs zijn gekomen. Afsluitend is begonnen met de opslag en ter beschikking stelling van de harvest. Het resultaat van de harvest is in het zaakstelsel geplaatst. Er is op het moment van schrijven nog geen besluit genomen over ter beschikking stelling aan het publiek. De harvest behoort namelijk tot het instellingsarchief van het NA en wordt binnen te zijner tijd overgebracht naar de collectie van het Nationaal Archief. Dan wordt de harvest als openbaar stuk overgebracht.

## <Uitklapmenu> Uitdagingen en complexiteit

### <Titel> Juridisch

Zoals eerder genoemd was het belangrijk dat de opslag van de bestanden voldeed aan EU-regelgeving. In artikel 3 van de Algemene Verordening Gegevensbescherming (AVG) staat vermeld dat alle EU-landen moeten voldoen aan dezelfde regels rondom gegevensbescherming. Bij landen buiten de EU gelden andere regels. Tijdens dit onderzoek is daarom besloten dat opslag in de EU een vereiste was wat onze keus in tools inperkte.

### <Titel> Context

Zonder veel vooronderzoek is de Rijksoverheid gestart met Twitter. Daardoor is er bij ingebruikname weinig aandacht besteed aan het archiveren van dit [type overheidsaccounts](#). Dit heeft er ook bij het account van het Nationaal Archief toe geleid dat er na twaalf jaar van gebruik nog niets van het platform was gearchiveerd. De overname van Twitter door Elon Musk heeft getoond hoe complex het is om bij een commercieel platform controle te blijven houden over de eigen content. Door beperkingen rondom toegang werd het lastig om het gehele account te archiveren. Ook de functie waarin Twitter zelf mogelijkheid tot het downloaden van de eigen data aanbiedt, is verre van optimaal. Zo verdwijnt de oorspronkelijke look and feel en zijn comments/reacties van anderen niet zichtbaar.

### <Titel> Techniek

Doordat harvesten via de API tevens werd belemmerd door de nieuwe beperkingen vanuit Twitter, zijn technieken die daar gebruik van maakten niet meer optimaal. Dat in combinatie met het beperken van inzage tot een maximum van 2000 tweets per dag zorgt ervoor dat accounts met meer

dan 2000 tweets niet naar behoren kunnen worden gearhiveerd (dus onbetaald en voor iedereen toegankelijk). Al deze beperkingen maakten het daarom tot een uitdaging om tot een volledige harvest te komen die aan de kwaliteitseisen van het Nationaal Archief voldeed. Voor toekomstige projecten zal hiermee rekening worden gehouden.

### **<Uitklapmenu> Lessons learned**

- De keuze voor de methode van archiveren is erg bepalend voor het resultaat. Voor het Nationaal Archief werkte het daarom goed om dit direct mee te nemen in de keuze voor de uiteindelijke techniek.
- Maak voordat een account op sociale media wordt aangemaakt een plan over hoe het betreffende account gearhiveerd kan worden. Doe dit bij voorkeur in overleg met de afdeling communicatie. Bij het Nationaal Archief is dit inmiddels besproken met communicatie en zal er onderling meer contact zijn hierover tussen de informatie- en communicatieprofessionals.
- Wacht niet te lang met het archiveren van content op sociale media. Ga hier proactief mee om.
- Betrek bij het archiveren van sociale media de afdeling communicatie. Zij hebben immers het account gecreëerd en beheerd en weten waar het uit bestaat.
- Houd de technische ontwikkelingen van het archiveren van sociale media in de gaten en blijf streven naar verbetering en oplossingen. Het is een dynamisch veld waarin veel gebeurt aan zowel de kant van de platformen (denk hierbij aan ontwikkelingen en beperkingen) als aan de kant van de archiveringstechnieken. Gedurende het project waren er veel veranderingen door de overname van Twitter, daarom is het belangrijk om tijdens het proces de ontwikkelingen in de gaten te houden.



## <Subpagina> Regionaal Archief Alkmaar

In 2021 zijn twee gemeenten in de regio Alkmaar gefuseerd. Zij wilden online graag een frisse start maken door hun oude sociale media accounts op te doeken en met één gezamenlijk account verder te gaan. Gezien de verplichting tot het archiveren van overheidsinformatie hebben zij de hulp van het [Regionaal Archief Alkmaar \(RAA\)](#) ingeschakeld voor het archiveren van deze accounts. Beide gemeenten hebben via diverse sociale media kanalen (Facebook, Instagram, Google Drive (YouTube) en Twitter X) contact gehad met bewoners en andere partijen. De accounts die beide gemeenten hiervoor gebruikten waren accounts die via het zakelijke e-mailadres van de medewerkers waren aangemaakt. Ze werden voor strikt zakelijke doeleinden ingezet. De accounts bevatten naast openbaar gedeelde informatie ook chatberichten tussen de gemeenten en bewoners. Hieronder worden de aanpak, uitdagingen en de geleerde lessen nader toegelicht.

## <Uitklapmenu> Aanpak

Tijdens de fusie tussen de twee gemeenten is er contact gezocht met het RAA over het archiveren van hun sociale media accounts. Het RAA heeft geen inspraak gehad over de wijze van archiveren of manier van aanleveren van de accounts.

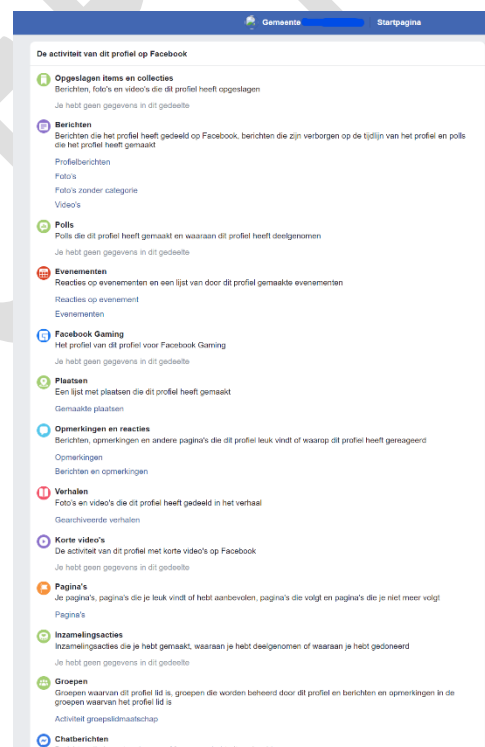
Het RAA kreeg de gearchiveerde sociale media accounts van de twee gefuseerde gemeenten op een harde schijf binnen. De ontvangen bestanden zijn daarna bekeken en in kaart gebracht door de adviseurs digitale informatie van het RAA. Hierbij werd duidelijk dat er gebruik was gemaakt van de archiveringsfunctie van diverse platformen: Facebook, Instagram, Google Drive (platform Youtube), en Twitter (nu X).

Er werd ook geconstateerd dat het om corporate accounts ging (zie [overheidsinformatie op sociale media](#)), die echter wel waren aangemaakt met de zakelijke e-mailadressen van individuele medewerkers.

Deze eerste inventarisatie leverde een plan van aanpak op waarbij de situatie werd geschetst, een uitgebreide analyse werd gedaan van de exports uit de diverse platformen, en aandachtspunten en risico's in kaart werden gebracht. De voornaamste punten die hieruit voort kwamen waren de omgang met persoonsgegevens, diversiteit in bestandsformaten en de mappenstructuur (zie uitdagingen en complexiteit).

Na het maken van een plan van aanpak is er gekozen om (nog) meer informatie in te winnen over sociale media archivering. Het RAA heeft diverse archiefinstellingen met ervaring op dit gebied bevroegd en ook meerdere gesprekken gevoerd. Bijvoorbeeld over hoe dit materiaal goed te preserven is. En hoe WARC en HTML zich verhouden tot elkaar. Daarnaast is er gebruik gemaakt van de handreiking voorkeursformaten van het Nationaal Archief en zijn diverse artikelen over het archiveren van sociale media als basis gebruikt.

Met de opgedane kennis konden de gearchiveerde accounts opgenomen worden in het informatiesysteem van het RAA. Dit is op twee manieren gedaan bij de accounts van de Facebook en Instagram. De eerste manier is in de vorm van een HTML waarbij de mappenstructuur is teruggebracht naar één maplaag. Dit was noodzakelijk, omdat de aangeleverde structuur niet kon worden opgenomen in het e-depot van het RAA. Deze structuur bestond uit verschillende mappen, die elk ook onderliggende mappen en bestanden konden bevatten. De oorspronkelijke mappenstructuur werd dus aangepast naar één maplaag, zodat het wel



*Startpagina gearchiveerde Facebook-account. Te zien wanneer de [index.html](#) wordt geopend (voor meer informatie, zie module 5 - archiveringsfunctie platform)*

opgenomen kon worden in het e-depot. Bovendien was het belangrijk voor het RAA om zo dicht mogelijk bij de oorspronkelijke mappenstructuur en bestandsformaten (HTML) te blijven. De tweede manier is een ZIP, waarin de export van het platform oorspronkelijk ook werd aangeboden. Bezoekers zullen daarbij de mogelijkheid hebben om de gearchiveerde accounts te bekijken via een beveiligde computer in de studiezaal van het RAA.

## <Uitklapmenu> Uitdagingen en complexiteit

### <Titel> Juridisch

Op juridisch vlak was het archiveren van het materiaal een uitdaging. Zo bevatten de gearchiveerde accounts privé informatie waar zorgvuldig mee om moet worden gegaan. In de gearchiveerde accounts zijn persoonsgegevens te vinden die variëren van persoonsnamen tot bijzondere persoonsgegevens, waarbij een medische status wordt genoemd. Het RAA heeft gekeken naar wetgeving zoals de Archiefwet, Algemene Verordening Gegevensbescherming en het Auteursrecht om informatie in te winnen rondom dit onderwerp. De opgenomen gearchiveerde accounts zullen worden aangeboden op de website van het RAA. Echter zal het grootste deel van de HTML niet beschikbaar zijn vanwege openbaarheidsbeperkingen, er zit namelijk heel veel persoonlijke informatie in de HTML. Ook zal de ZIP in zijn geheel niet te openen zijn. Wel zullen de accounts toegankelijk zijn op locatie vanaf een beveiligde computer op de studiezaal. In deze [blog](#) is meer te lezen over hoe er wordt omgegaan met persoonsgegevens in het sociale media archief.

### <Titel> Bestandsformaten

De exports bevatten een hoeveelheid aan bestandsformaten, waarbij sommige niet worden beschouwd als duurzaam en ook niet op de lijst van acceptabele formaten staan. Er is echter voor gekozen om deze allemaal op te nemen in plaats van te converteren. Hierbij kan namelijk informatieverlies ontstaan. Ook is er veel aandacht uitgegaan naar de verschillen tussen de bestandsformaten HTML (het oorspronkelijke formaat) en WARC (voorkeursformaat bij webarchieven). In dit geval is ervoor gekozen om de HTML niet om te zetten naar het WARC formaat in verband met mogelijk informatieverlies. Daarnaast heeft het opslaan van de bestanden op deze manier meer voordelen. Wanneer je de bestanden als een ZIP zou opslaan, preserveer je de individuele bestanden niet, hoewel dat wel de bedoeling is. Ook blijven de interne verwijzingen naar de verschillende bestanden behouden, zodat de informatie op de juiste manier getoond wordt wanneer je de berichten bekijkt. Tot slot, indien het voorkomt dat er beperking van openbaarheid op onderdelen van het sociale media archief zijn, hoeft dat niet te betekenen dat het gehele archief niet in te zien is. Indien de bestanden opgeslagen zijn in een ZIP, geldt deze beperking voor de gehele ZIP. Echter, op deze manier is het mogelijk om de beperking enkel op het desbetreffende bestand toe te passen.

### <Titel> Mappenstructuur

Met de downloadfunctie van de platformen krijg je een ZIP. Na het uitpakken van deze ZIP krijg je een mappenstructuur met daarin de eerder beschreven diversiteit aan bestandsformaten. Binnen de combinatie van systemen bij het RAA kan er maar één map worden opgenomen en daar mogen geen submappen in zitten. Er is door het RAA veel geëxperimenteerd met diverse manieren om dit op te lossen. Bij het terugbrengen naar één maplaag van de mappenstructuur krijg je namelijk dubbele bestandsnamen. Tevens moet er rekening gehouden worden met onderlinge verwijzingen in de diverse bestanden. Voor Facebook en Instagram is het aanpassen van de mappenstructuur gelukt,

```
N: .
  index.html
  oorspronkelijke_mappenstructuur.txt
+---activity_messages
      no-data.txt
+---affiliate
      no-data.txt
+---bonuses
      no-data.txt
+---comments_and_reactions
      comments.html
      posts_and_comments.html
+---icons
      anger.png
      confused.png
      dorothy.png
      flame.png
      haha.png
      like.png
      love.png
      plane.png
      selfie.png
      sorry.png
      toto.png
      wow.png
      yay.png
+---communities
      no-data.txt
+---events
      events.html
      event_responses.html
+---facebook_gaming
      no-data.txt
+---friends_and_followers
      people_who_follow_you.html
      who_you_follow.html
+---fundraisers
      no-data.txt
+---groups
      group_membership_activity.html
+---live_audio_rooms
      no-data.txt
```

*Mappenstructuur na archivering van het Facebook-account*

waarbij de context zoveel mogelijk is behouden door de mappenstructuur op te slaan als een boomstructuur en door alle aanpassingen te noteren (inclusief verwijzingen) in een apart bestand. Helaas bleek deze methode niet mogelijk voor de accounts van Google Drive en Twitter. Dit wordt nader onderzocht.

## <Uitklapmenu> Lessons learned

- Kennis rondom sociale media archivering staat nog aardig in de kinderschoenen. Het is nog veel pionieren momenteel, zeker met de diversiteit in mogelijkheden waarop sociale media gearchiveerd kan worden. Daarom is het handig om informatie te blijven vergaren, omdat er gedurende de tijd steeds meer komt. Dit is één van de redenen waarom het RAA ervoor gekozen heeft om de accounts zo te preserven dat er in de toekomst nog veel mogelijk is.
- Het is goed om voorbereid te zijn op de manieren waarop je sociale media kunt archiveren. Op basis van de tijd en context (e-Depot en collectiebeheersysteem/toegang) kan vervolgens dan de beste optie gekozen worden.
- Ga en blijf in gesprek met andere (archief)instellingen die bezig zijn met het archiveren van sociale media. Ondanks dat systemen verschillen blijft het heel nuttig om naar elkaars bevindingen te kijken.
- Probeer, indien mogelijk, zo snel mogelijk het gesprek over sociale media archivering en toegang aan te gaan met de archiefvormer. Zo kan er gezamenlijk gekeken worden naar de beste mogelijkheden voor beide kanten. Het liefst is dit al voordat er een sociale media account gestart wordt, om zo de best mogelijke keuzes te maken voor een toekomstige export.
- Zorg ervoor dat je actie onderneemt om de accounts veilig te stellen in je e-depot, ook al is iets niet perfect. Dit is beter dan iets op een locatie houden waar het kwetsbaarder is, zoals op een harde schijf.

## <Subpagina> Stadsarchief Amsterdam

In 2019 is het Stadsarchief Amsterdam gestart met een pilot voor het archiveren van de sociale media accounts van een aantal archiefvormers. De aanleiding om een pilot te beginnen kwam eind mei 2015. Een aantal voormalige bezetters van het Maagdenhuis, vroeg aan het Stadsarchief Amsterdam hoe zij hun sociale media accounts konden archiveren en of zij deze konden overbrengen naar het e-depot van het stadsarchief. Dit was geen vastomlijnd, doorlopend project, maar een pilot waar sporadisch aan gewerkt werd. Tijdens de coronapandemie is deze pilot voornamelijk vanuit huis en op privé laptops verder uitgevoerd, met behulp van open-source tools die relatief makkelijk te installeren en te gebruiken waren. De pilot richtte zich met name op hoe sociale media binnengehaald kon worden en hoe het materiaal vervolgens in het e-depot opgeslagen kon worden. Er was weinig voorkennis voor het archiveren van sociale media, maar wel enige ervaring met webarchivering. Uiteindelijk is er gebruik gemaakt van een deels geautomatiseerde aanpak, waar veel handmatige controles noodzakelijk waren. Hieronder worden de aanpak, uitdagingen en de geleerde lessen tijdens deze pilot nader toegelicht.

## <Uitklapmenu> Aanpak

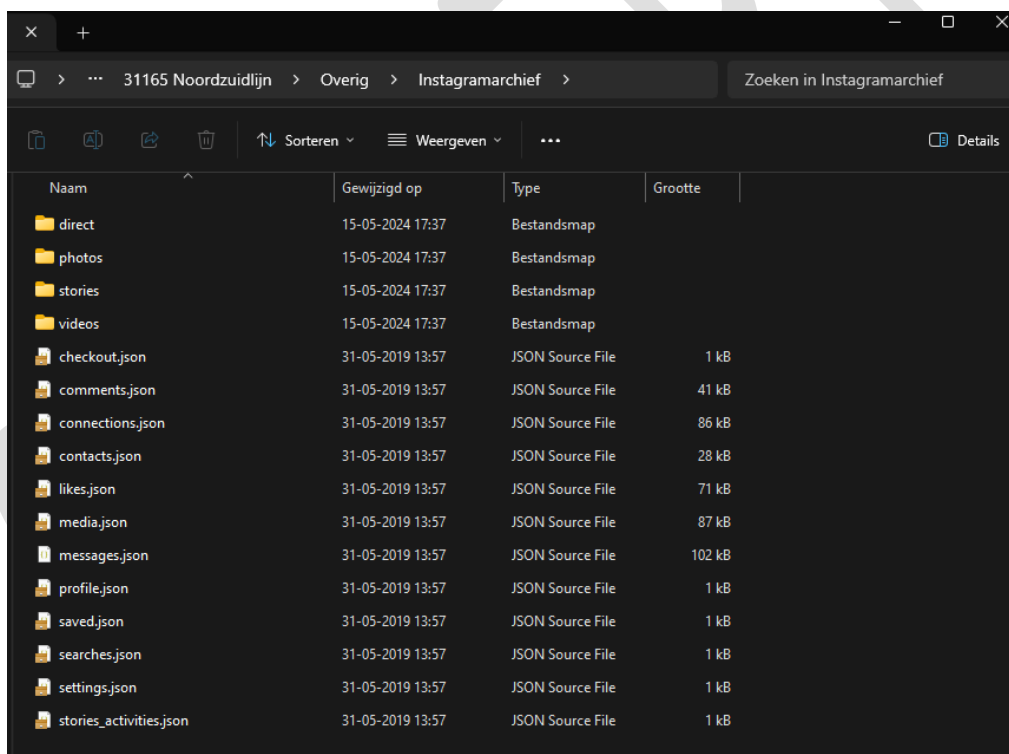
De pilot is uitgevoerd door medewerkers van het Stadarchief Amsterdam, werkzaam op het gebied van web archivering, collectiebeheer, preserving en wetgeving. Tijdens de pilot werd er samengewerkt met archiefvormers om te bepalen wat er opgenomen dient te worden in het archief, wat er niet opgenomen moet worden, en wat technisch überhaupt wel en niet mogelijk is. Er zijn zowel [zakelijke als privé accounts](#) door de archiefvormer overgedragen aan Stadsarchief Amsterdam, die binnengehaald zijn middels de archiveringsfunctie van het desbetreffende platform. Zoals de zakelijke Twitter, Instagram en Facebook accounts van de Noord/Zuidlijn (317 MB) en het privé Twitter archief rondom de Maagdenhuisbezetting (6 MB). Ook heeft het Stadsarchief zelf een aantal zakelijke accounts binnengehaald, bijvoorbeeld over onderwerpen betreffende burgemeester Van der Laan (30 MB). Hier waren geen specifieke selectiecriteria die werden gehanteerd. De meeste accounts draaien om bijzondere gebeurtenissen in Amsterdam en de vergankelijkheid van die informatie. Welke onderdelen binnen het archief wel of niet werden meegenomen was sterk afhankelijk van de techniek.



*Het gearchiveerde Twitter bericht, getoond in Replay Webpage, waarin wordt gedeeld dat Burgemeester Van der Laan is overleden.*

Stadarchief Amsterdam had al enige ervaring met webarchivering met behulp van de open-source tool Webrecorder. Tijdens de pilot bestond de functie om je eigen account te downloaden via een sociale media platform namelijk nog niet. Om deze reden is er gekozen om te kijken naar tools waar al ervaring mee was vanuit webarchivering. Webrecorder biedt de mogelijkheid voor screen capture. Je kan zowel handmatig als geautomatiseerd door een webpagina scrollen en dat wordt dan opgenomen. Naast Webrecorder werd er gebruik gemaakt van TAGS, en Social Feed Manager (eveneens een open-source tools). Tijdens de pilot is niet veel gedocumenteerd over de aanpak. Probleemoplossing was vooral een proces van *trial and error* en zoeken naar andermans ervaringen op het internet. Tijdens de pandemie is er contact geweest met collega-instellingen via online bijeenkomsten waar ervaringen werden uitgewisseld. Daar werd ook Browsertrix aanbevolen, wat nu nog steeds gebruikt wordt voor webarchivering. Er werd vooral gewerkt met tools die relatief simpel te installeren en gebruiken zijn, zonder dat daar een infrastructuur voor hoeft worden opgezet of dat er technische kennis voor nodig is.

Door gebrek aan ervaring met het archiveren van sociale media, werden geen eisen gesteld aan het materiaal dat gearchiveerd moest worden. Er werd bijvoorbeeld niet gedefinieerd welke elementen (zoals video's, foto's en emoji's) van belang waren. Er werd vooral geprobeerd om zoveel mogelijk binnen te halen, om een zo compleet mogelijk account te archiveren. Ook werd er nagedacht over welke platformen beschouwd werden als sociale media, namelijk Twitter, Facebook, LinkedIn, Instagram, YouTube en Whatsapp. Uiteindelijk zijn accounts van Twitter (inmiddels bekend als X), Facebook en Instagram daadwerkelijk gearchiveerd.



*Screenshot van de mappenstructuur van het Instagramarchief van het Noordzuidlijn project.*

## <Uitklapmenu> Uitdagingen en complexiteit

### <Titel> Techniek

Er is met name gebruik gemaakt van open-source tools die ook voor webarchivering worden ingezet, zoals Webrecorder en Browsertrix. Het voordeel hiervan is dat de tools relatief makkelijk zijn in gebruik en er geen uitgebreide technische vaardigheden nodig zijn om de tool werkend te krijgen. Ook is het niet nodig om hier een infrastructuur voor op te zetten.

De uitdaging om web content te archiveren zit vaak in de grootte. Mede doordat er gewerkt werd op eigen laptops, waren de capaciteiten beperkt. De tijd die je nodig hebt om een website of account binnen te halen kan over meerdere dagen verspreid zijn. Het is lastig om direct te zien of dat wat is binnengehaald wel volledig is. Bij sommige platformen, zoals Facebook, was het mogelijk om het als volledige pagina binnen te halen. Via deze techniek werden er ook pop-ups en cookies binnengehaald. Per keer werd bekeken wat er was verzameld en wat er wel en niet werkte.



*Screenshot van het Facebook account van De Nieuwe Universiteit, waar duidelijk te zien is dat het pop-up bericht om in te loggen ook gearcheveerd is.*

Het archiveren van sociale media en de kwaliteitscontrole die daarna handmatig is gedaan, is daarom een arbeidsintensief en tijdrovend proces. Het binnenhalen van video's lukte bijvoorbeeld niet altijd vanaf het platform. Dan werd er gezocht waar het beeldmateriaal nog meer te vinden was. Een YouTube fragment kan bijvoorbeeld als een apart bestand worden opgeslagen, en apart van het sociale media archief geraadpleegd worden.

Om zo volledig mogelijk te kunnen archiveren, werd een account opgedeeld in blokjes. Elke website en elk platform is namelijk anders. Als je bovenaan een tijdlijn begint, bestaat de kans dat er stukjes worden overgeslagen. Daarnaast kun je met de standaardinstellingen van tools redelijk veel binnenhalen, maar elk platform kent eigen limieten waar de tool op vast kan lopen. Door naar een specifiek stukje van de website of het platform te navigeren, verklein je je opgave en heb je een grotere kans van slagen om het succesvol binnen te halen.

## <Titel> Juridisch

Een deel van het sociale media archief is reeds veiliggesteld in het e-depot van het stadsarchief. Omdat sociale media archieven, en andere digital born archieven zoals webarchieven, nog niet raadpleegbaar zijn op de website van Stadsarchief Amsterdam, kunnen deze op aanvraag in de leeszaal beschikbaar worden gesteld.

## <Uitklapmenu> Lessons learned

- Verdiep je in de mogelijkheden die er zijn om sociale media binnen te halen en win advies en ervaringen in van partijen die ervaring hebben. Inmiddels zijn er al heel wat tools uitgetest, waardoor je zelf minder hoeft te testen en te vergelijken. Vooraf inzicht hebben in welke tool te gebruiken helpt je op weg.
- Tijdens de pilot is er voornamelijk op persoonlijke laptops geëxperimenteerd, met weinig capaciteit. Het is vaak niet mogelijk om met open-source tools te werken in de werkomgeving. Kijk of het mogelijk is om hier afspraken over te maken, zodat experimenteren mogelijk is op werk laptop. De ICT-afdeling kan hierbij helpen.
- Benader burgerbewegingen voor acquisitie van privé-accounts. Deze kunnen hun eigen account downloaden en overdragen aan het archief. Voor gebeurtenissen zoals bijvoorbeeld de Maagdenhuisbezetting kan dit veel waarde hebben. Lees [hier](#) meer over het doneren van een persoonlijk Twitterarchief bij Stadsarchief Amsterdam.

Wil je meer lezen over het onderzoek van Stadsarchief Amsterdam, lees dan [hier](#) meer over het onderzoek.

CONCEPT