

<Hoofdpagina> Module archiveren van overheidsinformatie op sociale media

Onder overheidsinformatie verstaan we alle informatie die de overheid zelf maakt of van een ander ontvangt bij het uitvoeren van haar taken. Ongeacht bijvoorbeeld vorm, inhoud, status, of vindplaats. Bij het archiveren van overheidsinformatie op sociale media is het dus ook belangrijk dat deze informatie duurzaam toegankelijk is. Overheidsinformatie is duurzaam toegankelijk wanneer deze vindbaar, beschikbaar, leesbaar, interpreteerbaar, betrouwbaar en toekomstbestendig is. Deze module licht toe hoe je dit kan bereiken voor overheidsinformatie die op sociale media is ontstaan.

<Blok> Advies: wees proactief

Sociale media archivering is een veld dat continu in beweging is. Sociale media platformen veranderen continu en dit geldt tevens voor hun voorwaarden.

Een concreet voorbeeld hiervan is Twitter (X) waar na de overname door Elon Musk in 2022 veel veranderingen kwamen. Zo werd toegang via de API steeds beperkter. Ook werd beperkt hoeveel berichten je per dag kan zien. Dit heeft uiteraard ook effect op het archiveren. Oudere berichten bij grote accounts dreigen daardoor verloren te gaan.

Het is noodzakelijk om tijdig en regelmatig te archiveren, in kleine intervallen, zodat je niet of minder afhankelijk bent van derde partijen. Zorg ervoor dat je op de hoogte bent van de laatste ontwikkelingen. Voor meer informatie, raadpleeg netwerken zoals het [praktijknetwerk sociale media archivering](#) en [KIA](#).

<Blok> Overzicht en beoordeling

Dit onderdeel biedt een overzicht van de technieken die nu bekend zijn op het gebied van duurzame toegankelijkheid. En in hoeverre de technieken tegemoet komen aan wat vereist wordt op het gebied van duurzame toegankelijkheid.

→ Bekijk het [overzicht en de beoordeling van archiveringsmethodes](#)

<Blok> Toelichting structuur module

<Uitklapmenu> Aanpak

Het streefdoel is [archivering by design](#). Maar door de afhankelijkheid van sociale media platformen kunnen overheidsorganisaties dit principe amper toepassen voor het registreren van overheidsinformatie op sociale media. Het is immers niet mogelijk om ontwerp- en beheerkeuzes te maken in de platformen zelf. Het is ook niet mogelijk om bepaalde functionaliteiten af te

dwingen met een contract of andere middelen. Om deze reden is het goed om de [aanbevelingen bij gebruik](#) te volgen en de relevante [DUTO randvoorwaarden](#) op organisatie- of procesniveau.

Vanwege deze afhankelijkheden is er gekozen voor een aanpak in lijn met het [DUTO-raamwerk](#), in combinatie met hoe de technieken voldoen aan de [essentiële kenmerken](#).

→ Lees hier meer over het [DUTO-raamwerk en essentiële kenmerken](#)

<Uitklapmenu> Registreren – technieken

De eerste stap in de aanpak is de overheidsinformatie van sociale-media-platformen halen om op te kunnen nemen in de eigen beheeromgeving. Dit valt onder het DUTO-proces [registreren](#). We bekijken de technieken die dit mogelijk maken. Hoe werken deze, hoe zit het met de uitvoerbaarheid en welke types informatieobjecten levert het op? Hierbij wordt rekening gehouden met de hieruit volgende DUTO-processen en de essentiële kenmerken.

→ Lees hier meer over [technieken](#)

<Uitklapmenu> Bewaren – informatieobjecten

De registratie levert een nieuw informatieobject op dat [overheidsinformatie op sociale media](#) zou moeten omvatten. De volgende stap is dat informatieobject zodanig [bewaren](#) dat er gedurende de gehele levensduur geen onaanvaardbaar informatieverlies optreedt, de authenticiteit wordt gewaarborgd en de leesbaarheid is gegarandeerd. Het is dus belangrijk dat de essentiële kenmerken intact blijven. Deze zijn te verdelen in de volgende vijf categorieën: inhoud, context, vormelementen, gedrag, en structuur.

→ Lees hier meer over [informatieobjecten](#)

<Uitklapmenu> Ter beschikking stellen – beheer en weergave

De derde stap is het [ter beschikking stellen](#) van het informatieobject. Hoe kan je de informatieobjecten goed beheeren en weergeven voor gebruikers? Zijn er specifieke applicaties nodig of andere afhankelijkheden aanwezig die beschikbaarstelling bemoeilijken? Ook daarmee moet al op het moment van registreren rekening mee gehouden.

→ Lees hier meer over [beheer en weergave](#)

<Uitklapmenu> Vernietigen

Een laatste mogelijke stap is de vernietiging van overheidsinformatie op sociale media als dat zo is [vastgelegd in een selectielijst](#). Voor het [vernietigen](#) van deze informatie bestaat (nog) geen reguliere inrichting, omdat er grote uitdagingen zijn. Vernietigen geldt namelijk op twee manieren, namelijk het verwijderen van het live web en het geregistreerde informatieobject in het eigen informatiesysteem. In het eerste geval moet dit ofwel handmatig ofwel geautomatiseerd per bericht gedaan worden, of door middel van het stopzetten van het desbetreffende account. Ook is totale vernietiging op het platform niet mogelijk, omdat het niet zeker is of de informatie ook

echt weg is bij het platform. Het DUTO-proces kan binnen sociale media platformen dus niet worden uitgevoerd.

In het tweede geval kunnen organisaties alvast het behulp van het DUTO-raamwerk kijken naar het passende [niveau van maatregelen](#). Voor het vernietigen van het archiefexemplaar in de eigen beheeromgeving van de overheidsorganisatie kan de [handreiking digitaal vernietigen](#) geraadpleegd worden.

CONCEPT

<Subpagina> Overzicht en beoordeling van archiveringsmethodes

<Kop> Overzicht

Hieronder is een compleet overzicht te zien van alle technieken gerelateerd aan de uitvoerbaarheid, essentiële kenmerken van het informatieobject, en de uiteindelijke beheer en weergave. De volgende onderdelen worden apart uitgelicht:

- [Uitvoerbaarheid van techniek](#)
- [Essentiële kenmerken van het informatieobject](#)
- [Uitvoerbaarheid van beheer en weergeven](#)

Legenda voor beoordeling:

Symbol	Betekenis
++	Voldoet ruim
+	Voldoet
-	Voldoet niet volledig
--	Voldoet niet

Technieken	Uitvoerbaarheid van techniek	Essentiële kenmerken van het informatieobject*		Uitvoerbaarheid van beheer	Uitvoerbaarheid van weergeven
		Inhoud	Context		
Application Programming Interface (API)	--	Inhoud	-	++	++
		Context	+		
		Vormelementen	--		
		Gedrag	--		
		Structuur	-		
Archiverings-functie platform	++	Inhoud	+	--	++
		Context	+		
		Vormelementen	++		
		Gedrag	++		
		Structuur	-		
Screen capturing	-	Inhoud	-	-	++
		Context	+		
		Vormelementen	++		
		Gedrag	--		
		Structuur	--		
Web scraping/ harvesting/ crawling	+	Inhoud	++	++	-
		Context	++		
		Vormelementen	++		
		Gedrag	++		
		Structuur	++		

* Over het algemeen genomen, kijk voor meer uitgebreide informatie bij [informatieobjecten](#).

<Kop> Advies: combineer de technieken

Duidelijk is dat geen van de technieken alles registreert wat je zou willen. Voor duurzaam toegankelijke overheidsinformatie op sociale media is het daarom noodzakelijk om verschillende technieken toe te passen. Met als beste combinatie de archiveringsfunctie platform en web scraping/harvesting/crawling. De archiveringsfunctie van het platform is namelijk het makkelijkste uit te voeren qua techniek. Tevens heb je dan in elk geval de eigen informatie. Web scraping/harvesting/crawling aan de andere kant scoort juist weer het beste op de essentiële kenmerken en behoudt ook de look-and-feel van de sociale media platformen. Door met beide technieken te werken kan je overheidsinformatie op het sociale media platform het beste duurzaam toegankelijk maken en houden.

CONCEPT

<Subpagina> DUTO-raamwerk en essentiële kenmerken

< Kop> DUTO-raamwerk

Het DUTO-raamwerk is een praktisch hulpmiddel om de passende maatregelen te bepalen waarmee je overheidsinformatie duurzaam toegankelijk kunt maken. Er zijn verschillende processen en functies gedefinieerd binnen het raamwerk. Degenen die expliciet van toepassing zijn op sociale media archivering worden in deze module [benoemd](#):

- [Registreren](#)
- [Bewaren](#)
- [Ter beschikking stellen](#)
- [Vernietigen](#)

< Kop> Essentiële kenmerken

Bij het onderdeel [Technieken](#) is te zien dat elke techniek een andere output levert qua bestandsformaten. Dit heeft vervolgens ook effect op wat zich in het informatieobject bevindt. Niet elk bestandsformaat is namelijk geschikt om elk type informatie zonder informatieverlies op te nemen. Zo zijn de formaten PDF/A en CSV niet geschikt om dynamische informatie zoals video's weer te kunnen geven.

Het is daarom van belang dat er voor een techniek wordt gekozen die de essentiële kenmerken intact houdt. Dit zijn de kenmerken die nodig zijn om het informatieobject goed te kunnen duiden in termen van authenticiteit, context, en functionaliteit. Het NA heeft geen uitputtend onderzoek uitgevoerd naar de essentiële kenmerken van sociale media. Wel is uitgegaan van het internationaal erkende [Joint Information Systems Committee](#) (JISC), die essentiële kenmerken definiëren als: '...kenmerken van een informatieobject die haar uiterlijk, gedrag (functionaliteit), kwaliteit en bruikbaarheid bepalen.'

Essentiële kenmerken zijn onder te brengen in vijf categorieën:

- Inhoud (*content*)
 - Het gaat hierbij om wat inhoud zoals de teksten of een emoji uitdrukken, niet om de vorm. Dus niet enkel dat je een duimpje omhoog ziet, maar dat duidelijk is wat dat duimpje betekent.
- Context (*context*)
 - Metagegevens, zoals auteur en datum, maken duidelijk door wie en wanneer iets geplaatst is. Het geeft dus context aan een bericht.
- Vormelementen (*appearance*)
 - Denk hierbij aan kleur en lay-out, waaraan je bijvoorbeeld kan zien op welk platform iets geplaatst is. Bijvoorbeeld de specifieke kleur blauw van Facebook, of de vogel/X van Twitter/X.

- Gedrag (*behaviour*)
 - Denk hierbij aan interactie en functionaliteit. Bijvoorbeeld een formulier dat je in kan vullen of een video die je kan afspelen.
- Structuur (*structure*)
 - Verschillende typen onderdelen in een bericht kunnen een relatie hebben, zoals berichten van een account die zijn gedeeld op een ander account (reposts).

Door essentiële kenmerken te koppelen aan de DUTO-proces Bewaren en Registreren, kun je concreter afwegen waaraan het informatieobject moet voldoen. En dus welke techniek hier het beste op aansluit.

< Kop> Voorbeeld essentiële kenmerken

Om de toepassing van de essentiële kenmerken te verduidelijken, wordt dit hier aan de hand van een concreet voorbeeld toegelicht. Hieronder is een bericht van het Nationaal Archief te zien. Dit bericht is op Twitter (X) geplaatst op 27 september 2023. Het bericht bevat tekst, emoji's, een externe link en een afbeelding.



De rode lijn en blauwe lijn staan om alles heen. Het gaat hier om het volledige bericht. Rood staat voor **inhoud**. Hier gaat het om datgene dat de inhoud overbrengt, dus de tekst en afbeelding in dit geval. De blauwe lijn staat voor de **vormelementen** van de hele post, waar het gaat om wat

het uitdrukt en dat dat goed weergegeven en gelezen wordt door de computer en gezien kan worden door mensen.

De groene lijn toont de elementen die **context** bieden, zoals wie het bericht geplaatst heeft (de gebruiker), wanneer (datum) en het bereik van de post van 1.5K (weergaven). Let wel, het gaat ook om metadata die je niet in de post zelf kan zien, zoals geolocatie. Ook hashtags kunnen context bieden, maar zijn geen onderdeel van dit voorbeeld.

De gele lijn staat voor **gedrag**. Het gaat hier niet om wat je ziet, maar of datgene wat je ziet zich ook gedraagt zoals wordt verwacht. In dit voorbeeld zien we een afbeelding, maar als hier een video zou staan, zou het als video moeten afspelen. Ook de link valt onder gedrag, deze verwijst namelijk ergens naar als je erop klikt. Een ander voorbeeld is een 3D-afbeelding, die je meedraait op de manier hoe je de telefoon beweegt.

De oranje lijn staat voor **structuur**. Het gaat om de relatie tussen bepaalde elementen. In dit geval is de afbeelding onderdeel van de post. Dat noemen we *embedded content*, vandaar dat deze oranje lijn rond de afbeelding in de post staat. Structuur is bij sociale media ook van toepassing bij reacties. En dan met name hoe reacties op reacties bij elkaar horen. Verlies je die structuur, dan weet je niet meer welke berichten waarop reageerden.

<Subpagina> Registreren - Technieken

Dit onderdeel licht de meest gebruikte technieken voor het registreren van informatieobjecten uit sociale media-platformen toe. Dat wordt in alfabetische volgorde gedaan. Daarbij is niet alleen aandacht voor de techniek zelf, maar ook hoe uitvoerbaar deze is en in welke vorm het zo gegenereerde informatieobject wordt opgeslagen.

< Uitklapmenu> Application Programming Interface (API)

<Titel> Omschrijving

De afkorting API staat voor Application Programming Interface. Een sociale media API is een stukje code dat het mogelijk maakt om sociale media platformen te integreren met apps en tools van anderen. In zekere zin is een API een communicatiemiddel. Verschillende programma's gebruiken een API om met elkaar te 'praten' en data te delen.

De API kan ook gebruikt worden om overheidsinformatie van sociale media platformen te halen. Door te verbinden met de API van een sociaal media platform, is het tevens mogelijk om (meta)gegevens te verzamelen die eindgebruikers niet standaard te zien krijgen. Denk hierbij bijvoorbeeld aan:

- Geaggregeerde aantallen actieknoppen (likes) en reposts
- Metagegevens over de locatie
- Unieke identifiers voor elke post

<Titel> Uitvoerbaarheid

- Sociale media platformen hebben de afgelopen jaren steeds meer de toegang tot de API gelimiteerd. Om toegang te krijgen moet je vaak betalen en een account hebben op het specifieke platform, maar zelfs dan kan er sprake zijn van restricties.
- Sociale media platformen geven geen inzicht in welke data wel en niet worden gedeeld via de API. Dat betekent dat de uitwisseling van data niet transparant is en de volledigheid, authenticiteit en integriteit niet verzekerd kan worden.
- Sociale media platformen veranderen regelmatig de API waardoor telkens een nieuwe tijdsinvestering nodig is om weer aan te sluiten op de meest recente versie van de API.
- De toegang tot en het gebruik van de API vraagt technische kennis.

<Titel> Output

Deze techniek genereert gestructureerde, tekstuele data. Deze data is eenvoudig te analyseren en te verwerken met behulp van een computer. De gestructureerde data wordt vaak opgeslagen in bestandsformaten zoals JSON, XML, CSV of XLS(X).

< Uitklapmenu> Archiveringsfunctie platform

<Titel> Omschrijving

Sociale media platformen bieden gebruikers de mogelijkheid om de eigen informatie van het eigen sociale media account te exporteren en zo te archiveren. De wijze van ter beschikking stellen kan verschillen per platform. Eerst krijgt de gebruiker de mogelijkheid om ofwel alle soorten gegevens te downloaden, ofwel een selectie. Een afbakening in de tijd is daarentegen niet mogelijk. Na een keuze te hebben gemaakt dient de gebruiker een verzoek in bij het platform. De gebruiker kan vervolgens het 'archief' downloaden.

Bij het gebruik van de archiveringsfunctie komt informatie van andere gebruikers, zoals opmerkingen en gedeelde berichten, niet mee. Of deze wordt geanonimiseerd. Ook (het aantal) reacties (via actieknoppen) wordt bij diverse platformen niet mee geëxporteerd en gearchiveerd

<Titel> Uitvoerbaarheid

- Het gebruik van deze archiveringsfunctie vraagt geen grote tijdsinvestering. Het verwerken van de output op correcte wijze kost echter wel tijd. Sommige preserveringssystemen kunnen de gedownloade mappenstructuur niet in zijn geheel opnemen, dus daar moet een oplossing op maat voor gemaakt worden. Zie ook het [praktijkvoorbeeld RAA](#).
- Het gebruik van deze archiveringsfunctie vraagt geen extra software of technische kennis.

<Titel> Output

Deze techniek genereert een ZIP-bestand. Dit ZIP-bestand bevat doorgaans een HTML-bestand dat dient als index. De gebruiker kan met behulp van deze HTML de informatie bekijken in een browser. Daarnaast kan het ZIP-bestand JavaScript-bestanden, afbeeldingen (zoals JPG), video's (zoals MP4), en zelfs gestructureerde data (zoals CSV) bevatten.

< Uitklapmenu> Screen capturing

<Titel> Omschrijving

Screen capturing is het opnemen van een scherm. Dit kan een statisch beeld zijn, zoals een screenshot, of een dynamische opname van activiteit op het scherm. Alles wat op het scherm te zien was tijdens het maken van de screen capture, wordt gearchiveerd. Dus als comments zijn geopend of bijbehorende (account)namen te zien zijn, zullen deze ook in het archief komen.

Hoewel het uiterlijk van het platform deels behouden blijft, is het een beperkte weergave van het origineel. De interactie, zoals het openen van links of het afspelen van video's, is met deze techniek na registratie niet meer mogelijk.

<Titel> Uitvoerbaarheid

- Screen capturing is een tijdrovende techniek omdat deze grotendeels handmatig uitgevoerd moet worden. De gebruiker moet zelf alle functionaliteiten en inhoud (handmatig) selecteren. Met programmeerkennis zou dit (deels) opgelost kunnen worden.
- Output in de vorm van een video is foutgevoelig. Bij te snel scrollen tijdens het registreren kan het beeld onscherp worden, waardoor het niet meer leesbaar is.
- Screen capturing vraagt weinig technische kennis. Het is een laagdrempelige manier om informatie vast te leggen.

<Titel> Output

Deze techniek genereert een afbeelding (bijvoorbeeld een PNG- of JPEG-bestand), video (bijvoorbeeld MP4) of een PDF.

< Uitklapmenu> Web scraping/harvesting/crawling

<Titel> Omschrijving

Bij webarchivering worden verschillende termen - web scraping, harvesting of crawling – door elkaar gebruikt voor het systematisch extraheren van delen van het web. Tijdens dit geautomatiseerde proces wordt een software agent, een web robot, of een script ingezet om de interactie tussen webserver en de mens na te bootsen en zo gegevens binnen te halen. De gearchiveerde webpagina's worden getoond in een browser en maken doorgaans gebruik van (X)HTML.

Ook sociale media kun je met deze techniek archiveren. Voor deze techniek geldt dat zowel de eigen informatie als informatie van andere gebruikers, zoals comments en likes, worden meegenomen.

<Titel> Uitvoerbaarheid

- Deze techniek kun je zowel handmatig als geautomatiseerd toepassen.
- Er zijn veel tools beschikbaar. Zowel commerciële als opensourcetools, die door een actieve community van gebruikers worden onderhouden.
- Sommige tools vergen meer technische kennis, omdat programmeerkennis vereist is. Andere tools bieden een eenvoudig te gebruiken plug-in voor de browser. Zorg ervoor dat je een keuze maakt die past bij je eigen kennisniveau.

<Titel> Output

Deze techniek genereert doorgaans output in de vorm van een [WARC](#). Dit is een open bestandsformaat voor het archiveren van websites.

<Subpagina> Bewaren - Informatieobjecten

Dit onderdeel focust op de essentiële kenmerken van informatieobjecten op sociale media. Deze kenmerken zijn belangrijk voor de keuze voor een bepaalde techniek, omdat de gekozen techniek effect heeft op het informatieobject dat je registreert. Kenmerken behoren altijd tot een van deze vijf categorieën: inhoud, context, vormelementen, gedrag, en structuur.

In de volgende onderdelen gaan we op elk van deze categorieën in. Met de volgende legenda voor beoordeling:

Symbool	Betekenis
+	Voldoet aan categorie
-	Voldoet niet aan categorie
?	Is afhankelijk van het sociale media platform

< Uitklapmenu> Inhoud

<Titel> Omschrijving

Dit is de inhoud van het informatieobject. Denk hierbij bijvoorbeeld aan tekst, emoji's, afbeeldingen, en video's die de primaire boodschap of informatie overbrengen. Bij een video gaat het als zodanig niet om hoe deze eruit ziet, maar om wat deze uitdrukt.

<Titel> Inhoud en technieken

De onderstaande tabel verduidelijkt voor elke techniek of de inhoud wel of niet wordt binnengehaald.

Techniek	Inhoud
API	+ tekst + reacties + actiekноп(pen)* - afbeeldingen - video - emoji's
Archiveringsfunctie platform	+ tekst + afbeeldingen + video's + emoji's + actiekноп(pen) - reacties
Screen capturing	+ afbeeldingen

Web
scraping/harvesting/crawling****

- + emoji's
- + actiekноп(pen)
- tekst**
- video's***
- reacties**

- + tekst
- + afbeeldingen
- + video's
- + emoji's
- + reacties
- + actiekноп(pen)

* Met actiekноп(pen) wordt bedoeld op likes en reposts.

** Dit kan incompleet zijn omdat deze handmatig uitgekapt moeten worden. Wees hier alert op tijdens het toepassen van deze techniek.

*** Dit is veelal een statische screenshot, behalve als de video actief is gestart en meegenomen tijdens het toepassen van deze techniek.

**** De volledigheid is afhankelijk van de instellingen van de tool die je gebruikt en de limieten van het platform.

< Uitklapmenu > Context

<Titel> Omschrijving

Context is de informatie die nodig is om te beschrijven hoe en wanneer de content is gemaakt en wat de intentie van die content was.

De context van een informatieobject wordt vastgelegd in metagegevens. Denk bijvoorbeeld aan auteur, hashtags, geolocatie, datum en tijd van een post. Het aantal volgers en aantal weergaven zegt eveneens iets over de context van een account of post.

Ook hoe en wanneer content is binnengehaald, kan worden vastgelegd in metagegevens. Dit wordt ook wel *crawl* of *capture metadata* genoemd.

<Titel> Context en technieken

De onderstaande tabel verduidelijkt voor elke techniek welke metagegevens worden aangeleverd.

Techniek

Context

API*

- + auteur
- + hashtags
- + datum en tijd
- + geolocatie

Archiveringsfunctie platform	+ aantal volgers + weergaven - <i>crawl/capture metadata</i>
Screen capturing	+ auteur + hashtags + datum en tijd + aantal volgers + weergaven ? geolocatie - <i>crawl/capture metadata</i>
Web scraping/harvesting/crawling	+ auteur + hashtags + geolocatie + datum en tijd + weergaven + <i>crawl/capture metadata**</i>

* Is afhankelijk van de mogelijkheden die het platform biedt en wat je zelf aanvraagt.

** Is afhankelijk van de instellingen van de tools.

< Uitklapmenu > Vormelementen

<Titel> Omschrijving

Dit is de informatie die nodig is om het uiterlijk van het informatieobject correct weer te geven. Zijn alle vormelementen van het informatieobject binnengehaald? Dat omvat zowel de inline lay-out als de externe lay-out. Het is hierbij een wisselwerking tussen de computer en de mens. De computer moet kunnen begrijpen dat het bijvoorbeeld om een afbeelding of video gaat om dit correct weer te kunnen geven. Vervolgens moet de mens dit ook kunnen zien en kunnen identificeren als een afbeelding of video.

Inline lay-out is de uitstraling van de inhoud van de pagina. Het is onderdeel van de pagina/broncode. Voorbeelden hiervan zijn:

- Stijlelementen
- Lijsten
- Tabellen
- Formulieren
- Scalable vector graphics (SVG)

- Client-side scriptinstructies, zoals pull-down menu's, navigatiehulpmiddelen.

Externe lay-out: dit omvat externe bestanden die de lay-out en designelementen van de pagina bepalen. Hoewel deze bestanden zich niet in de pagina bevinden, moeten ze wel worden meegenomen om het uiterlijk en lay-out te behouden. Voorbeelden hiervan zijn gelinkte objecten en bestanden die geïntegreerd zijn in de lay-out, zoals:

- Afbeeldingen
- Video
- Audio
- Server scripts
- Cascading Style Sheets (CSS).

<Titel> Vormelementen en technieken

De onderstaande tabel verduidelijkt voor elke techniek of de inline en externe lay-out wordt meegenomen.

Techniek	Vormelementen
API*	- inline lay-out - externe lay-out
Archiveringsfunctie platform**	+ inline lay-out + externe lay-out
Screen capturing***	+ inline lay-out + externe lay-out
Web scraping/harvesting/crawling****	+ inline lay-out + externe lay-out

* Enkel de URLs naar *embedded* content, zoals video's worden binnen gehaald. Daardoor is het niet duidelijk dat het om een video gaat.

** Het gedownloadte indexbestand van diverse platformen, zoals bijvoorbeeld Facebook en Twitter (X), geeft geen authentieke weergave van het live platform.

*** De output in een PDF en afbeelding (bijvoorbeeld PNG) is een statische weergave van het platform. Het uiterlijk van het platform en de diverse vormelementen komt mee, dit geldt ook bij de output in de vorm van een video waarbij een afspeel-symbool te zien is.

**** Dit is afhankelijk van de instellingen van de tools die wordt gebruikt.

< Uitklapmenu> Gedrag

<Titel> Omschrijving

Dit verduidelijkt hoe content op sociale media interacteert met andere content zoals bijvoorbeeld hyperlinks. Bij de vorige categorie, vormelementen, werden ook formulieren genoemd. Waar het bij vormelementen gaat om hoe een formulier eruit ziet, gaat het bij de categorie gedrag erom dat deze ook nog steeds functioneert als een formulier. Ook gegevensvalidatie is een voorbeeld van gedrag. Dus dat het bijvoorbeeld niet mogelijk is om tekst in te vullen in een datumveld.

<Titel> Gedrag en technieken

De onderstaande tabel verduidelijkt voor elke techniek of het gedrag van het platform wordt meegenomen.

Techniek	Gedrag
API	-
Archiveringsfunctie platform	+
Screen capturing*	-
Web scraping/harvesting/crawling	+

* Dit is afhankelijk van de output. Bij output in de bestandsformaten PDF en PNG/JPG komt het gedrag niet mee omdat dit een statische weergave is. Bij output in de vorm van een video (bijvoorbeeld MP4) kan gedrag wel meekomen indien deze is geactiveerd ten tijde van het toepassen van de techniek. Echter geeft dit nog steeds enkel een weergave van gedrag en interactie, maar niet het daadwerkelijke gedrag en interactie op het sociale media platform (je kan zelf nergens op klikken in de gearchiveerde versie).

< Uitklapmenu> Structuur

<Titel> Omschrijving

De categorie structuur doelt op de informatie die vereist is om de relatie tussen twee of meer typen content te kunnen duiden. Denk hierbij bijvoorbeeld aan de relatie tussen een origineel bericht en een reactie op dit bericht (tweet en retweet) zowel als de relatie met *embedded content*.

<Titel> Structuur en technieken

De onderstaande tabel verduidelijkt voor elke techniek hoe deze de structuur en de onderlinge relaties tussen verschillende typen content meeneemt in het informatieobject.

Techniek	Structuur
API	-
Archiveringsfunctie platform*	-
Screen capturing**	-
Web scraping/harvesting/crawling	+

* De structuur (onderlinge relaties tussen verschillende types content) is onvolledig hier omdat informatie van derden bij deze techniek ontbreekt. Je krijgt enkel de eigen data.

** Anders dan een visuele reconstructie (vormelement) is de structuur niet meer te reconstrueren uit de data bij deze techniek.

<Subpagina> Ter beschikking stelling - Beheer en weergave

In het onderdeel [Technieken](#) is aangetoond dat de methode die je gebruikt, van invloed is op het informatieobject dat wordt gecreëerd en in welk bestandsformaat dat terecht komt. Dat geldt ook voor de volgende stap, het beheren en weergeven van het informatieobject.

In onderstaande tabel wordt voornamelijk gekeken naar de uitvoerbaarheid van deze twee aspecten. Het gaat bij weergave er dus niet om of het authentiek weer wordt gegeven, maar hoeveel vervolgstappen er nodig zijn om het te kunnen weergeven. Dit is uiteraard over het algemeen genomen, elk informatiesysteem werkt anders. Denk vooral ook na over de eigen context in relatie tot beheer en weergave.

Met de volgende legenda voor beoordeling:

Symbol	Betekenis
+	Vergt weinig vervolgstappen
-	Vergt meer vervolgstappen

	Beheer	Weergave
Application Programming Interface (API)	+	+
Archiveringsfunctie platform	-	+
Screen capturing	-	+
Web scraping/harvesting/crawling	+	-

< Uitklapmenu> Toelichting

<Titel> Application Programming Interface (API)

Beheer

De output van de API vraagt weinig extra aandacht. De gestructureerde data kan eenvoudig in beheer genomen worden.

Weergave

Een JSON-bestand kun je in principe openen en bekijken met elk programma dat *plain text* kan lezen, zoals kladblok, Notepad of zelfs Word. Een CSV bestand - eigenlijk gestructureerde data in tabelvorm – kun je in principe openen met elke spreadsheetprogramma, zoals Excel.

<Titel> Archiveringsfunctie platform

Beheer

Specifiek in het geval van Twitter (X) is bekend dat de links en afbeeldingen in het indexbestand verwijzen naar het live sociale media platform. Wanneer het account van het live web is verwijderd, of het platform of bepaalde content offline zou zijn gehaald, werkt dat dus niet meer.

Ook is er sprake van verkorte links (t.co). Bovendien zijn de afbeeldingen niet de oorspronkelijke, maar veel kleiner dan wat origineel online is gezet.

Het is mogelijk om een verwerkingslag over je Twitter archief te doen, om veel van bovenstaande zaken op te lossen. Voor het probleem van verwijzingen naar het live web, de verkorte links en de kleine afbeeldingen zijn oplossingen online te vinden, zoals een Twitter parser. Dit is echter wel een extra bewerkingslag die je zelf moet uitvoeren.

Weergave

Het indexbestand kan je openen en bekijken in een webbrowser. Dit toont de structuur van het platform, zoals de tijdlijn. Ook afbeeldingen, emoji's en video's worden weergegeven in de index, maar zijn dus tevens terug te vinden in afzonderlijke bestanden.

<Titel> Screen capturing

Beheer

De output van de screen capture vraagt weinig extra aandacht om op te nemen in je informatiesysteem. Echter, context en metagegevens (denk hierbij aan geolocatie) raken vrijwel volledig verloren bij screen capturing. Toevoegen hiervan vraagt veel aandacht.

Weergave

De output zoals video's (bijvoorbeeld MP4), afbeeldingen (bijvoorbeeld PNG), of een PDF kan je openen en bekijken zonder bijzondere applicaties.

<Titel> Web scraping/harvesting/crawling

Beheer

De output van het web scraping/harvesting/crawling vraagt weinig extra aandacht. Een WARC-bestand kun je valideren met behulp van de [handreiking WARC-validatie](#).

Weergave

Een volledige WARC kan je alleen openen en bekijken met een speciale reader, zoals replayweb.page, Webrecorder Player of OpenWayBack. De meeste preserveringsystemen hebben een dergelijke functie ingebouwd. Het is mogelijk om een WARC te openen in een tekst editor. Je kunt dan de metadata en HTML bekijken. Afbeeldingen en video's zijn echter als binaire data opgeslagen en geven dus geen authentieke weergave.