

Tijs Boesschen Hospers & Merijn Meijntz

‘Als websites er niet meer zijn, raak je ook een stuk van je cultuur kwijt’

Interview met Sophie Ham

Fotograaf: Jacqueline van der Kort



Sophie Ham is conservator Digitale Collecties bij de Koninklijke Bibliotheek (KB). Na haar opleiding Nederlandse taal- en letterkunde aan de Universiteit Utrecht is ze zich bij de KB gaan bezighouden met hoe (digitaal) erfgoed geselecteerd, bewaard, en beschikbaar gesteld moet worden. Haar focus bij het conserveren ligt op born-digital-publicaties, zoals websites, blogs en sociale media. Redacteuren Tijs Boesschen Hospers en Merijn Meijntz gingen met haar in gesprek over haar ervaringen als digitaal conservator, het belang van webarchivering, de (on)eindigheid van het bewaren van internet-publicaties en de toekomst van de bibliotheek.

Hoe ben je conservator Digitale Collecties bij de Koninklijke Bibliotheek (KB) geworden?

Ik ben ooit, al in 2003, bij de KB begonnen als stagiair. Als je bij de KB werkt, kom je met allerlei soorten bronnen in aanraking. Eerst waren dat vooral boeken, maar de KB was al snel bezig met digitalisering. Vanaf 2004 begonnen we met massadigitalisering, zoals het op grote schaal scannen van kranten – de basis van het latere Delpher (2013). Ook kwam er sinds het begin van dit millennium steeds meer *born-digital*-materiaal bij. Dat zijn digitale bronnen waar geen papieren equivalent van is, zoals websites. Het webarchiveringstraject is toen – in 2007 – begonnen, omdat de KB al snel zag dat die bronnen ook opgeslagen moeten worden. Als er een papieren versie van een publicatie is, is de dreiging van het verdwijnen ervan minder groot. Van een boek van bijvoorbeeld Arnon Grunberg zijn namelijk veel papieren kopieën in diverse bibliotheken. Ook als we de e-versie niet meer hebben, zou zo'n boek niet snel helemaal weg zijn. Een website die offline gaat, is daarentegen hooguit nog door de eigenaar opgeslagen, dus die is gewoon weg. Daar zit voor de KB een gevoel van urgentie.

Ik heb een tijdje aan de achterkant van een born-digitalstroom gewerkt als coördinator, met name van het webarchief. Toen de vacature van digitaal conservator vrijkwam, heb ik daar vanwege mijn affiniteit met de inhoud op gereageerd. Bovendien ben je als digitaal conservator niet alleen maar bezig met het kiezen van nieuwe stukken voor de collectie, maar krijg je ook steeds meer te maken met de technische en juridische randvoorwaarden. Kan en mag ik iets wel binnenhalen? Daar had ik door mijn voorgeschiedenis bij de digitaliseringstrajecten al ervaring mee. Ik had al veel met auteursrecht te maken gehad. Dat is bij digitale bronnen van groter belang dan wanneer je naar een veiling gaat en een boek koopt om in de kast te zetten. Als vroegmodern letterkundige heb ik me altijd vooral geïnteresseerd in de cultuurhistorische context van teksten: wat zeggen ze over de tijd waarin ze zijn geschreven? Voor mij zijn websites dé cultuurdragers van onze tijd en daarom verdienen ze het om goed bewaard en bestudeerd te worden.

Voor mij zijn websites dé cultuurdragers van onze tijd en daarom verdienen ze het om goed bewaard en bestudeerd te worden

Waarom vind je webarchivering interessant?

Ik denk dat mensen over twintig jaar terugkijken en denken: jeetje, wat er toen allemaal gebeurde... We zien nu al in dat het heel erg jammer is dat we pas in 2007 begonnen zijn met het archiveren van het web. Toen het internet begon hadden mensen niet zo door dat het zo revolutionair zou zijn en ons leven in grote mate zou veranderen. Nu zien we dat veel mensen geïnteresseerd zijn in hoe het internet eigenlijk begonnen is.

Wat ik extra interessant vind aan het internet is dat mensen sinds de introductie hiervan voor het eerst zonder tussenkomst van een uitgever konden gaan publiceren. Mensen kunnen zelf bepalen wat ze met de wereld willen delen zonder dat iemand

zegt: 'Dit is goed, dit publiceer ik wel, maar dit niet.' Als je wilt volgen hoe bubbels en polarisering werken, welke ideeën daar eigenlijk achter liggen, dan heb je webpagina's nodig. Hoe mensen over bepaalde onderwerpen denken, kun je allemaal terugvinden op webpagina's die niet per se van de NOS of een uitgever zijn, maar van mensen of groeperingen die zelf iets over dat onderwerp willen zeggen. Ik denk dat dat veel zegt over onze tijd.

Wat doet een digitaal conservator precies? Hoe verschilt dat van het werk van een 'traditionele' conservator?

Een deel van mijn taken gaat om het zoeken van stukken voor de collectie. Daarnaast moet ik goed weten wat je met de collectie kan en begeleid ik onderzoekers die er gebruik van willen maken. Dat lijkt nog best een beetje op wat mijn collega's die verantwoordelijk zijn voor middeleeuwse handschriften doen.

De technische en juridische factoren maken het echter anders. Je kan wel bedenken dat je heel Wikipedia wilt binnenhalen, maar je moet ook bedenken hoe je dat gaat doen. Kan de software dat wel aan? Kan je het goed opslaan? Bovendien gelden voor webarchivering andere regels dan voor boeken. Het kan zijn dat de website van een museum rechten heeft om bepaalde plaatjes op hun website te tonen. Maar volgens de Nederlandse wet is het zo dat als de KB iets opslaat, dat eigenlijk een kopie is en voor die kopie moet je voorafgaand toestemming hebben. Dat werkt dus heel anders dan met papier. Een deel van mijn werk bestaat nu ook uit het voeren van gesprekken met het ministerie van Onderwijs, Cultuur en Wetenschap over modernisering van de wetgeving. We zien namelijk dat al die wetten en onze taakstelling erg uitgaan van het papieren landschap en niet van het digitale.

Als we een website willen opnemen, sturen we vooraf een bericht dat we de website willen opnemen in het archief en dat contactpersonen binnen vier weken moet reageren als ze het daar niet mee eens zijn. We krijgen wel eens een 'nee'. Dat is bijvoorbeeld veel gebeurd in het proces rondom de collectie maatschappijkritische websites, waarin we heel veel websites hebben opgenomen van mensen die wat kritischer zijn ten opzichte van de overheid en ook de KB zien als een overheidsachtige organisatie. Daar willen ze geen onderdeel van zijn.

We moeten die ingewikkelde procedure volgen en dat kost veel administratie. Daarnaast hebben we te maken met beperkte capaciteit en zijn er technische randvoorwaarden waardoor we bewust moeten kiezen welke internetbronnen we archiveren. Daar doen we ook kwaliteitscontrole op. De rest hebben we niet, dus heel veel van wat er nu op het internet is, gaat verloren.

Er staat heel veel op het internet. Waar ligt de grens van wat we wel en niet willen bewaren? Is alles de moeite waard om te bewaren?

Daar kunnen we binnen de KB echt discussie over krijgen. Sommige mensen willen het liefst alles bewaren, elke dag, de hele tijd. Maar we moeten keuzes maken. Wat we vooral proberen, is nagaan of iets in het verlengde van onze collectie ligt. Websites die

echt 'publicaties' zijn krijgen bijvoorbeeld voorrang ten opzichte van iets wat toevallig óók op het internet staat maar eigenlijk geen publicatie is, zoals webwinkels. En dan kom je meteen op de vraag: wat is een publicatie? Wij zien iets als een publicatie wanneer het geopenbaard is aan een publiek. Steeds meer vinden we dat er ook een intentie moet zijn om te publiceren. Sociale media zijn daarin een twijfelgeval. We zagen Twitter lang als een soort microblog, maar inmiddels denken we: niet iedereen die iets op Twitter zet, beschouwt dat als een publicatie. Terwijl je van bijvoorbeeld Geert Wilders of Sander Schimmelpenninck die intentie wél kan verwachten.

*Sommige mensen willen
het liefst álles bewaren,
elke dag, de hele tijd. Maar
we moeten keuzes maken*

Het blijft zoeken en afwegen. Je wilt ook iets van de gewone internetcultuur bewaren. Maar sommige dingen horen meer bij andere instellingen, zoals Beeld en Geluid voor video, of het Nieuwe Instituut voor digitale kunst. Onze taak ligt bij tekst: boeken, kranten, tijdschriften. Dingen die daar het meest op lijken, krijgen prioriteit. Je weet nooit wat over honderd jaar belangrijk blijkt. Kranten werden ooit bewaard voor het nieuws, maar nu doen mensen onderzoek naar familieberichten en advertenties. Met het internet zal dat net zo gaan. Je wilt niet álles opslaan, maar je wilt ook niet achteraf denken: waarom hebben we dát niet bewaard? Je wilt niet alle kattenfilmpjes voor de eeuwigheid behouden, maar misschien toch wel een paar.

Er moet dus geselecteerd worden. Hoe kies je wat je wel of niet gaat archiveren?

We zijn op een gegeven moment gaan kijken wat de meest bezochte websites in Nederland zijn. Oftewel, welke websites zijn belangrijk voor mensen? Daar zitten wel allemaal beperkingen aan. Buienradar is bijvoorbeeld een van de meest gebruikte websites, maar live satellietbeelden zijn niet zo archiveerbaar. Hetzelfde geldt voor de website van NS. Je kunt nog wel iets zien van oude versies van de NS-website, maar de functies zijn niet goed vast te leggen. We zijn die populaire websites best structureel afgegaan en hebben ook gekeken naar welke andere sites die verwezen. Die informatiebronnen konden dan ook weer vastgelegd worden. Zo kom je websites op het spoor die je zelf niet kent, maar die wel bij zo'n onderwerp horen.

Mijn voorganger, Kees Teszelszky, heeft later het idee van de speciale collecties uitgewerkt. Dat is een heel mooi idee, waarbij je een onderwerp kiest om vervolgens zo breed mogelijk te kijken welke websites bij dat thema aansluiten. Dat doe je dan zo representatief mogelijk, dus je selecteert een heel mainstream website en een niche, een heel professionele site en een uit de hobbysfeer, eentje aan de rechterkant en eentje aan de linkerkant.

Zo proberen we een soort evenwicht te krijgen. Vaak hangen keuzes bij de opbouw van collecties ook af van de interesse van een stagiair of collectiespecialist of van wat we kunnen laten aansluiten op iets anders wat de KB aan het doen is. We proberen

daarom in collectiebeschrijvingen heel goed uit te leggen welke keuzes er zijn gemaakt. Daarnaast betrekken we, als het lukt, ook experts van buiten. Iemand met een ander perspectief kan dan weer blinde vlekken ontdekken.

We kregen vroeger van onze ouders te horen dat alles wat we plaatsten voor altijd op het internet zou staan. Als dat waar zou zijn, zou het waarschijnlijk niet interessant zijn om internetpagina's te archiveren. Hoe zit dat?

Aan de ene kant kan er een kopie gemaakt worden van iets wat jij als privépersoon schrijft. Die kan dan eeuwig blijven rondcirculeren, ook al denk je dat je het zelf hebt weggehaald. Aan de andere kant zien we dat de gemiddelde webpagina na negentig dagen weg of veranderd is, dus dan is de informatie niet meer hetzelfde. We hebben bijvoorbeeld de website Thuisarts gearchiveerd, wat een van mijn favoriete websites is. Daar kijken namelijk veel mensen naar, bijvoorbeeld omdat ze een kwaaltje hebben en willen weten wat ze moeten doen. De adviezen die er nu op staan zijn heel anders dan tien jaar geleden. Daardoor kun je zien wat een gemiddeld iemand bijvoorbeeld wist over griep; het grieplemma van tien jaar geleden rept niet over corona, maar nu natuurlijk wel. Oude versies van dat soort sites zijn er vaak niet meer, terwijl veel mensen daar gebruik van maakten. Die veranderingen zijn heel interessant, al maak je nog altijd momentopnames omdat je niet alles continu kunt vastleggen. Als oude versies van veelgebruikte websites er niet meer zijn, raak je een stuk van je cultuur kwijt. Om diezelfde reden staan de oude edities van de Medische Encyclopedie hier ook allemaal keurig in het magazijn.

Als oude versies van veelgebruikte websites er niet meer zijn, raak je een stuk van je cultuur kwijt

Is er een concrete publicatie waarvan je vindt dat we die moeten archiveren maar waarbij het helaas nog niet is gelukt?

Ja, Wikipedia! Die is te groot, en dat vind ik echt heel erg. We willen hem heel graag archiveren, maar dat is technisch erg lastig. Met de standaard manier van binnenhalen, via *crawling*, moeten we heel precies instellen wat we wel en niet meenemen. Je wilt alleen de Nederlandstalige Wikipedia, maar niet *Wikimedia Commons*, want daar zitten alle media in en die zijn gigantisch. We zoeken nog naar de juiste instellingen om Wikipedia goed binnen te halen zonder dat het krankzinnig groot wordt. We maken daar wel stappen in, maar het had eigenlijk al gebeurd moeten zijn.

Wikipedia staat hoog op mijn verlanglijstje omdat die site zo enorm belangrijk is. Het is dé plek waar mensen – jongeren, volwassenen, iedereen – even snel iets opzoeken. Er komen steeds meer automatisch gegenereerde lemma's bij. Hierdoor weten we straks niet meer zeker of iets door een mens is geschreven. Daarom vind ik het belangrijk om nu een snapshot vast te leggen voordat er te veel is veranderd. Bovendien past Wikipedia perfect bij een bibliotheek. We bewaren alle encyclopedieën sinds ze bestaan, en Wikipedia is daar een logische opvolger van. Het laat zien wat we weten en

wat we belangrijk vinden. De revisiegeschiedenis is ook fascinerend: welke lemma's komen erbij, welke verdwijnen, en wat is al tien jaar niet aangepast? We kijken nu: kunnen we dat ook meenemen? Of maken we gewoon elk jaar een nieuwe versie? Als ik dus één frustrerend voorbeeld moet noemen, dan is het zeker Wikipedia.¹

Wat vind jij het topstuk van de digitale collectie van de KB?

Een van onze topstukken is De Digitale Stad (DDS), een platform uit 1995 dat echt pionierswerk verrichtte op het internet. Destijds had bijna niemand, slechts een paar duizend mensen, in Nederland toegang tot het internet. DDS, gestart met subsidie van de gemeente Amsterdam, gebruikte de stad als metafoor. Je kon er een huisje bouwen, er waren pleinen zoals een Filmplein of een Homoplein waar mensen afspraken konden maken. Ook instituties zoals de bibliotheek maakten een digitale versie van zichzelf. Het was echt een manier om het internet te verkennen voordat zoekmachines bestonden. DDS had sociale functies die later pas gemeengoed werden, zoals chatrooms en online cafés. Mensen konden daar informatie opzoeken en met elkaar praten over bijvoorbeeld sport of televisie. Het was een hele nieuwe manier van een gemeenschap vormen.

In 1996 hebben ze een 'freeze' gemaakt, dat wil zeggen dat ze een snapshot bewaard hebben op harde schijven. Dat bleek later essentieel, omdat die stad echt verdwenen is. DDS is vorig jaar aan ons overgedragen, wat echt een aanwinst is. Het geeft een uniek inzicht in hoe mensen internet gebruikten vóór het bestaan van zoekmachines. Destijds moest iedereen, zonder een omgeving als DDS, een precies webadres kennen; navigeren was veel minder vanzelfsprekend. Met de opkomst van zoekmachines zoals Google werd DDS minder relevant, maar het blijft fascinerend om te zien hoe internetgebruik zich ontwikkelde.

Daarnaast hebben we via XS4ALL ook veel oude webpagina's kunnen archiveren, vaak van gewone mensen die voor het eerst online gingen. Ik vind het persoonlijk erg interessant hoe mensen het internet toen nog als een heel andere ruimte zagen, vol experimenten en mogelijkheden.

Mensen hebben nu nog een heel fysiek idee bij een bibliotheek. Gaat dat in de komende jaren veranderen? Wordt een bibliotheek eerder een database?

We hebben lang gedacht dat de bibliotheek volledig zou digitaliseren, maar dat blijkt genuanceerder. Ik denk dat bibliotheken zowel fysiek als digitaal blijven bestaan. Er is steeds meer

We hebben lang gedacht dat de bibliotheek volledig zou digitaliseren, maar dat blijkt genuanceerder

¹ Inmiddels is het de KB gelukkig deels gelukt om Wikipedia te archiveren (zie: 'KB neemt Nederlandstalige Wikipedia op in archief', Wikimedia Nederland, 16-06-2025, <<https://www.wikimedia.nl/actueel/persberichten/kb-neemt-nederlandstalige-wikipedia-op-in-archief/>>).

behoefte aan digitale detox: even weg van het scherm, gewoon een fysiek boek in je handen hebben. Tegelijkertijd zie je dat nieuws en communicatie wél steeds meer digitaal worden. De vraag is dan: hoe ziet die digitale ruimte eruit? Is dat volledig in handen van techgiganten, of hebben we nog eigen, veilige plekken? Daar ligt een rol voor bibliotheken.

Interessant is dat boeken juist weer terug in de bibliotheek komen. Waar veel bibliotheken eerst dachten dat ze alleen nog maar touchscreens nodig hadden, merk je nu dat mensen het fysieke boek missen. Zelf heb ik veel liefde voor zowel het digitale als het fysieke. Een middeleeuws handschrift vasthouden geeft een historische sensatie die je digitaal niet zomaar nabootst. Bovendien denk ik dat je als je een oude website op een oude computer bekijkt, een beetje datzelfde gevoel kunt krijgen: oh ja, zo was het internet in de jaren negentig.

We hebben een lezerspubliek dat bestaat uit veel onderzoekers en toekomstige onderzoekers. Wat kunnen zij voor jullie betekenen?

Wat we vooral nodig hebben, zijn onderzoekers die het webarchief willen gebruiken. Er zit zoveel in waarvan we zelf niet eens precies weten wat er allemaal te vinden is. Mijn oproep zou daarom zijn: kom hierheen, maak gebruik van de collectie en laat weten wat werkt en wat mist. Het is niet perfect, maar het bevat bronnen die je niet zomaar ergens anders vindt. En als er dingen beter kunnen, willen we dat graag horen.