

Collectiebeschrijving

Webcollectie maatschappijkritische websites

Auteur: Laurie Hasselt

Redactie: Peter de Bode en Kees Teszelszky

Versie: definitief

Datum: 03-01-2018

Redactedatum: 25-07-2018

Inhoud

0. Inleiding.....	3
1. Webarchivering: het hoe en wat.....	5
1.1 Waarde als erfgoed: weg is weg.....	5
1.2 Webarchiveringslandschap in Nederland.....	6
1.3 Webarchivering bij de KB	7
1.4 Hoe werkt webarchivering	8
1.5 Harvestingstrategie.....	9
2. Maatschappijkritische webcollectie	10
2.0 Inleiding.....	10
2.1 Beknopte omschrijving	11
2.2 Uitgebreide omschrijving.....	13
2.2.1 Aanleiding collectie.....	13
2.2.2 Selectiecriteria	14
2.2.3 Totstandkoming.....	15
2.2.4 Reacties	15
3. Wat niet wordt gearhiveerd.....	17
3.1 Algemeen	17
3.2 Juridische beperkingen	17
3.3 Technische beperkingen	18
3.4 Overzicht niet-gearhiveerde websites.....	18
4. Collectie	20
4.1 Algemeen	20
4.3 Websfeer	26
4.4 Toegang.....	38
5. Conclusies	39
6. Aanbevelingen.....	40

0. Inleiding

'I don't know, what do I know about it? All I know is what's on the internet.'¹ Aldus sprak Donald Trump, destijds presidentskandidaat voor de Verenigde Staten, een halfjaar voor hij tot president verkozen werd. Maar wat er op het internet te vinden is, verschilt per seconde. Webpagina's verschijnen en verdwijnen in ras tempo. De gemiddelde webpagina is slechts negentig dagen online. En aangezien "de geschiedenis van vandaag [...] vooral online [wordt] geschreven"², betekent dat dat we het risico lopen dat er grote gaten in ons collectief geheugen ontstaan.

Personen en instituties die zich bezighouden met webarchivering, waarvan de Koninklijke Bibliotheek (KB) er één is, weten dat natuurlijk al veel langer. Dat is dan ook precies de reden dat de KB in 2007 gestart is met het aanleggen van een webarchief, waarin websites over de Nederlandse taal, cultuur en geschiedenis worden gearchiveerd. Anno december 2017 maken er meer dan 13.000 Nederlandse websites deel uit van dit webarchief.³

In de tien jaar dat het KB-webarchief inmiddels bestaat, is er tweemaal grootschalig onderzoek gedaan naar het webarchiveringsproces aldaar. Van 2012-2016 werd in het project WebART, een samenwerkingsproject van de KB, de Universiteit van Amsterdam (UvA) en het Centrum Wiskunde en Informatica (CWI), onderzocht hoe het webarchief bruikbaar gemaakt kan worden voor wetenschappers.⁴ In 2016-2017 heeft Kees Teszelszky onderzoek gedaan naar verschillende aspecten van webarchivering bij de KB, te weten: selectie van, opslag van en toegang tot de gearchiveerde websites.⁵

Uit deze onderzoeken zijn verschillende aanbevelingen ten behoeve van het optimaliseren van het KB-webarchief gekomen. Zo leverden wetenschappers in WebART het commentaar dat het "webarchief van de KB wel een beetje braaf is".⁶ In het rapport dat uit zijn onderzoek is voortgevloeid besteedt ook Kees Teszelszky aandacht aan de aard van de inhoud van de websites in het archief. Hij concludeert onder meer dat de collectie "inhoudelijk aan waarde [zou] winnen als meer sites worden gearchiveerd met inhoud die politiek, religieus of anderszins abject of omstreden is en daarom juridisch bezien op het randje, maar die wel van groot historisch belang in de toekomst kunnen worden."⁷

Twee andere belangrijke aanbevelingen die uit onderzoek van laatstgenoemde voortvloeiden waren dat gebruikers beter geïnformeerd moeten worden over het webarchief – hoe de inhoud tot stand is gekomen, wat het bevat en wat hiervan de waarde is als erfgoed en bron voor de wetenschap – en dat het webarchief van de KB in context geplaatst moet worden, zoals dat al wel

¹ https://www.youtube.com/watch?v=l9IEIP3k_8c

² <https://www.kb.nl/organisatie/onderzoek-expertise/e-depot-duurzame-opslag/webarchivering>

³ Peildatum mei 2017: <https://www.kb.nl/organisatie/onderzoek-expertise/e-depot-duurzame-opslag/webarchivering>

⁴ Zie voor meer informatie: <http://www.webarchiving.nl/>

⁵ Barbara Sierman en Kees Teszelszky, 'How can we improve our web collection? An evaluation of webarchiving at the KB National Library of the Netherlands (2007-2017)': <http://journals.sagepub.com/doi/10.1177/0955749017725930>

⁶ <https://www.kb.nl/organisatie/onderzoek-expertise/onderzoek-digitalisering-en-digitale-duurzaamheid/webart-het-kb-webarchief-bruikbaar-maken-voor-de-wetenschap>

⁷ Kees Teszelszky, 'Onderzoek webarchivering. Deel I: selectie van de websites': p. 66.

gebeurd met andere bijzondere delen van de collectie.⁸ Dit is een belangrijke voorwaarde om een collectie op te slaan in een e-depot.

De collectie die ten grondslag ligt aan dit verslag, evenals de collectiebeschrijving zelf, vloeien direct voort uit de aanbevelingen van bovengenoemde twee kritische beschouwingen van het KB-webarchief. Voor het eerst in de tien jaar dat de KB zich bezighoudt met webarchivering is er ruimte gekomen voor een ingrijpende inhoudelijke verruiming van de selectie. Er is een collectie samengesteld van circa vijfhonderd websites die politiek of maatschappelijk omstreden of ‘op het randje’ zijn of maatschappijkritische inhoud bevatten, en die een bron zijn van huidige maatschappelijke tendensen en in die hoedanigheid als belangrijke bron van de huidige tijdgeest kunnen dienen voor toekomstige historici.

Maar, om terug te keren bij Donald Trump, als het enige wat we weten datgene is wat zich op het internet bevindt, is met het archiveren van deze websites de kous nog niet af. Zonder contextinformatie zal de toekomstige onderzoeker of andere archiefgebruiker immers met alleen een link naar de gearchiveerde site nog steeds niet afdoende weten ‘what’s on the internet’. Om de gebruiker van deze webcollectie een helpende hand te bieden bij het begrijpen wat er anno 2017 zoal op het wereldwijde web te vinden is, dient deze collectiebeschrijving, waarin de benodigde informatie voor het in context plaatsen van deze bronnen wordt geboden.⁹

Deze collectiebeschrijving biedt een eerste, doch geen definitief antwoord op de geconstateerde lacunes in het webarchiveringsproces bij de KB. Niettemin wil ik graag een poging doen de inspanningen die verricht zijn in dit project van duurzaam nut voor de KB te maken. Om die reden heb ik getracht een model te ontwikkelen voor collectiebeschrijvingen dat ook in de toekomst gebruikt kan worden als basis voor beschrijvingen van speciale webcollecties die in de toekomst samengesteld zullen worden. Daartoe is besloten de collectiebeschrijving in twee delen te splitsen: één algemeen deel waarin uitleg wordt geboden over wat webarchivering is en hoe het in zijn werk gaat, en een inhoudelijk deel over de betreffende collectie waarin onder meer de aard van de opgenomen websites in de collectie wordt toegelicht, de manier waarop deze tot stand is gekomen en de selectiecriteria worden toegelicht.¹⁰

⁸ <https://www.kb.nl/bronnen-zoekwijzers/kb-collecties>

⁹ https://www.theatlantic.com/technology/archive/2017/12/it-might-be-impossible-for-future-historians-to-understand-our-internet/547463/?utm_source=feed

¹⁰ Tenzij expliciet anders aangegeven zijn alle onderzoekswerkzaamheden in dit verslag verricht door Laurie Hasselt.

1. Webarchivering: het hoe en wat

1.1 Waarde als erfgoed: weg is weg

Tallose erfgoedinstellingen houden zich bezig met het verzamelen en bewaren van culturele producten die bewaard moeten blijven voor toekomstige generaties. Met de voortschrijdende inzichten op technologisch gebied van de afgelopen decennia, is er een nieuw type erfgoed ontstaan: *born-digital erfgoed*. Het gaat hier om erfgoed dat van origine digitaal is, en dus geen analoge equivalent kent. Te denken valt aan digitale kunst, digitale foto's waarvan geen fysieke negatieven bestaan en websites of delen daarvan. 'Born digital cultureel erfgoed is bedreigd erfgoed', stelde de Culturele Coalitie Digitale Duurzaamheid (CCDD) in haar rapport dat voortvloeide onderzoek naar de stand van zaken van duurzame toegankelijkheid van born digital bronnen dat in 2014 werd uitgevoerd.¹¹ De CCDD richtte zich daarbij weliswaar enkel op erfgoed op het gebied van film, fotografie, architectuur en kunst, maar voor websites is deze constatering van minstens even zo groot belang.

Alomtegenwoordig is inmiddels de waarschuwing dat men moet oppassen welke informatie men online deelt, want wat eenmaal op het web is gepubliceerd, is daar voor altijd terug te vinden. Maar is dat wel zo? Piet Bakker, die onderzoek doet naar massamedia en digitalisering aan de Hogeschool Utrecht, nuanceert dat beeld van 'het eeuwige internet'.¹² Soms verdwijnen hele stukken van het web, en, waarschuwt hij: "Het digitale hiernamaals heeft geen *undo*-functie. Weg is weg."¹³ Van dat soort verdwenen websites of delen daarvan zijn enkele voorbeelden te noemen: Dichtbij.nl; VK-blog.nl en NUjij.nl, het zijn alle Nederlandse discussieplatforms die inmiddels opgeheven zijn en niet meer online vindbaar.¹⁴ Dit is slechts een greep, zo zijn er nog tal van situaties denkbaar waarin online content verdwijnt in een digitaal zwart gat: websites worden verbouwd, organisaties en hun websites worden opgeheven, media fuseren, krijgen een andere naam, of gaan werken met andere software waardoor online content verdwijnt of onzichtbaar wordt omdat de links niet langer werken.

Werelderfgoedorganisatie UNESCO erkende al in 2003 het belang van het behoud van websites als digitaal erfgoed.¹⁵ Langzaam maar zeker dringt dit besef ook in bredere lagen van de maatschappij door. Zo schreef *NRC* in 2016 een artikel over het belang van webarchieven,¹⁶ en besteedde ook *de Volkskrant* in december 2017 aan webarchiveringsprojecten, waarin ook aandacht was voor de activiteiten op dit gebied bij de KB.¹⁷

¹¹ http://www.den.nl/art/uploads/files/20150622_CCDD-BornDigitalOnderzoek-def.pdf

¹² <http://www.journalismlab.nl/mythe-eeuwige-internet-geheim-grotten/>

¹³ Ibid.

¹⁴ Ook zijn er voorbeelden van het bewust en doelmatig verwijderen van online content. Zo verwijderde de Britse Conservative Party in 2013 zijn speecharchieef, dat speeches bevatte van 2000-2010, van de eigen website, en sommeerde de partij ook het Internet Archive om de gearchiveerde versies van diezelfde pagina's niet langer toegankelijk te maken voor publiek:

<https://www.theguardian.com/politics/2013/nov/13/conservative-party-archive-speeches-internet>

¹⁵ 'Charter on the Preservation of Digital Heritage': http://portal.unesco.org/en/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html

¹⁶ <https://www.nrc.nl/nieuws/2016/05/31/vechten-tegen-digitaal-geheugenverlies-1623567-a539397>

¹⁷ <https://www.volkskrant.nl/tech/de-digitale-stad-duikelt-oude-websites-op-zo-kun-je-baanbrekende-journalistieke-experimenten-reconstrueren~a4542706/>

Tegelijkertijd zijn er echter nog maar weinig onderzoekers die ook de waarde van de gestaag groeiende webarchieven als rijke bron van erfgoedmateriaal inzien.¹⁸ Dat terwijl het internet een dagelijkse en alomtegenwoordige bron van informatie is in het leven van een enorme hoeveelheid mensen wereldwijd. Er is inmiddels veel geschreven over methodologische en andere uitdagingen in webarchivering, maar webarchieven zijn nog maar weinig gebruikt als primaire bron voor historisch onderzoek.¹⁹ De verwachting is echter dat onderzoek naar het web onvermijdelijk zal toenemen, constateren Niels Brügger en Ralph Schroeder: 'In any event, the web, in view of the fact that it is a large and accessible source of data and increasingly important in peoples' lives, is bound to grow as a topic of research.'²⁰ Juist daarom is het zo belangrijk dit onderzoek te faciliteren en ervoor te zorgen dat websites als bron van onderzoek duurzaam opgeslagen worden.

1.2 Webarchiveringslandschap in Nederland

Alvorens het webarchiveringslandschap in Nederland uiteengezet kan worden, is het van belang helderheid te verschaffen over de gebruikte terminologie. Want waar hebben we het eigenlijk over als we spreken over webarchivering? Het proces van webarchivering kan plaatsvinden vanuit archiefperspectief, of vanuit een collectieperspectief.²¹ Dit onderscheid is van belang omdat het gevolgen heeft voor het proces van selectie, opslag, presentatie en gebruik.

Webarchivering vanuit archiefperspectief wordt in Nederland gedaan door de overheid, conform de Archiefwet 1995.²² Die houdt in dat overheidsorganen verplicht zijn hun eigen websites te archiveren om verantwoording af te kunnen leggen aan de burger over het handelen door de overheid. In Nederland dienen overheidsinstellingen hiervoor zelf zorg te dragen. In veel gevallen is kennis omtrent webarchivering niet in huis, dus wordt deze taak uitbesteed aan externe partijen. Archiefweb is daar een voorbeeld van, dat websites van overheidsorganisaties archiveert en toont.²³ Binnen 25 jaar moet de overheid deze gearchiveerde data vervolgens overdragen aan het Nationaal Archief, dat betrokken is bij de opslag van het materiaal. De Erfgoedinspectie concludeert in 2016 echter dat dit in de praktijk niet of nauwelijks gebeurt.²⁴

Het archiveren van websites vanuit collectieperspectief wordt veelal gedaan door erfgoedinstanties die in feite geen webarchief maar een webcollectie beogen samen te stellen: een verzameling van gelijksoortige objecten, die als zodanig opzettelijk bijeengebracht zijn.²⁵ Verschillende instanties in Nederland houden zich bezig met het aanleggen van webcollecties. Hierna volgt een opsomming met enkele grotere webarchiveringsprojecten in Nederland, daarbij nadrukkelijk opgemerkt dat dit overzicht niet compleet is.

¹⁸ Ralph Schroeder & Niels Brügger: *The Web as History* (2017): p. 1

¹⁹ *The Web as History* (2017) van Niels Brügger en Ralph Schroeder is een van de weinige boeken die zich expliciet richt op webarchieven als primaire bron.

²⁰ Ralph Schroeder & Niels Brügger: *The Web as History* (2017): p. 5

²¹ <http://digitaalduurzaam.blogspot.nl/2011/04/>

²² <http://wetten.overheid.nl/BWBR0007376/2015-07-18>

²³ <http://www.archiefweb.eu/>

²⁴ <https://www.erfgoedinspectie.nl/actueel/nieuws/2016/12/7/webarchivering-bij-de-centrale-overheid-gebeurt-nauwelijks>

²⁵ Zie: <http://ingmarbladertenschrijft.blogspot.nl/2011/04/websites-archiveren-of-websites.html>

Het webarchiveringsproject Archipol werd in 2000 opgezet door het Documentatiecentrum Nederlandse Politieke Partijen (DNPP) in Groningen en omvat inmiddels ongeveer 1.000 websites van politieke partijen, politieke bewegingen en politici.²⁶ Het Nederlands Instituut voor Beeld en Geluid in Hilversum archiveert sinds 2014 structureel een selectie websites van omroepen.²⁷ Verder zijn er in Nederland enkele lokale webarchiveringsinitiatieven, zoals die van het Regionaal Archief Dordrecht, het Stadsarchief Rotterdam en de Groninger Archieven. De grootste webcollectie in Nederland is die van de KB, die zich sinds 2007 bezighoudt met het archiveren van websites over de Nederlandse taal, cultuur en geschiedenis. Sindsdien heeft de KB meer dan 13.000 gearchiveerd.²⁸ In totaal zijn er circa 16.000 websites uit het Nederlandse Nationale Domein (NND) gearchiveerd – waarvan het overgrote merendeel dus door de KB – op een totaal van ca. 10 miljoen websites die onder het NND vallen.

1.3 Webarchivering bij de KB

Het KB-webarchief bevat een unieke selectie van digitaal bronmateriaal over de Nederlandse cultuur, die nergens anders ter wereld in deze vorm wordt verzameld, bewaard en kan worden geraadpleegd. Bovendien is dit materiaal binnen afzienbare tijd niet meer op het web te vinden, maar alleen nog maar bij de KB te bestuderen. Voorheen werd het KB-webarchief beschouwd als een bijzondere collectie, maar vanaf 2017 worden websites als onderdeel van de Nederland Collectie beschouwd.

Tot nu toe wordt gestreefd representatieve websites uit Nederland of over Nederland te selecteren, waarbij de primaire focus berust op websites (waarbij tekst overheersend is) die betrekking hebben op Nederlandse taal, cultuur, politiek, geschiedenis en samenleving en de output van de Nederlandse overheid. Onder Nederland wordt verstaan het huidige Koninkrijk der Nederlanden, dus inclusief de BES-eilanden, Aruba, Curaçao en St. Maarten. Voor werken met een historische strekking geldt dat als het onderwerp van de publicatie betrekking heeft op het toenmalige Nederlandse grondgebied of Nederlandse koloniën, deze ook binnen het opnamecriterium vallen. Uiteindelijk is de selectie van de afgelopen jaren afhankelijk geweest van wat op het moment van archivering technisch, organisatorisch en juridisch mogelijk was.²⁹

In de praktijk bewaarde de KB tot nu toe vooral websites over geschiedenis of geschiedenisbronnen, zoals websites van musea, historische verenigingen en erfgoedinstellingen. Veel minder aandacht ging uit naar sites die in de toekomst als bron kunnen worden gebruikt van de geschiedenis. De belangrijkste oorzaak was een gebrek aan tijd en middelen: het totale NND omvat naar schatting 10 miljoen websites. Het is bijzonder veel werk om hier een relevante en representatieve selectie van te maken.

²⁶ <http://www.archipol.nl/>

²⁷ <http://www.beeldengeluid.nl/blogs/researchanddevelopment/201710/dynamischewebcontenthetarchief>

²⁸ Peildatum december 2017

²⁹ Koninklijke Bibliotheek, 'Contentstrategie 2017-2020'

Om aan de behoeftes van toekomstige onderzoekers tegemoet te komen, heeft de KB nu een speciale webcollectie van ‘maatschappijkritische websites’ opgebouwd. Deze collectie omvat sites uit het Nederlandse webdomein die een bron van of reflectie op een actueel en misschien zelfs wel omstreden fenomeen in de maatschappij zijn. Hierbij valt te denken aan maatschappijkritische, al dan niet extreme of onalledaagse denkbeelden, nepnieuws en sites van complotdenkers. Deze websites kunnen online voorbodes zijn van nieuwe ontwikkelingen in de maatschappij.

1.4 Hoe werkt webarchivering

Webarchivering is in essentie het verwerven van dynamische digitale objecten van internet die zo veel mogelijk in hun oorspronkelijke samenhang worden opgenomen in een (bibliotheek-)collectie om te dienen als historische bron van de digitale cultuur van het moment van vastlegging. In de praktijk komt webarchivering in welke vorm dan ook neer op het maken van een onvolledige momentopname van een bepaald deel van het live web.³⁰ Deze activiteit transformeert het dynamische web tot een statische bron van het digitale collectieve geheugen.

Kort gezegd komt het proces van webarchiveren dus neer op het vastleggen voor de lange termijn in een webcollectie van een dynamisch digitaal object als een website, zoals dat op een bepaald moment online was op het wereldwijde web. Desondanks is de website in de gearchiveerde toestand zelden tot nooit identiek aan de originele of live website vanwege het ontbreken van de kenmerkende dynamiek.³¹ Webarchivering is daarom niet zozeer het vastleggen van een historische bron, maar de constructie van een nieuwe bron die zoveel mogelijk kenmerken heeft van de oorspronkelijke live website.³² Hierdoor kan een webcollectie ook nooit een getrouwe afspiegeling zijn van een deel van het live web op een bepaald moment.

De KB archiveert in principe individuele websites in hun geheel, met alle webpagina’s met een URL die daartoe behoren. Voor speciale collecties wordt soms een uitzondering gemaakt en worden ook individuele webpagina’s gearchiveerd. Voor de collectie maatschappijkritische websites is voor het eerst ook aandacht besteedt aan de *webomgeving*: alle sites die met een bepaalde website verbonden zijn door middel van hyperlinks. De ambitie is om ook in de toekomst meer aandacht te hebben voor de webomgeving.

³⁰ Helen Hockx-Yu, The Unknown Aspects of Web Archives. [RESAW Conference, “Web Archives as Scholarly Sources: Issues, Practices and Perspectives”, 8 – 10 June 2015, Aarhus University, Denmark.] <https://hhockx.wordpress.com/2015/08/11/7/>; <https://web.archive.org/web/20160517114541/https://hhockx.wordpress.com/2015/08/11/7/>

³¹ Helen Hockx-Yu, The Unknown Aspects of Web Archives. [RESAW Conference, “Web Archives as Scholarly Sources: Issues, Practices and Perspectives”, 8 – 10 June 2015, Aarhus University, Denmark.] <https://hhockx.wordpress.com/2015/08/11/7/>.

³² Kees Teszelszky, ‘De oogst van de digitale Hollandse akker. Het landschap van webarchivering in Nederland’. In: Overheidsdocumentatie (Od), juli-augustus 2017 (5): p. 14.

1.5 Harvestingstrategie

Voor het binnenhalen van de voor archivering geselecteerde websites gebruikt de KB het programma Heritrix. Dit programma zoekt op het niveau van de domeinnamen en de homepages van de ingevoerde URL naar zoveel mogelijk links naar onderliggende pagina's en verzamelt daarna een deel van de inhoud die met of door deze links verbonden is. De Heritrixversie 1.10.1 die de KB gebruikt, beschikt niet over een gebruikersvriendelijke gebruikersomgeving voor curatoren. De KB gebruikt daarom de open-source applicatie Web Curator Tool (WCT) met een daarin geïntegreerde versie van Heritrix. WCT is gericht op het stroomlijnen van het werkproces van webarchivering en dit gebruikersvriendelijk te maken voor diegenen die de invoer verrichten. Via de WCT kan bovendien metadata over de toestemming en herkomst aan de gearchiverde bestanden worden toegevoegd.³³

³³ Voor meer informatie over de werking van Heritrix zie: <https://github.com/internetarchive/heritrix3>; meer informatie over de WCT: <https://github.com/DIA-NZ/webcurator>

2. Maatschappijkritische webcollectie

2.0 Inleiding

Voor het beschrijven van een bibliotheek- of archiefcollectie wordt vaak een inventaris gemaakt van alle stukken die zich in de collectie bevinden. Maar als een collectie erg groot is, kan dat ervoor zorgen dat door de bomen het bos niet meer zichtbaar is. Webcollecties kunnen al gauw erg groot worden, omdat deze bestaan uit een lijst van URL's, links, losse bestanden en bijbehorende metadata. De webcollectie van de KB bevat na tien jaar webarchiveringsactiviteiten bijvoorbeeld al 13.000 websites, die samen meer dan 30 terabyte aan geheugen beslaan en meer dan 300 miljoen links bevatten.

Zeker als het gaat om speciale collecties die zijn samengesteld rond een bepaald onderwerp of thema, is de samenhang van de collectie, zowel inhoudelijk als op het niveau van de hyperlinks, een belangrijk aspect. Daarom is ervoor gekozen deze speciale webcollectie te beschrijven op collectieniveau.

De collectiebeschrijving biedt inzicht in de webcollectie als geheel. De beschrijving geeft een globaal overzicht van de inhoud ervan en focust op het verband tussen de afzonderlijke stukken in de collectie en de achtergrond van de verwerving, zonder dat de gebruiker overweldigd wordt door een grote hoeveelheid informatie over de afzonderlijke stukken waaruit die samenhang niet direct zou blijken. Een beknopte omschrijving van de collectie, met informatie over de omvang, collectietype, verwerving, vindplaats en samensteller biedt een eerste ingang tot de collectie voor een globaal idee over de inhoud ervan.³⁴ De beknopte omschrijving is gebaseerd op het door model voor collectiebeschrijvingen *Cometa*, dat door verschillende Vlaamse erfgoedinstellingen (Vlaamse Erfgoedbibliotheek, VVBAD, Archiefbank Vlaanderen, FARO) in 2009 werd ontwikkeld om collectiebeschrijvingen op een gestandaardiseerde manier bredere ingang te doen vinden in de erfgoedsector.

Voor de gebruiker die daarna meer wil weten over de collectie en de totstandkoming ervan volgt daarna een uitgebreidere beschrijving van de inhoud van de collectie en de wijze van verwerving.

³⁴ Zie: <http://wiki.cometamodel.be/pagina/Cometa>

2.1 Beknopte omschrijving

algemeen

- Naam: maatschappijkritische webcollectie
- Collectietype: bibliotheekcollectie (web)
- Beschrijving: Maatschappijkritische websites uit of over Nederland, uit de periode 1993-2017 (online in het vierde kwartaal van 2017)
- Doelgroep: onderzoekers en andere geïnteresseerden in het bezit van een KB-pas
- Verwerving:
 - o Linkextractie uit wetenschappelijke publicaties over complotdenkers³⁵
 - o Terugbrengen links tot niveau websites, selectie relevante links
 - o Constructie linkwolk en extractie Nederlandse links uit linkwolk
 - o Werkcollectie van ongeveer 500 sites, aangevuld met suggesties.
- Materialen: websites met een URL
- Omvang: ca. 525 websites
- Taal: Nederlands, met een enkele in het Engels
- Verzamelperiode: 2016-2018
- Ouderdom: 1993-2017
- Locatie: <http://webaccess.kb.nl:8080/archived/> (link werkt alleen onsite; zie bij *Ontsluiting*)

Agenten

- Eigenaar: Koninklijke Bibliotheek
- Beheerder: dienstcoördinator en collectiespecialist webarchivering KB, Peter de Bode
- Collectievormer: Kees Teszelszky, Peter de Bode, Laurie Hasselt

Inhoud

- Concept: maatschappijkritiek
- Plaats: Nederland
- Periode: Post-truthtijdperk, 1993-2017, nadruk op 2012-2017

Collectiebeleid

- Acquisitiebeleid: Afgesloten. Websites worden gedurende 3 jaar iedere 3 maanden geharvest (tot 1 januari 2021).
- Waardering: Deze deelcollectie maakt deel uit van de Nederland Collectie
- Collectiebeleidsplan: Contentstrategie 2017-2020

Ontsluiting

- Raadpleegbaarheid: de gehele collectie is vanaf een jaar na archivering binnen de muren van de KB vrij raadpleegbaar voor jaarpashouders
- Raadpleegbaarheid URL: <http://webaccess.kb.nl:8080/archived/>

Rechten

- De inhoud van de websites valt onder het Nederlands auteursrecht.

³⁵ Jelle van Buuren (2 november 2016), Doelwit Den Haag?: complotconstructies en systeemhaat in Nederland 2000- 2014 (Dissertatie. Institute of Security and Global Affairs, Governance and Global Affairs, Leiden); Jaron Harambam & Stef Aupers (2015). 'Contesting Epistemic Authority: Conspiracy Theories on the Boundaries of Science. Public Understanding of Science', 24 (4), 466-480. doi: 10.1177/0963662514559891

Behoud en beheer

- Percentage van het digitaal materiaal in de deelcollectie waarvan de lange- termijnbewaring voldoende is gegarandeerd door opslag in een digitaal depot op basis van het OAIS-model of een ander gevestigd model voor langetermijnbewaring: op dit moment nul procent, dit is een kwestie waar de KB momenteel volop mee bezig is om dit mogelijk te kunnen maken in de toekomst.

Administratieve gegevens

- Creatie: 25 september 2017
- Laatste wijziging: 21 december 2017
- Auteur: Laurie Hasselt
- Organisatie: Koninklijke Bibliotheek (KB)

2.2 Uitgebreide omschrijving

2.2.1 Aanleiding collectie

Al sinds langere tijd bestaat de wens bij dienstcoördinator webarchivering Peter de Bode om een collectie aan te leggen van zogenaamde ‘abjecte’ websites. Kenmerkende eigenschappen van het internet zijn immers de vrijheid van meningsuiting en publicatie: abjecte opvattingen, aanstootgevende informatie en illegale activiteiten worden makkelijker online dan offline gepubliceerd. Aangezien de KB ernaar streeft een representatief deel van het Nederlandse web met betrekking tot de Nederlandse taal, cultuur en geschiedenis te bewaren voor toekomstig onderzoek, zou zij er dus goed aan doen óók dit deel van de online Nederlandse cultuur te bewaren. Lange tijd was dit echter niet mogelijk vanwege de mogelijke problemen die daarbij zouden kunnen komen kijken. Omdat beschikbaarstelling door de KB van gearchiveerd materiaal met illegale of aanstootgevende inhoud neerkomt op herpublicatie, is het mogelijk dat de KB hiermee een strafbare of in ieder geval een laakbare daad verricht met het aanbieden van dit materiaal aan de gebruiker.³⁶

Om deze reden heeft de KB er tot nu toe voor gekozen om dit soort risicovolle websites niet op te nemen in haar webcollectie. Het gaat daarbij bijvoorbeeld om websites met een politiek of religieus extreem karakter, zoals websites van extreemlinkse of extreemrechtse groeperingen of websites waarop geweld verheerlijkt wordt, groepen beledigd worden, of aangezet wordt tot geweld of haat.

Een andere reden om dit soort websites niet op te nemen in de KB-webcollectie is het gevaar dat opname van een website in de collectie gezien kan worden als erkenning door een overheidsorganisatie als de KB van de doelstellingen of activiteiten van de organisatie of persoon achter de website. Dat is in het verleden al eerder gebeurd: de website *Sekswerkerfgoed* plaatste in 2015 een mededeling op de homepage met de tekst “Door de Koninklijke Bibliotheek erkend digitaal erfgoed.”³⁷ Dat zou tot imagoschade voor de KB kunnen leiden.

Kees Teszelszky vergelijkt dit dilemma met het probleem van de opname en beschikbaarstelling van ‘foute’ kranten uit de Tweede Wereldoorlog via Delpher. Een belangrijk verschil is echter de vluchtigheid van de bron: papieren publicaties worden doorgaans veel langer bewaard dan digitale publicaties. Waar het papieren publicaties betreft, kan de selectie doorgaans dus veel langer uitgesteld worden, tot het moment waarop duidelijk is geworden of de inhoud ook een cultuurhistorische waarde vertegenwoordigt. Maar waardering en selectie van digitale archieven kan niet uitgesteld worden tot dat moment, zeker niet waar het websites met een extreme inhoud betreft. Zo’n website, of delen daarvan, kan bijvoorbeeld op last van de rechter verwijderd worden.³⁸ “In een digitale omgeving dient de waardering bij voorkeur al toegekend te worden vóór of bij de creatie van archieven en dient de selectie plaats te vinden bij of net na de creatie van

³⁶ Rapport 1 Kees: 58

³⁷ Zie: <https://sekswerkerfgoed.nl/>

³⁸ Een voorbeeld hiervan is de website van de linkse activist Joke Kaviaar: <http://jokekaviaar.nl>. Op last van het Openbaar Ministerie werden er delen van haar website verwijderd.

archieven”, constateert ook hoogleraar archiefwetenschap Charles Jeurgens (Universiteit van Amsterdam).³⁹ Door bepaalde websites te selecteren voor archivering anticipeert de KB er al op dat daarop in de toekomst materiaal zal komen te staan dat bewaard moet worden, en wordt in feite dus al toekomstig materiaal binnengehaald.

Het archiveren van dit soort websites brengt dus naast complexe juridische en ethische vraagstukken ook een grote verantwoordelijkheid, en daarmee een aanzienlijke hoeveelheid werk met zich mee. Om deze redenen heeft de KB zich tot op heden niet beziggehouden met het archiveren van dit type omstreden websites. Het onderzoek naar webarchivering bij de KB door Kees Teszelszky heeft echter het belang aangetoond om dit soort websites tóch te archiveren. Bovendien staan dit soort websites vandaag de dag zowel in de maatschappelijke als de wetenschappelijke belangstelling.⁴⁰

2.2.2 Selectiecriteria

De selectie van websites voor het KB-webarchief is gebaseerd op de contentstrategie van de KB.⁴¹ De selectiecriteria worden hieronder puntsgewijs toegelicht:

- Domein: .nl en andere domeinen waarvan de website een nationaal aspect heeft
- Nationale aspect:
 - a) Website in het Nederlands en geregistreerd in Nederland
 - b) Website in een andere taal, geregistreerd in Nederland
 - c) Website geregistreerd in het buitenland, gepubliceerd door een Nederlander
- Content: Websites met betrekking tot Nederlandse taal, cultuur en samenleving; met speciale nadruk op maatschappij- en overheidkritische websites en ‘abjecte’ websites die een actueel beeld geven van de digitale cultuur in Nederland en websites.
- Protocol: http en indien technisch mogelijk https
- IPR: Via opt-out. Auteur/beherende organisatie/persoon moet digitaal bekend en traceerbaar zijn.
- Toegang: Alleen openbaar toegankelijke (delen van) websites, websites achter een inlog worden niet gearchiveerd.
- Formaat: Alle gebruikelijke bestandsformaten die met standaardbrowsers en standaard plug-ins geïnterpreteerd kunnen worden.

³⁹ Charles Jeurgens, ‘Gewaardeerd Verleden’: p. 15

⁴⁰ Zie bijvoorbeeld: *de Volkskrant*, ‘Alt-right in Nederland: Hoe Erkenbrand zich opmaakt voor de strijd om een blanke natie’, 19-11-2017; Jelle van Buuren (2 november 2016), *Doelwit Den Haag?: complotconstructies en systeemhaat in Nederland 2000- 2014* (Dissertatie. Institute of Security and Global Affairs, Governance and Global Affairs, Leiden); Jaron Harambam & Stef Aupers (2015). ‘Contesting

Epistemic Authority: Conspiracy Theories on the Boundaries of Science’. *Public Understanding of Science*, 24 (4), 466-480. doi: 10.1177/0963662514559891

⁴¹ Zie ‘Contentstrategie 2017-2020’.

- Brontype: In de meeste gevallen worden de websites in hun geheel geharvest, in uitzonderingsgevallen worden er ook afzonderlijke pagina's van websites geharvest, bijvoorbeeld bij het ontbreken van een homepage.

2.2.3 Totstandkoming

De speciale KB-webcollectie is zorgvuldig tot stand gekomen. Als vertrekpunt voor de collectie is er onderzoek gedaan naar recent wetenschappelijk onderzoek naar de verspreiding van maatschappijkritische ideeën op het Nederlandse web.

- Hierbij is het proefschrift van Jelle van Buuren en een artikel van Jaron Harambam en Stef Aupers over complotdenken geanalyseerd.⁴² Met behulp het programma pdfx⁴³ zijn er website-URL's geëxtraheerd uit het proefschrift en het artikel.⁴⁴
- Vervolgens zijn daaruit die links gefilterd die naar primaire complotwebsites verwijzen. Links die bijvoorbeeld naar secundaire literatuur en wetenschappelijke teksten verwijzen zijn buiten beschouwing gelaten voor het opbouwen van het corpus.
- Vervolgens zijn de links teruggebracht tot de homepage van de website. Links die verwijzen naar dieper liggende pagina's zijn dus ingekort tot alleen de URL van de startpagina van de website.
- Via netwerkanalyses op het web is er vastgesteld naar welke andere sites er vaak verwezen wordt door de websites uit het corpus, of welke websites er juist veel náár de websites uit het corpus verwijzen. Op deze manier is er getracht een netwerk te reconstrueren van samenhangende websites die met elkaar in verband staan, en om die reden een coherente collectie vormen.
- Diverse deskundigen hebben bovendien suggesties gegeven voor aanvullende websites die ook in connectie staan met het corpus. De groep bestond uit wetenschappers – voornamelijk historici – een OSINT⁴⁵-medewerker van de landelijke politie en een deskundige van de Anne Frank Stichting.
- Tot slot hebben diverse interne medewerkers van de KB suggesties aangedragen voor relevante websites.

2.2.4 Reacties

Het merendeel van de reacties van de website-eigenaren op het archiveringsverzoek door de KB was opvallend positief. De verwachting vooraf was dat er met meer afwijzing en scepsis gereageerd zou worden op het bericht van de KB: veel van de websites zijn uitgesproken kritisch en wantrouwend ten aanzien van de overheid, en de KB is een overheidsinstelling. In de reacties die

⁴² Jelle van Buuren (2 november 2016), *Doelwit Den Haag?: complotconstructies en systeemhaat in Nederland 2000-2014* (Dissertatie. Institute of Security and Global Affairs, Governance and Global Affairs, Leiden); Jaron Harambam & Stef Aupers (2015). 'Contesting Epistemic Authority: Conspiracy Theories on the Boundaries of Science'. *Public Understanding of Science*, 24 (4), 466-480. doi: 10.1177/0963662514559891

⁴³ <https://www.metachris.com/pdfx/>

⁴⁴ <http://journals.sagepub.com/doi/10.1177/0963662514559891>

⁴⁵ Open Source Intelligence

er kwamen op het opt-outbericht, gaven de website-eigenaars vaak te kennen blij en vereerd te zijn dat hun website als belangrijk genoeg beschouwd wordt om te bewaren voor het nageslacht. Sommigen vroegen om nadere informatie over hoe het archiveringsproces technisch in zijn werk gaat, en of er kosten aan verbonden zijn. Een neveneffect van het aanschrijven van de website-eigenaren was dat er bewustwording ontstond over het belang van webarchivering.

Slechts een half dozijn⁴⁶ website-eigenaren gaven geen toestemming voor het archiveren van hun website door de KB. Redenen die daarvoor gegeven werden verschilden van aard: één eigenaar maakte bezwaar tegen archivering omdat er op de betreffende website voornamelijk berichten van derden worden gepubliceerd; een ander kon zich niet voorstellen dat de selectie van de betreffende website door de KB weloverwogen had plaatsgevonden en ging ervanuit dat de selectie middels een automatisch proces had plaatsgevonden, en maakte om die reden bezwaar tegen archivering. Een andere reden voor bezwaar was dat de KB het robots.txt-bestand negeert bij het harvesten van de website, terwijl de beperkingen die dit bestand oplegt er (in dit geval) met een doelbewuste reden zijn, die door de KB genegeerd wordt. Het beleid dat de KB voert met betrekking tot het negeren van de robots.txt getuigt volgens deze website-eigenaar niet van fatsoen. Een laatste bezwaar dat werd aangevoerd is dat de KB niet alle informatie die zich onder een hyperlink bevindt op de website mee-archiveert: dat zou leiden tot een onvolledige archiefversie van de website en daar wilde de eigenaar niet aan meewerken.⁴⁷ De bezwaren die worden gemaakt tegen archivering komen vaker voor en zijn niet typerend voor deze collectie.

Ten slotte was er nog een opvallende negatieve reactie die niet afkomstig was van een website-eigenaar zelf, maar van een lezer van een van de geselecteerde websites. Nadat de eigenaar in een e-mail positief heeft gereageerd op het opt-outbericht van de KB, publiceert hij op zijn website een aankondiging waarin hij meedeelt dat zijn site gearchiveerd zal worden door de KB. In deze periode loopt er een strafrechtelijk onderzoek naar de eigenaar, omdat er smadelijke en lasterlijke uitlatingen op zijn site gepubliceerd zouden worden. Naar aanleiding van de door hem gedane archiveringsaankondiging, heeft de aanklager per e-mail contact opgenomen met de KB. Hij geeft te kennen dat verdere verspreiding van de smadelijke en lasterlijke teksten van de website-eigenaar door de KB er mogelijk toe kunnen leiden dat hij verdere gerechtelijke stappen zal ondernemen tegen de nieuwe verspreider van die teksten, in dezen de KB.

⁴⁶ Peildatum eind december 2017. Op dit moment waren nog niet alle websitteenaars benaderd, het kan dus zo zijn dat er hierna nog meer eigenaars geweigerd hebben.

⁴⁷ Let wel: de hyperlinks zelf op een website worden wel degelijk meegenomen in de harvest, maar de webpagina waarnaar de hyperlink verwijst (tenzij die pagina tot dezelfde website behoort) niet. Dit omdat de KB, zoals gezegd, iedere website-eigenaar afzonderlijk moet aanschrijven, en dus hyperlinks die verwijzen naar andere websites om die reden niet zonder meer mag archiveren zonder ook die eigenaren te benaderen.

3. Wat niet wordt gearhiveerd

3.1 Algemeen

Volgens het selectiebeleid zijn websites met games (digitale spelletjes), webcams en andere live content (bijvoorbeeld camera's op internet), intranets en andere online content die zich achter een inlog bevindt, radio- en televisieprogrammasites, pagina's op sociale mediasites zoals Twitter, Facebook en LinkedIn uitgesloten van opname in de webcollectie. Ook commerciële websites zoals webshops worden doorgaans in principe niet opgenomen in de collectie, tenzij daar een gegronde reden voor is. In de collectie maatschappijkritische websites is bijvoorbeeld wel een webshop voor eindtijdpreppers opgenomen, omdat die vanwege de inhoud wel tot deze collectie behoort.⁴⁸

Daarnaast worden er volgens het selectiebeleid geen websites gearhiveerd die door andere instellingen worden gearhiveerd, zoals websites van politieke partijen (die gearhiveerd worden door Archipol in Groningen) of websites van omroepen. Voor deze webcollectie is bijvoorbeeld de website van de Nederlandse Volks-Unie (NVU) geselecteerd, wat een politieke partij is en daarom dus niet door de KB gearhiveerd zal worden. Wel blijven alle websites die geselecteerd zijn voor de collectie op de selectielijst staan die beschikbaar gesteld wordt aan het publiek. Naast dit soort websites die niet gearhiveerd worden omdat ze uitgesloten zijn van het selectiebeleid van de KB, zijn er nog juridische en technische beperkingen die het archiveren van bepaalde websites bemoeilijken of onmogelijk maken, zoals hieronder nader uiteengezet.

3.2 Juridische beperkingen

Naast het gebrek aan een wettelijke depotfunctie in Nederland – dat ervoor zorgt dat de KB maar een fractie van de Nederlandse websites kan archiveren – zijn er nog enkele andere juridische beperkingen van kracht die ervoor zorgen dat bepaalde websites niet gearhiveerd mogen worden. Websites kunnen over het algemeen worden geraadpleegd op het live web zonder dat daar juridische consequenties aan verbonden zijn, tenzij het om illegaal downloaden gaat via een website of het doelbewust zoeken naar en bewaren van strafrechtelijk verboden materiaal. Voor deze collectie maatschappijkritische sites is een aantal websites geselecteerd waarop illegale content gepubliceerd is. Het gaat in alle gevallen om uitingsdelicten.⁴⁹ De illegale content heeft veelal betrekking op het beledigen van een groep mensen, discriminatie of aanzetten tot geweld. Ook Holocaustontkenning komt meermaals voor, of het tonen van hakenkruizen. Bij twijfel over de toelaatbaarheid van de inhoud van een site, is er in overleg met de juridische afdeling van de KB en de afdeling communicatie besloten deze sites ook *niet* te archiveren.

Wat echter wel mag, is het selecteren van webadressen op het live web, deze in een database zetten en de lijst publiceren. Met name voor deze collectie, die niet alleen abject, maar waarschijnlijk in sommige gevallen ook illegaal materiaal bevat, is dit van belang. De URL's van de websites die niet gearhiveerd worden, blijven dus wel bestaan op de selectielijst en worden

⁴⁸ <https://www.prepz.nl>

⁴⁹ <https://nl.wikipedia.org/wiki/Uitingsdelict>

meegenomen in de collectiebeschrijving, zodat voor de gebruiker van de webcollectie wel informatie behouden blijft over welke websites er tot deze collectie behoren. Ook de hyperlinks die deze websites bevatten zijn geogost en meegenomen in de netwerkanalyse van de collectie. Zolang de niet-gearchiveerde websites online zijn, kunnen die sites alsnog bekeken en bestudeerd worden samen met de rest van de collectie.

3.3 Technische beperkingen

Een van de grootste technische beperkingen voor het archiveren van websites is dat het bijzonder lastig is om webpagina's met complexe JavaScript (JS) te archiveren. JS is een veelgebruikte scripttaal bij webdesign om webpagina's interactief te maken en webapplicaties te ontwikkelen. Het archiveren van webpagina's met JS levert vaak grote hoeveelheden onbruikbare bijvangst op. Bovendien wordt deze JS alleen geladen door zeer specifieke gebruikersinteracties die niet te dupliceren zijn. Een website wordt meestal interactief voor de bezoeker door het gebruik van JS, maar kan juist daardoor meestal niet worden gearcheveerd of worden afgespeeld. Een site met JS kan daarom in veel gevallen worden beschouwd als dood digitaal erfgoed. Vrijwel alle moderne websites bevatten JS: uit een onderzoek van de British Library blijkt dat 95% van de bestaande websites in 2010 JS bevat. Dat betekent dat vrijwel alle sites met complexe JS die op dit moment online zijn niet of niet volledig kunnen worden gearcheveerd.⁵⁰

Websites met een HTTPS-protocol (Hypertext Transfer Protocol Secure) kunnen ook technische moeilijkheden opleveren voor het archiveren van de websites. HTTPS is een protocol dat wordt gebruikt om informatie op een website te versleutelen. Het is de tegenhanger van het minder beveiligde HTTP (Hypertext Transfer Protocol). De versie van het softwareprogramma Heritrix waarmee de KB de websites harvest, kan met bepaalde HTTPS-protocollen niet overweg. Voordat een site geharvest wordt is echter niet te achterhalen of de betreffende site die protocollen gebruikt die het archiveren van de site onmogelijk maken. Daarom wordt toch iedere website met een HTTPS-protocol geharvest. Pas nadat de website geharvest is met Heritrix kan met kwaliteitscontrole achterhaald worden of de website goed is binnengekomen of niet.

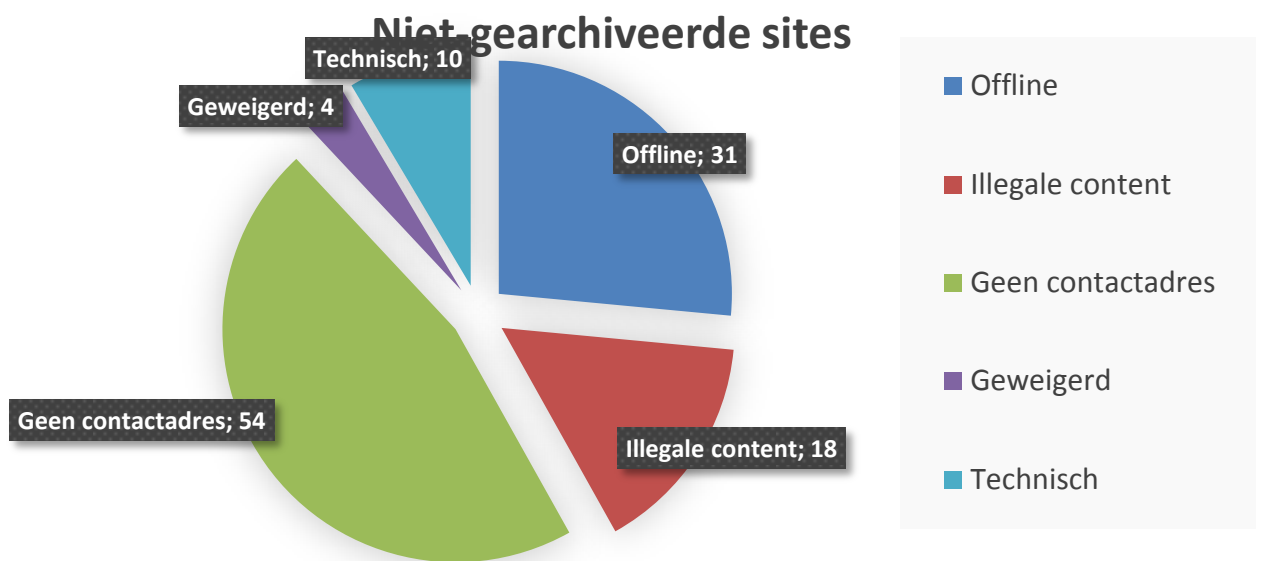
3.4 Overzicht niet-gearchiveerde websites

Van de ca. 500 websites die geselecteerd zijn voor de speciale collectie maatschappijkritische sites kunnen er een hoop niet gearcheveerd worden, om uiteenlopende redenen. Veel sites bleken in de periode die tussen de selectie en het moment dat begonnen werd met het archiveren van de websites offline te zijn gegaan. Daar het invoeren van de websites uit deze collectie in de WCT anno december 2017 nog niet afgerond is, zal dit getal naar verwachting nog oplopen. Het laat eens te meer zien hoe urgent het is websites te archiveren, en hoe snel er moet worden beslist en gehandeld om websites niet verloren te laten gaan. Sommige (delen van) websites die offline zijn, zijn nog elders te raadplegen. De stukken van de website <http://nachthemel.nl/>, inmiddels offline, zijn bijvoorbeeld doorgeplaatst op <http://www.vrijspreker.nl/wp/author/nachthemel/>. Het is niet bekend of daarbij alle content is overgeplaatst of slechts een deel daarvan. Een andere trend die

⁵⁰ <http://blogs.bl.uk/webarchive/2014/08/web-archiving-in-the-javascript-age.html>

zichtbaar is is dat websites offline gaan, maar als Facebookcommunity blijft bestaan. Een voorbeeld daarvan is Geldgeweld.⁵¹

Het ontbreken van een wettelijk depot in Nederland is het tweede grootste obstakel bij het archiveren van de sites. De KB moet iedere website-eigenaar afzonderlijk inlichten dat het de intentie heeft de betreffende website te archiveren. Daarvoor is het noodzakelijk dat er een contactadres te vinden is waarop de eigenaar benaderd kan worden. Bijna 10% van de geselecteerde sites had geen contactmogelijkheid, waardoor de website-eigenaar niet benaderd kon worden en de websites ten gevolge daarvan niet gearchiveerd konden worden.



Schematisch overzicht redenen waarom websites niet gearchiveerd konden worden

⁵¹ URL van de website (inmiddels offline): <http://geldgeweld.nl/>

URL van de nog bestaande Facebookcommunity: <https://www.facebook.com/Geldgeweld/>

4. Collectie

4.1 Algemeen

De collectie ‘maatschappijkritische websites’ is een speciale thematische collectie binnen de gehele KB-webcollectie van ca. 13.000 websites. Met de selectie van deze sites beoogt de KB allerm minst volledigheid na te streven, maar wel representativiteit van het brede scala aan denkbeelden dat er vandaag de dag circuleert in Nederland. Dat levert een bonte stoet aan websites op: van anarchisten tot nationalist en van ‘eindtijdpreppers’ tot aanhangers van de platte-aarde-theorie. Achter die ogenschijnlijke veelzijdigheid schuilt echter ook een gemene deler: het gros van de websites is kritisch ten aanzien van de maatschappij, de overheid en vrijwel alles waar het predicaat ‘mainstream’ aan gegeven kan worden. Omdat de aard van de kritiek die de websites leveren zo veelzijdig is, is de collectie middels trefwoorden opgedeeld in subcategorieën. De categorieën zijn thematisch van de aard en hebben betrekking op de inhoud van de teksten op de websites.

De inhoud van de websites is onderhevig aan veranderingen door de tijd heen. De categorisering op basis van thematische inhoud zal dus wellicht in de toekomst niet meer overeenkomen met de inhoud van de websites op moment van beschrijven. Dit probleem is echter niet te omzeilen, enkel kan de disclaimer gegeven worden dat de classificatie van de inhoud van de sites geldt op het moment van dit onderzoek (september-december 2017).

Iedere website van de lijst is afzonderlijk bestudeerd. De categorisering is grotendeels gedaan aan de hand van de doelstelling, missie of gelijksoortige informatie waarmee de website-eigenaar zijn of haar eigen website omschrijft (informatie die vaak wordt gegeven in een rubriek ‘over’, ‘informatie’, ‘doelstelling’ of ‘missie’). Als die informatie niet aanwezig was op een website, is de categorisering gedaan op basis van hoofdthema’s van de artikelen op een site, die meestal blijkt uit de titels van subrubrieken. Deze categorisering is vanzelfsprekend niet sluitend: sommige categorieën lijken thematisch erg op elkaar, soms is er sprake van overlap en vallen websites onder meerdere categorieën. Ze dienen dan ook vooral om nadere globale informatie te geven over wat voor soort websites er tot de collectie behoort en over welke thema’s deze handelen, zoals hieronder nader toegelicht.

4.2 Categorieën

a. Complot

Websites van complotdenkers zijn met 66 websites ruim vertegenwoordigd in de collectie. Complottheorieën hebben met name betrekking op schokkende gebeurtenissen “die schreeuwen om een grote verklaring”.⁵² De officiële verklaringen die gegeven worden door de autoriteiten worden beschouwd als ontoereikend of ze worden helemaal niet geloofd. De websites in de collectie schrijven over allerlei verschillende complotten. Zo wordt er veel geschreven over de MH17-vliegtuigramp in 2014, de vermeende pedofilie van voormalig Nederlands topambtenaar Joris Demmink, de aanslag op de Twin Towers op 9 september 2001, de moord op Marianne Vaatstra, de Bijlmerramp (1992) en de moord op Theo van Gogh (2004).

Naast dit soort specifieke gebeurtenissen, wordt er veel geschreven over buitenaards leven en ufo's, graancirkels, 'geo-engineering' (het ingrijpen in natuurlijke systemen met als doel klimaatverandering), 'chemtrails' (chemische stoffen die uitgestoten worden door vliegtuigen in opdracht van de overheid) en de theorie dat de aarde plat en niet rond is. Op abstracter niveau is de retoriek hoofdzakelijk dat wij allen gevangen zitten in een kwaadaardig systeem, waarbij er achter de schermen duistere machten proberen kritische burgers de mond te snoeren en totale controle uit te oefenen. Complotdenkers roepen dan ook vaak op om verklaringen van autoriteiten niet te geloven, maar vooral zelf onderzoek te doen. Een belangrijke term is 'wakker worden': “de complotdenkers zijn al wakker, de grote opgave is om de rest van de bevolking ook wakker te maken.”⁵³ Zij zijn op zoek naar de échte waarheid: bij meerdere complotwebsites komt zelfs in de URL het woord 'waarheid' voor.⁵⁴

b. Anti-complot

Naast de complottheoretici zijn er websites die juist die samenzweringen proberen te bestrijden. Met “mondsnoerende logica”⁵⁵ en scepsis worden complottheorieën, pseudowetenschap, mystiek en alternatieve geneeskunde bekritiseerd en soms belachelijk gemaakt door de verspreiders van die ideeën te bestempelen als 'aluhodjes': sommige complotdenkers dragen hoedjes gemaakt van aluminiumfolie die de hersenen zou beschermen tegen elektromagnetische straling of gedachtenlezen.

c. Rechts

Websites die deel uitmaken van deze subcategorie hebben enkele gemeenschappelijke kenmerken. Met name nationalisme, kritiek op de islam en moslims en een afwijzende houding ten aanzien van migratie en vluchtelingen zijn onderwerpen in deze categorie. Ook zijn er websites die tot het extreem-rechtse spectrum behoren, afgaande op de kenmerken: racisme, antisemitisme, haat tegen homo's, moslims en/of allochtonen, ultra-nationalisme, verwerpen van democratische

⁵² Jelle van Buuren in de Groene

⁵³ ibid. Zie bijvoorbeeld de website: <http://www.slaaptgijnog.nl/>

⁵⁴ <http://www.waarheid911.nl/>; <http://www.deverborgenwaarheid.nl/>; <http://hetuurtvandewaarheid.info/>; <http://www.alleendewaarheid.nl/>; <http://www.deaardsewaarheid.com/>.

⁵⁵ <https://barracudanls.blogspot.nl/>

grondbeginselen en een positieve associatie met het Derde Rijk.⁵⁶ Op sommige websites worden bijvoorbeeld hakenkruizen afgebeeld, wordt de Holocaust ontkend en/of wordt er gesproken over rassenzuivering. Veel van deze extreem-rechtse content is strafbaar en om die reden niet gearhiveerd. Ook zijn er een paar sites die behoren tot de zogenaamde alt-rightbeweging, een recente beweging die met name in de Verenigde Staten leeft en grotendeels op internet is begonnen en uitgegroeid tot een bredere beweging.⁵⁷ De Nederlandse website <http://altrechts.com> is daar een voorbeeld van, die website is inmiddels offline. Een mogelijke aanleiding daartoe zijn meldingen die er bij het Meldpunt Discriminatie Internet zijn gedaan naar aanleiding van lijsten met Joodse Nederlanders en ‘volksvijanden’ die de website publiceerde.⁵⁸

d. Links

Websites in het linkerspectrum bevatten sites van en over kraken, anti-fascisme, anti-racisme, anti-militarisme, anarchisten en arbeidersbewegingen. De anti-fascistische bewegingen bestaan bijna alle uit lokale groepen. Op het eerste gezicht is het wellicht opvallend dat websites zo lokaal georiënteerd zijn: het internet heeft immers weinig boodschap aan grenzen en geografie.⁵⁹ De websites zijn dan ook eerder een middel dan een doel: de anti-fascistische groepen hebben als voornaamste doel om demonstraties en protestacties te organiseren, meestal tegen extreem-rechtse groepen als Pegida of de NVU, en sympathisanten oproepen daaraan deel te nemen. Die demonstraties vinden niet online maar op straat plaats en zijn daarmee dus wel gebonden aan geografie, wat wellicht een verklaring biedt voor de plaatsgebonden oriëntatie van deze groepen en hun websites.

Websites met betrekking tot de kraakbeweging hebben vaak betrekking op een specifiek kraakpand of een groep kraakpanden die soms functioneren als een min of meer autonoom ‘dorp’.⁶⁰ Veel van deze kraakpanden zijn inmiddels ontruimd, omdat met de Wet Kraken en Leegstand die op 1 oktober 2010 in werking trad kraken illegaal werd. De panden werden veelal bewoond door kunstenaarscollectieven die daar kunst produceerden en tentoonstelden. Daarnaast zijn er websites van meerdere kraaktijdschriften te vinden in de collectie. Deze tijdschriften hebben vaak een opmerkelijk oud archief: het gaat meestal om tijdschriften die in de jaren tachtig en negentig op papier werden uitgegeven en verspreid in de kraakbeweging en inmiddels alleen online voortleven.⁶¹

⁵⁶ Deze kenmerken zijn ontleend aan het rapport ‘Rechtsextremisme op het internet’ van Meldpunt Discriminatie Internet uit 2009.

⁵⁷ Een voorbeeld daarvan is <http://www.erkenbrand.nl/>, een Nederlandse groep die zichzelf omschrijft als een ‘alt-right studiegenootschap’. In *de Volkskrant* verscheen een uitgebreid artikel over deze beweging: <https://www.volkskrant.nl/binnenland/alt-right-in-nederland-hoe-erkenbrand-zich-opmaakt-voor-de-strijd-om-een-blanke-natie~a4539593/>

⁵⁸ <https://nos.nl/artikel/2170558-cidi-wil-site-met-lijsten-joodse-nederlanders-en-volksvijanden-offline.html>; https://www.reddit.com/r/thenetherlands/comments/683a65/cidi_wil_site_met_lijsten_joodse_nederlanders_en/

⁵⁹ In de URL’s van de groepen komt vaak een plaatsbepaling voor: <https://www.afafrislan.nl/>; <https://afadenhaag.wordpress.com/> of <https://afanoordbrabant.noblogs.org/>

⁶⁰ Bijvoorbeeld de websites <https://stuiverstraat.dse.nl/>; <https://schoolantarctica.wordpress.com/>; <http://bajesdorp.nl/> of <http://deslang.nl/>

⁶¹ Zo is op <http://www.ravagedigitaal.org> het archief van tijdschrift *Ravage*, voorheen de *Nixnut* te raadplegen tot 1988 en op <http://www.konfrontatie.nl/> is het archief van links maandblad *Konfrontatie* tot 1993 te vinden.

e. Gaza

Het conflict tussen Israël en Palestina in de Gazastrook wordt regelmatig gethematiseerd: soms vormt dit het hoofdthema waarover de website handelt, in andere gevallen is het slechts een van de vele onderwerpen die voorkomen op een site. Wat in het laatste geval opvalt, is dat websites die zich in berichten over het Gaza-conflict aan de zijde van Israël scharen, daarbij veel schrijven over de negatieve invloed van asielzoekers en in het algemeen kritisch zijn op de islam.⁶² Websites die zich echter tégen Israël keren en positief berichten over Palestina, schrijven daarnaast veel positieve berichten over feminisme, socialisme en andere ‘linkse’ thema’s.⁶³

Daarnaast zijn er enkele websites die niet zozeer duidelijk ‘een kant kiezen’, maar vooral beogen uitleg en informatie te geven over het ontstaan en verloop van het conflict Israël-Palestina en/of ondersteuningshulp te verlenen in de Gazastrook zelf.⁶⁴

f. Gender

In de categorie gender bevinden zich zowel feministen als anti-feministen, maar ook sites die juist mannelijkheid thematiseren. Zo vraagt de <https://themasculinistsite.wordpress.com> zich af ‘Wat [het] betekent om man te zijn in onze huidige maatschappij?’. Ook is er een christelijke mannenbeweging⁶⁵ en wordt er op <https://alfanl.com> gepleit voor de ‘alfaman’. Op de banner van laatstgenoemde website zijn afbeeldingen te zien van de Amerikaanse president Donald Trump, maar ook van de Nederlandse politicus Thierry Baudet. Aan de andere kant van het spectrum bevinden zich de feministische websites, die vaak niet alleen feminisme maar ook *queer* thematiseren. Zo omschrijft <https://feministischverzet.noblogs.org/> zichzelf als ‘we are queer, we are radical, we are feminist yeah!’.

g. Alternatieve media

Websites onder het kopje alternatieve media lijken soms veel op de websites over complottheorieën: ze geven vaak te kennen ook op zoek te zijn naar ‘de echte waarheid’ die verzwegen (al dan niet bewust) zou worden in andere media. De sites profileren zich expliciet als ‘alternatieve’ en ‘onafhankelijke’ nieuwsplatforms en zetten zich af tegen ‘mainstream media’: denk aan landelijke dagbladen als *NRC*, *Trouw* of *de Volkskrant* en het *NOS Journaal*. Nieuws afkomstig van deze media wordt weggezet als ‘gekleurd’ of ‘propaganda’. Sommige houden zich voornamelijk bezig met het analyseren en weerleggen van wat deze media berichten.⁶⁶ Daarnaast zijn er enkele overzichtssites en platforms die vooral links naar allerlei alternatieve nieuwssites publiceren. Veel van deze alternatieve nieuwssites verwijzen naar elkaar.

h. Spiritualiteit

Websites die schrijven over spiritualiteit lopen thematisch uiteen van persoonlijke spirituele ervaringen en innerlijke bewustwording, te bereiken via yoga, mindfulness of spirituele

⁶² Zoals bijvoorbeeld <https://martienpennings.wordpress.com/>

⁶³ <http://www.anjameulenbelt.nl/weblog/>; <http://grenzeloos.org>

⁶⁴ <http://www.vecip.nl/>; <http://kifaia.nl/>

⁶⁵ <https://de4emusketier.nl/>

⁶⁶ Zoals <http://www.nrcombudsman.nl/>; <http://blikopnosjournaal.blogspot.nl/> en <http://www.leugens.nl/>.

groepsreizen, tot websites die zich meer bezighouden met bovennatuurlijke zaken als geesten, energiestromen en de kosmos. Zo wordt er geschreven over helderziendheid, contact maken met Moeder Aarde of wordt er voorspeld dat er een nieuw tijdperk voor de aarde zal aanbreken die zal leiden tot een ‘galactische federatie of samenleving’ en zijn er websites die workshops wichelroedelopen aanbieden. Veel van deze websites hebben een sterrenhemel of planetenstelsel als achtergrond van hun site. De teksten op deze sites zetten zich af tegen de gangbare, rationele manier van denken en zijn in die zin als maatschappijkritisch te beschouwen.

i. Gezondheid

Websites op het gebied van gezondheid vallen uiteen in verschillende subcategorieën, maar zijn in ieder geval alle tegen de reguliere gezondheidszorg en vaak op zoek naar alternatieven daarvoor. Enkele websites laten zich kritisch of ronduit afwijzend uit over vaccinaties. Veel van deze sites worden beheerd door ouders die aangeven ongerust te zijn over de gezondheid van hun kinderen naar aanleiding van het krijgen van vaccinaties. Ze missen wetenschappelijke onderbouwing voor de goede en betrouwbare werking van vaccins en belichten de negatieve bij-effecten die inentingen volgens hen kunnen hebben (zoals het ontwikkelen van autisme).⁶⁷ Andere websites in deze categorie pleiten meer in het algemeen voor het wantrouwen van de reguliere gezondheidszorg en promoten verschillende methoden uit de alternatieve geneeskunde. Daarnaast zijn er websites gericht op de schadelijke effecten van elektromagnetische stralen op het lichaam en websites over de schadelijke effecten van genetisch gemanipuleerd voedsel. Op veel van deze websites wordt dus met scepsis gereageerd op recente technologische en wetenschappelijke ontwikkelingen.

j. Religie

Veel websites hebben een religieus karakter. Het gaat daarbij vooral om kritische beschouwingen van gangbare opvattingen in een bepaalde religie. Net als bij de complottheoretici gaat het vooral om zelf onderzoek doen en niet zomaar aannemen wat wordt verkondigd, bijvoorbeeld door de kerk. Het merendeel van de sites heeft betrekking op het christendom. De sites bevatten kritische bijbellezingen, verkondigen het evangelie of bepleiten het christen-anarchisme. Enkele sites verkondigen vooral tegen bepaalde dingen te zijn: tegen abortus, drugs, alcohol, housemuziek, occult, Pokémon en/of Harry Potter. Daarnaast is er een aantal profeten die beweren berichten door te krijgen uit hemelse sferen om die vervolgens op hun website verder te verspreiden. Een heel klein aantal websites heeft betrekking op de islam of het jodendom.

k. Crimesites

Op crimesites wordt bericht over criminele activiteiten, al dan niet met betrekking tot een bepaald circuit of een bepaalde locatie. Het zijn sites die handelen over activiteiten die buiten de reguliere maatschappij vallen en zich daarom kritisch daartoe verhouden. Sommige van deze sites worden

⁶⁷ Bijvoorbeeld <http://verontrustemoeders.nl/> of <https://www.vaccinvrij.nl/>.

gemaakt door ex-criminelen die zelf ooit veroordeeld zijn geweest.⁶⁸ De inhoud van deze sites is soms omstreden: zo sprak Martin Kok op <http://www.vlinderscrime.nl/> over criminele activiteiten waarbij hij daders met naam en toenaam noemde. Op last van de rechter heeft hij bepaalde teksten van zijn website moeten verwijderen. Kok werd op 8 december 2016 geliquideerd, waarna er een foto van zijn opgebaarde lichaam door bekenden van hem op zijn website werd gepubliceerd. Later werd de website helemaal uit de lucht gehaald. Meerdere website-eigenaren van crimesites liggen in de clinch met de bekende televisiemisdadjournalist Peter R. de Vries en refereren daaraan op hun site.⁶⁹

l. Natuur

Websites met betrekking tot het thema natuur lopen uiteen van websites over kernenergie tot websites over het klimaat en klimaatverandering, tegen steenkool en biopolitiek en websites van radicale milieuactivisten. Ze richten zich nadrukkelijk tegen de manier waarop de mens in de huidige maatschappij omgaat met dier, natuur en milieu.

m. Economie

Bij de economische websites bevinden zich sites die zich richten op specifieke thema's, zoals een belangenorganisatie voor mensen met een minimuminkomen, meerdere websites die ageren tegen het gebruik van de pinpas en andere digitale betaalmiddelen en pleiten voor contant geld en websites die schrijven over de restschuld, de economische crisis en de woningmarkt. Daarnaast zijn er websites die zich in bredere zin bezighouden met economie en pleiten voor een hervorming van het economische stelsel. De sites zetten zich op verschillende manieren af tegen de huidige invulling van het economische stelsel.

n. Overheid

Websites in deze categorie richten zich tegen verschillende aspecten van de overheid. Sommige websites zijn kritisch op de jeugdzorg in Nederland, meestal vanuit eigen ervaring. Anderen stellen in het algemeen misstanden in de rechtspraak aan de kaak. Daarnaast zijn er enkele websites die in het algemeen tegen de overheid of in ieder geval overheid-kritisch zijn. Een enkele site is kritisch op het koningshuis en pleit voor een republiek. Dan zijn er nog sites die zich niet tegen de Nederlandse overheid maar tegen de Europese regering keren en een handvol websites die kritisch is op het Nederlandse onderwijssysteem en/of pleit voor thuisonderwijs.

o. Eindtijd: preppers en religieus

In de categorie websites die schrijven over de eindtijd zijn er twee subcategorieën te onderscheiden: religieuze websites en websites van en voor 'eindtijdpreppers', zij die zich voorbereiden (van het Engelse 'prepare') op de eindtijd. De religieuze sites beroepen zich op de Bijbel en geloven op basis daarvan dat Jezus spoedig zal wederkeren en de uitverkorenen

⁶⁸ Bijvoorbeeld de twee websites van Steven Brown: <https://stevenbrownsblog.wordpress.com/> en <https://demaffiajournalist.wordpress.com/> en de website van Martin Kok: <http://www.vlinderscrime.nl>, die inmiddels offline is.

⁶⁹ De eerdergenoemde Steven Brown is daar een voorbeeld van, de website <http://www.antivries.eu/> is een website die volledig gewijd is aan het bekritisieren en zwartmaken van De Vries.

opgenomen zullen worden in de hemel, waarna de aarde zal vergaan. Ze zijn niet zozeer bezig zich fysiek voor te bereiden op de eindtijd, maar vooral mentaal. Daarnaast zijn ze vooral bezig met het verkondigen van de nakende eindtijd, en proberen ze met bewijzen uit bijbelteksten en tekenen als planeetstanden de lezer te overtuigen dat de eindtijd nabij is. De preppers houden zich niet zozeer bezig met de oorzaak of precieze vorm waarin de eindtijd zal verschijnen, maar zijn vooral bezig met hoe te overleven als die eenmaal is aangebroken. Zonder uitzondering komt er een afleiding van het woord preppen voor in de URL van deze sites. Ze geven elkaar survivaltips op fora over hoe zelfvoorzienend te leven en hebben webshops opgericht waarin preppers de daarvoor benodigde materialen kunnen kopen.

p. Privacy

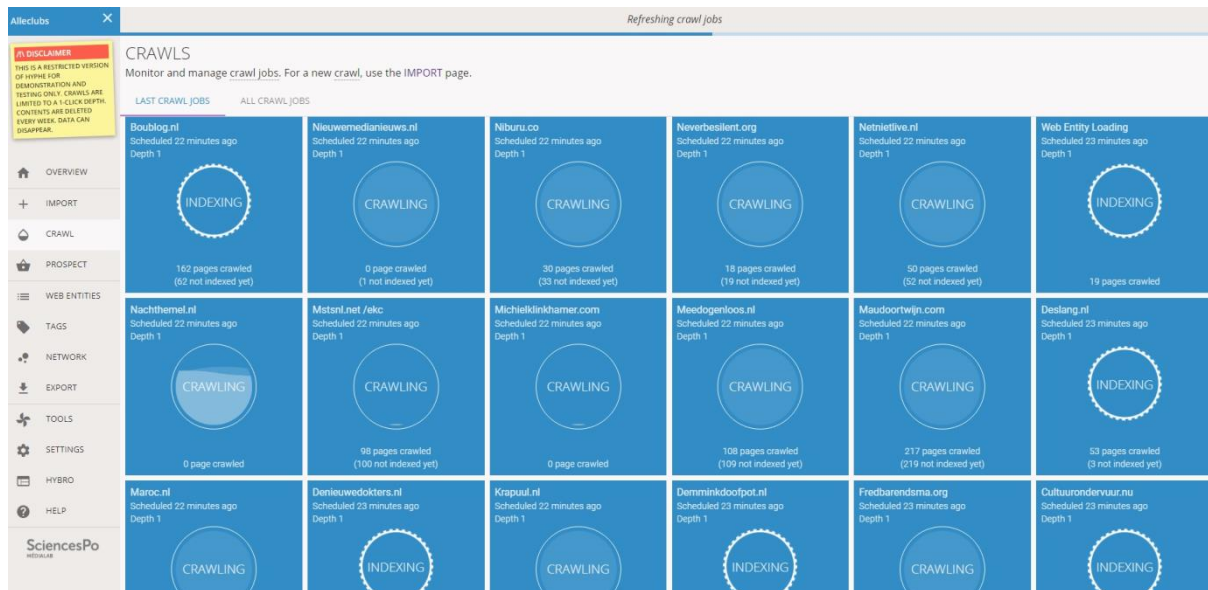
Websites die privacy thematiseren zijn onder andere tegen het elektronisch patiëntendossier, tegen de identificatieplicht en tegen de zogenaamde ‘slimme meters’ van energiebedrijven die makkelijk te hacken zouden zijn en waarmee persoonsgegevens gestolen zouden kunnen worden.

4.3 Websfeer

Met de tool Hyphe⁷⁰ zijn verschillende analyses uitgevoerd waarmee onderzocht is hoe de websites onderling met elkaar verband staan, maar ook hoe de collectie zich verhoudt tot ander digitaal materiaal. Het gaat hierbij nadrukkelijk om verkennend onderzoek, waarmee is nagegaan welke links naar webpagina’s er op een website te vinden zijn. Daarvoor is een demoversie van het programma gebruikt, dat enkele beperkingen kent. De demo onderzoekt slechts 1 laag van de URL die ingevoerd wordt. Dat betekent dat als we bijvoorbeeld zouden willen onderzoeken naar welke webpagina’s er allemaal wordt gelinkt op de website van de KB en daarvoor de URL van de homepage <https://www.kb.nl/> ingevoerd zou worden in Hyphe, het programma alleen onderzoekt welke links er allemaal op de homepage van de KB te vinden zijn, en *niet* welke links er allemaal te vinden zijn op dieperliggende pagina’s van de KB-website. Niettemin resulteert dit verkennende onderzoek al in interessante inzichten in de collectie die niet op een andere manier te verkrijgen zijn.

Het doel van deze netwerkanalyse is het leveren van *halffabricaten* voor onderzoekers: door dit eerste verkennende onderzoek dat al gedaan is door de KB, krijgen onderzoekers sneller inzicht in en overzicht over de collectie en een idee over wat voor onderzoek er potentieel mogelijk is naar deze collectie. Een interessant resultaat dat hierdoor naar voren kwam was bijvoorbeeld de aanwezigheid van Russische websites die veel in contact stonden met complotdenkers. Dit kan een aanknopingspunt zijn voor bijvoorbeeld datawetenschappers om verder onderzoek te verrichten naar hoe dit netwerk er precies uit ziet en hoe het werkt. Bovendien is het makkelijk dupliceerbaar onderzoek: het gebruikte programma is relatief eenvoudig om mee te werken en iedereen kan de uitgevoerde analyses nogmaals uitvoeren, mits de websites nog online zijn en niet al teveel veranderd zijn.

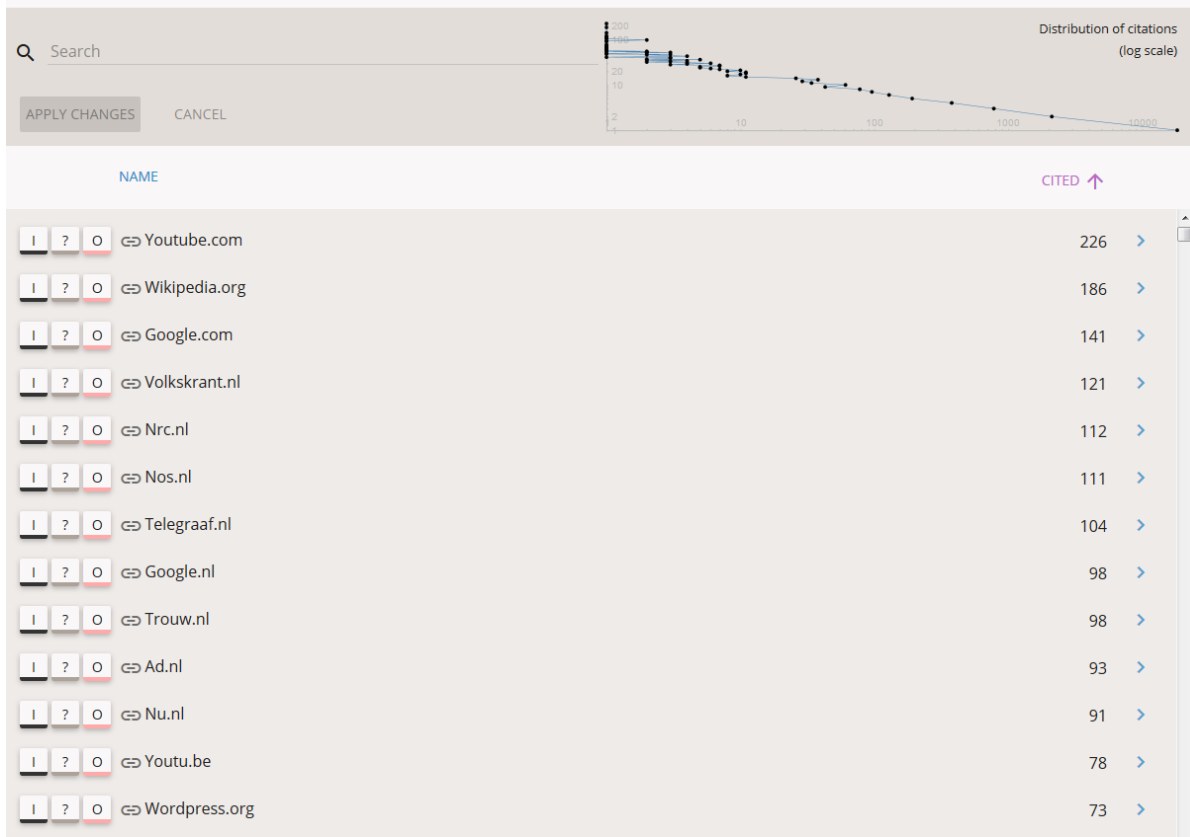
⁷⁰ <http://hyphe.medialab.sciences-po.fr/>



Afbeelding 1: Hyphe is bezig met het binnenhalen van alle links

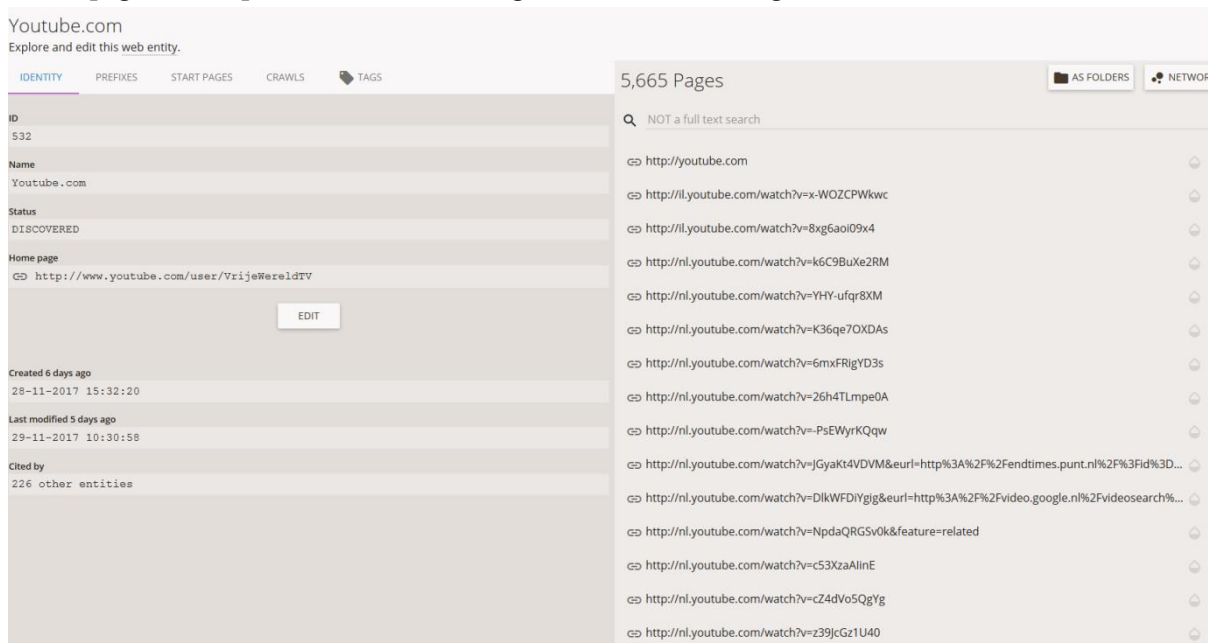
Bij het invoeren van alle URL's van de websites die deel uitmaken van deze webcollectie is Hyphe urenlang bezig met het binnenhalen van alle links die zich op deze websites bevinden (zie afbeelding 1). Uiteindelijk resulteert dat in 22.546 links die het programma heeft weten te extraheren van deze websites. In het Prospect-overzicht van het programma is vervolgens te vinden naar welke websites er het meest gelinkt wordt en door hoeveel verschillende websites (zie afbeelding 2).

PROSPECT 22,546 DISCOVERED



Afbeelding 2: websites waarnaar het meest gelinkt wordt

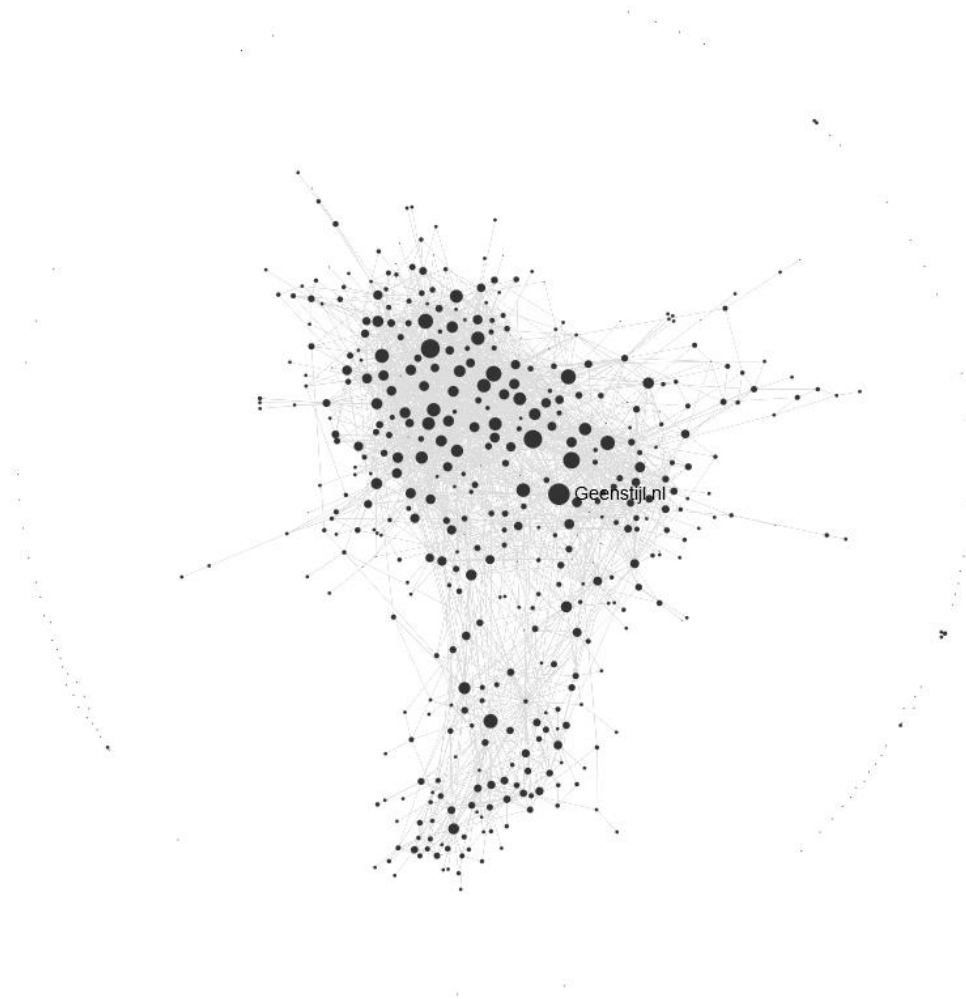
De meest gelinkte website in het hele corpus is YouTube, waar door maar liefst 226 websites verwezen wordt. Door op het pijltje naast de geciteerde website te klikken, is ook te vinden naar welke pagina's er op die website worden gelinkt (zie afbeelding 3).



Afbeelding 3: verschillende pagina's waarnaar wordt gelinkt

Zo is te zien dat er naar 5.665 verschillende pagina's op YouTube verwezen wordt. Wat voor de hele collectie opvalt is dat de meest geciteerde websites allereerst de bekende internationale online media zijn (YouTube en Wikipedia), gevolgd door de oude (Nederlandse) media die ook op papier verschijnen (*de Volkskrant*, *NRC*, *de Telegraaf*, *Trouw*, *Algemeen Dagblad*), de Nederlandse omroep (*NOS*) en de nieuwe media (NU.nl). Deze uitkomsten laten meteen de relevantie van de netwerkanalyses zien: dit zijn allemaal sites die niet gearchiveerd kunnen of mogen worden door de KB. Deze analyse laat dus zien dat we belangrijke en prominente sites missen in ons archief. Bij mogelijke verandering van procedures en wetgeving omtrent webarchivering kan dit aangedragen worden als argument. Tegelijkertijd kunnen deze websites door de netwerkanalyse toch tot op zekere hoogte onderdeel zijn van de collectie, omdat ze weliswaar niet zelf geharvest worden, maar wel genoemd worden en aangetoond kan worden dat ze in verband staan met de rest van de collectie.

Vervolgens kan er een netwerk gecreëerd worden van de websites in de collectie waarmee gevisualiseerd wordt hoe de websites zich zowel intern als extern tot elkaar verhouden op basis van gedeelde links. Hieronder is het interne netwerk gevisualiseerd. Daaruit blijkt dat de grootste website binnen de collectie, dat wil zeggen, de website waarnaar het meest gelinkt wordt door andere websites in de collectie, Geenstijl.nl is.



Afbeelding 4: netwerkvisualisatie intern van de gehele collectie

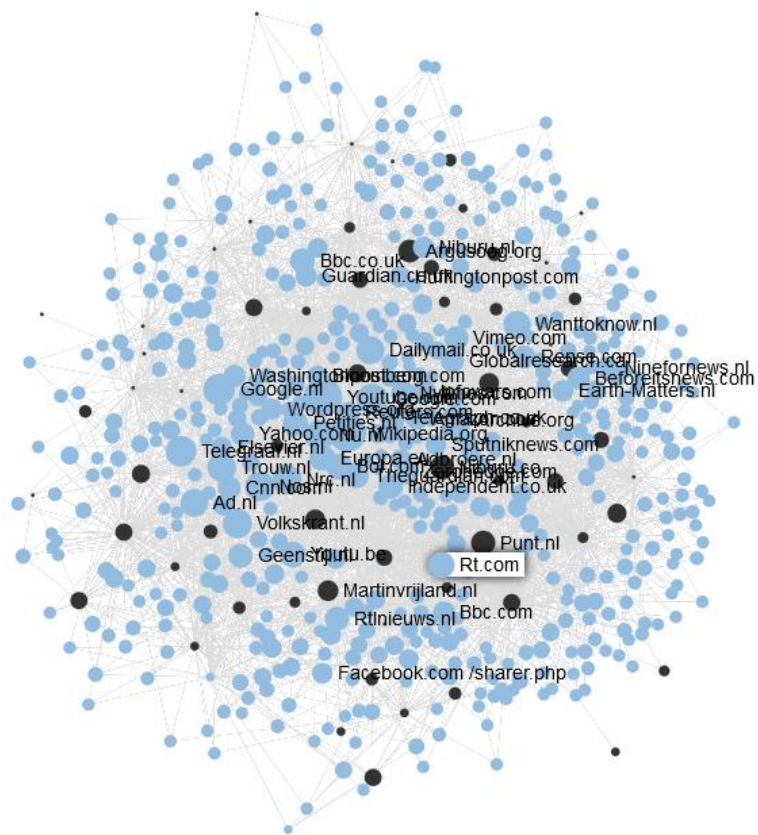
De losse stipjes aan de buitenrand van de visualisatie zijn websites die wat betreft interne links niet in verband staan met andere websites in de webcollectie, ofwel het zijn websites die inmiddels offline zijn.



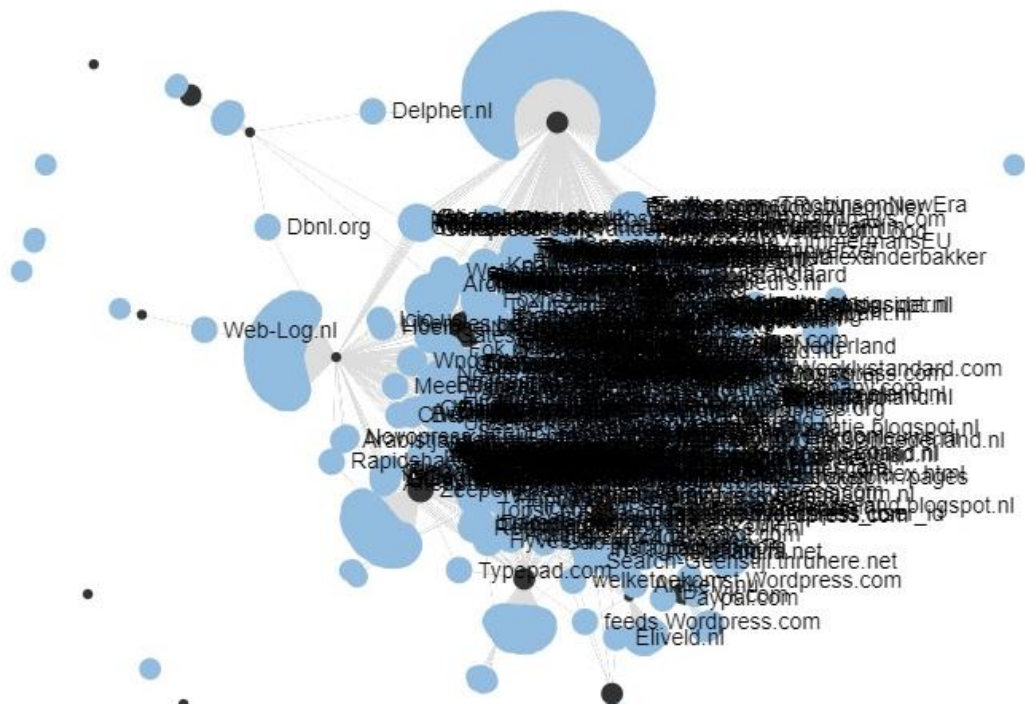
Afbeelding 5: netwerkvisualisatie extern van de gehele collectie

Hierboven is het netwerk zichtbaar met daarbij ook alle links naar externe websites die niet tot deze collectie behoren. De zwarte stippen zijn websites binnen de collectie, de blauwe stippen zijn websites buiten de collectie. In deze visualisatie is ook te zien dat er minder stippen aan de buitenrand van het netwerk zijn. Dat betekent dat sommige websites wellicht niet naar andere websites binnen de collectie linken, maar wel naar dezelfde externe webpagina's linken als andere websites in de collectie en op die manier dus toch in verband staan met andere sites uit de collectie.

Als er een netwerk gemaakt wordt van de subcategorieën in de collectie worden er weer andere dingen zichtbaar. Bij de complotwebsites springen er bijvoorbeeld enkele interessante links naar buiten in het oog. Complottheoretici linken naar meerdere Russische websites: Russia Insider (zie afbeelding 6), RT.com (voorheen Russia Today; zie afbeelding 7) en Pravda.ru (zie afbeelding 8).



Afbeelding 7: link naar RT.com



Afbeelding 9: links naar Delpher.nl en Dbnl.org

Vervolgens kan nagegaan worden naar welke specifieke pagina's er dan gelinkt wordt. Op DBNL wordt er gelinkt naar *Het nationaal-socialisme als rancuneleer* (1937) van Menno ter Braak.⁷¹ Op Delpher wordt er tweemaal naar artikelen over Sinterklaas en Zwarte Piet gelinkt. Eenmaal naar een strip uit het *Nieuwsblad van het Noorden* van zaterdag 30 november 1940 (zie afbeelding 10).

⁷¹ http://www.dbnl.org/tekst/braa002nati01_01/braa002nati01_01_0001.php

N HET NOORDEN VAN ZATERDAG 30 NOVEMBER 1940

HET ZWARTE PIET-PROBLEEM



„Ik zal je maar een strekje geven, Piet, anders
raak ik je in het donker kwijt”

Afbeelding 10: Nieuwsblad van het Noorden

De andere link verwijst naar een artikel in *De Telegraaf* van 22 november 1993: 'Sinterklaas heeft nu ook witte Pieten' (zie afbeelding 11).

Sinterklaas heeft nu ook witte Pieten



Jongen komt om

VERVOLG VAN PAGINA 1

De moeder reed. Volgens de politie raakte zij de macht over het stuur kwijt toen een van de kinderen haar aandacht probeerde te trekken door haar arm te pakken. De auto begon te slingeren, kwam in de rechter berm terecht en reed het talud af.

De kinderen, die geen veiligheidsgordel droegen, werden uit de auto geslingerd. Ze waren niet ernstig gewond. Vanaf de vluchtstrook probeerden ze het verkeer te waarschuwen. Toen dat niet lukte, begaf de oudste zich op de rijbaan. De ouders van de verongelukte jongen en zijn broertje zijn opgenomen in een ziekenhuis in Woerden.

Gladheid

Bij een ongeval op de autosnelweg A4 bij Rijswijk kwam omstreeks dezelfde tijd

een tienjarig meisje uit Rotterdam om toen de auto van haar ouders door gladheid van de weg raakte en over de kop sloeg.

Het meisje was met haar ouders, een tante en twee andere kinderen in de auto op weg naar huis. Bij Rijswijk vloog de auto uit de bocht. De auto schoot over de vangrail, klapte tegen een lichtmast en sloeg dertig meter verder over de kop. Er ontstond een enorme verkeerschaos op de A4 richting Rotterdam.

De snel invallende winter heeft het gehele weekeinde het verkeer ontregeld. Zaterdag werden veel automobilisten verrast door sneeuwbuien en ijzel, die in de loop van de dag van noord naar zuid trokken. De Algemene Verkeersdienst in Driebergen kreeg rond vijftig aanrijdingen te verwerken, waarbij ruim honderd voertuigen waren be-

● Sinterklaas is weer in het land. Dit jaar nam hij niet alleen zwarte Pieten mee uit Spanje, maar ook een aantal witte Pieten, die de belangrijke taak hebben gekregen voor het paard van de Sint te zorgen.

De kindervriend had het afgelopen zaterdag na zijn aankomst reuze druk met het bezoeken van tal van steden in Nederland, waar in totaal tienduizenden kinderen op de been waren om de Sint te verwelkomen. Een Amsterdams jongetje (foto) mocht even met bij de Goedheiligman op de schimmel plaats nemen.

Het bezoek van de goede Sint verliep niet overal vlekkeloos. In Boxmeer raakten twee kinderen en een man gewond toen het paard van Sinterklaas op hol sloeg. In Tiel werd een zwarte Piet beroofd van zijn fototoestel en tijdens de landelijke officiële intocht van de Goedheiligman in Ravenstein tuimelde een meisje in het ijskoude water van de jachthaven. Ze kon ongedeerd weer uit het water worden gehaald.

Foto: Richard Mouw

Abbeelding 11: link naar *De Telegraaf*

4.4 Toegang

De huidige Auteurswet biedt weinig ruimte voor vrije toegang tot de webcollectie verzameld in het KB-webarchief. Iedereen met een KB-jaarpas kan de webcollectie binnen de muren van de KB bestuderen (onsite), maar de content mag niet worden gekopieerd (mits voor privégebruik of studie) of openbaar worden gemaakt.

5. Conclusies

1. De reacties waarvoor enigszins gehuiverd werd, vielen alleszins mee. De meeste website-eigenaren waren juist opmerkelijk positief en blij met het archiveren van hun website. De bezwaren die er wel gemaakt werden, zijn niet typerend voor deze specifieke collectie. Op één uitzondering⁷² na maakte geen enkele website-eigenaar bezwaar vanuit zijn eigen ideologie of stond afwijzend tegenover het archiveren door een overheidsorganisatie.
2. Netwerkanalyse biedt inzichten in de webcollectie die niet op een andere manier te verkrijgen zijn. Zo viel bij handmatige bestudering van de websites op dat er opmerkelijk weinig over Zwarte Piet werd geschreven (ondanks het feit dat de discussie hieromtrent juist in deze periode in reguliere media weer vrij hoog oplaaide). De linkanalyse liet echter zien dat er via gedigitaliseerd materiaal wel degelijk aandacht aan Sinterklaas en zijn omstreken hulp werd besteed.
3. Bovendien kan een netwerkanalyse wellicht bijdragen aan een beter met elkaar samenhangende collectie. Er kan in kaart gebracht worden hoe websites met elkaar in verband staan, of juist niet. Tegelijkertijd is dit geen zaligmakend hulpmiddel: dat websites niet met elkaar in verband staan omdat ze geen onderlinge hyperlinks bevatten, betekent natuurlijk niet dat ze op geen enkele manier met elkaar in verband staan.

⁷² <http://vitalifeiten.nl>

6. Aanbevelingen

1. Informeer de (potentiële) gebruiker beter over wat zich in het webarchief bevindt, vooral inhoudelijk. Het bestuderen en beschrijven van iedere afzonderlijke website is zeer tijdrovend en zal wegens gebrek aan tijd en middelen niet voor iedere collectie haalbaar zijn. Maak dan beknopte omschrijvingen over de inhoud van de collectie en biedt daarnaast met behulp van simpele tools snel en visueel overzicht in de collectie met bijvoorbeeld wordclouds of domeinnaamanalyse. Dit is over het algemeen vrij simpel te doen en geeft in één oogopslag een beeld van de collectie.
2. Probeer met verschillende tools zo veel mogelijk content van de gearchiveerde websites beschikbaar te stellen. Dat biedt onderzoekers een schat aan materiaal dat wellicht op korte termijn niet meer verkrijgbaar is als websites offline gaan en biedt een heleboel nieuwe mogelijkheden voor het onderzoeken van de websites met behulp van digital humanities tools. Denk aan het scrapen van tekst, links en afbeeldingen van de sites. Het bijeen zoeken van zo'n corpus is voor onderzoekers vaak bijzonder tijdrovend, dit zou mooie hapklare corpora opleveren die klaar zijn voor analyse. Ook voor studenten biedt dit een goede handreiking om onderwezen te worden met digitale content en analysetools.
3. Raadpleeg voor het opbouwen van inhoudelijke deelcollecties experts die gespecialiseerd zijn in dit specifieke thema. Voor deze collectie zijn er bijvoorbeeld veel religieuze sites, waarvan het merendeel betrekking heeft op het christendom. Er is simpelweg te weinig kennis in huis om bijvoorbeeld ook maatschappijkritische islamitische websites in beeld te brengen. Ga in dit soort gevallen met een islamoloog in gesprek. Idem uiteraard voor andere specifieke thema's: is de kennis niet in huis, haal het ergens anders.