

Inleiding

HET DIGITALE VERLEDEN EN DE TOEKOMST VAN DE MEDIAGESCHIEDENIS

‘Digitaal erfgoed is doolhof’, kopte *NRC Handelsblad* in september 2011 op de voorpagina van haar wetenschapskatern. De wat mistroostige onderkop ‘Wanhopig zoeken in digitale bronnen’ werd nog eens bevestigd door de kop ‘Het digitale drama’ boven het eigenlijke artikel. Hierin werd beweerd dat, ondanks de naar schatting honderd tot tweehonderd miljoen euro die sinds 2004 is uitgegeven aan digitalisering, wetenschappers door ongecoördineerd en slordig digitaliseren ‘tallose bronnen’ niet kunnen vinden of doorzoeken. ‘Waar bètawetenschappers met een sportauto over de digitale snelweg razen, hobbelen beoefenaars van de humaniora met een brik door een doolhof van websites en databanken.’¹

In het afgelopen decennium zijn archieven, bibliotheken en andere erfgoedinstellingen op nationaal, regionaal en lokaal niveau materiaal gaan digitaliseren. Een goed en toegankelijk centraal overzicht, waarin wordt bijgehouden welk materiaal al onder de scanner is doorgehaald, ontbreekt echter. Het gevolg is dat gebruikers nauwelijks weten welke bronnen beschikbaar zijn en waar zij terecht kunnen. Potentieel blijft dus veel materiaal ongebruikt. Bovendien zijn er ongeveer net zo veel standaarden en specificaties voor digitalisering gehanteerd als er projecten zijn. Daardoor wordt het – ook in de toekomst – onmogelijk om bestanden te koppelen en als één geheel te onderzoeken. Wie bijvoorbeeld de berichtgeving in kranten wil raadplegen, moet met dezelfde zoekvraag langs tientallen archieven en portals, terwijl het maar zeer de vraag is of de uitkomsten – methodologisch gezien – vergelijkbaar zijn, door de verschillende opzet van databases en afwijkende zoekalgoritmes.

De aard van gedigitaliseerd materiaal is zeer divers. Er is archiefmateriaal gescand via individuele archiefinstellingen die het aanbieden op hun eigen website. Hetzelfde geldt voor fotoarchieven. Boeken en tijdschriften zijn beschikbaar via de Digitale Bibliotheek der Nederlandse Letteren (DBNL), via Early Dutch Books Online en in toenemende mate ook via Google Books dat in samenwerking met de Koninklijke Bibliotheek (KB) miljoenen Nederlandse boeken wil digitaliseren. Kranten zijn inmiddels gedigitaliseerd door talrijke bibliotheken en archieven, mede omdat ze zo’n mooie bron zijn voor (regionaal en lokaal) historisch onderzoek. Ambitieuze is de database die de KB, met steun van NWO, heeft gebouwd en waarin een grote steekproef van Nederlandse kranten vanaf 1614 tot nu beschikbaar wordt gemaakt. Het Nederlands Instituut voor Beeld en Geluid doet veel aan het digitaliseren van

* Marcel Broersma coördineerde dit themanummer.

audiovisueel materiaal en participeert ook in diverse Europese projecten en portals (waaronder EUScreen) om het beschikbare materiaal te contextualiseren. Met onder meer het EYE Film Instituut Nederland neemt het tevens deel aan het project Beelden voor de Toekomst waarin ook film beschikbaar wordt gesteld. Deze opsomming heeft niet de minste pretentie van volledigheid – bij zoveel activiteit past bescheidenheid. Digitaliseren lijkt welhaast het nieuwe toverwoord in het land van de bibliothecarissen en archivarissen.

De doelen van deze projecten lopen uiteen. Soms gaat het primair om conservering van kwetsbaar materiaal, soms staat *in-depth* beschikbaarstelling van materiaal voorop, in andere gevallen lijkt ook de wens *traffic* te genereren naar de eigen website mee te spelen. Het uitgangspunt voor digitalisering blijkt wezenlijk voor de uitvoering en daarmee het nut en de bruikbaarheid van projecten voor onderzoekers. In sommige gevallen lijkt de horizon niet verder te reiken dan het digitaliseren zelf; over het gebruik is nauwelijks nagedacht. Bij andere projecten wordt in nauw overleg met onderzoekers en andere gebruikers gekeken naar de wensen en mogelijkheden voor onderzoek. De kwaliteit van de beschikbare databases is dan ook wisselend. Scannen is relatief eenvoudig, maar het toevoegen van metadata en het toegankelijk maken van teksten via *optimal character recognition* (OCR) is een stuk lastiger. Wanneer dit niet of niet goed gebeurt, ligt er weliswaar een prachtige stapel plaatjes, maar een die slecht – en soms zelfs helemaal niet – doorzoekbaar is. Daardoor zijn deze bestanden nauwelijks bruikbaar voor onderzoekers.

Digitalisering lijkt op het eerste gezicht dus een moeras vol problemen waarbij tegengestelde belangen en onafgestemde ambities de mogelijkheden voor gebruikers reduceren. Gelukkig is er ook goed nieuws. Hetzelfde artikel in *NRC Handelsblad* meldt dat het onderzoek naar de entree van de Romantiek in Nederland, dat de Amsterdamse neerlandicus Wim van den Berg meer dan vijftien jaar kostte in de jaren zestig en zeventig van de vorige eeuw, nu in één avond kan worden verricht. Waar Van den Berg stapels boeken en artikelen doorspitte op zoek naar het woord “romantisch” is dat in de Digitale Bibliotheek der Nederlandse Letteren (DBNL), een database met gedigitaliseerde boeken, met één druk op de knop te achterhalen. Volgens de Utrechtse hoogleraar Wijnand Mijnhardt een ‘fantastisch voorbeeld’ van de mogelijkheden die digitalisering biedt. ‘Iedereen kan het experiment herhalen, het is controleerbaar, het is objectief. Het is een kwantitatieve ondersteuning van kwalitatief onderzoek.’²

Van den Berg zelf was minder gelukkig met het gekozen ‘exempel van dolgedraaide wetenschappelijke idiotie en wereldvreemdheid’, waarin hij in zijn perceptie figureerde als een ‘enigszins meelijwekkende figuur, zo niet lachwekkende figuur’. Hij wees er op dat zijn onderzoek veel breder was en ook in een theoretische en literatuurhistorische context was geplaatst.³ Van den Berg had natuurlijk gelijk. Met één enkele zoekvraag laat zijn onderzoek zich natuurlijk niet repliceren. Het laat echter wel zien dat het doorzoeken van grote hoeveelheden digitaal materiaal nieuw licht kan werpen op ‘oude’ vragen. In het lopende debat over de Romantiek in

Nederland bevestigde de queeste in de DBNL Van den Berg's stelling. Maar, zoals René van Stipriaan die het onderzoekje deed, stelt:

‘Veel inventariserend werk dat in de voorbije decennia door neerlandici is verricht zal de komende jaren met behulp van zoekmachines overgedaan kunnen worden, en ik durf de voorspelling wel aan dat niet veel studies zo mooi overeind zullen blijven als die van Wim van den Berg. Het wordt een slagveld, en veel *idées reçues* uit ons vak zullen sneuvelen. De geesteswetenschappen zullen er wat exacter door worden, of misschien is het reëler om te zeggen, dat ze nog minder exact blijken te zijn dan we al dachten.’⁴

De schat aan materiaal die inmiddels in gedigitaliseerde vorm op onderzoekers ligt te wachten, biedt geweldige mogelijkheden voor onderzoek.

In een andere ingezonden brief in *NRC Handelsblad* stelde de Utrechtse historicus en alfa-informaticus José de Kruif dat geesteswetenschappers ondertussen nog slecht of zelfs helemaal niet zijn toegerust voor deze digitalisering van bronnenmateriaal. Het is volgens haar een *idée-fixe* te veronderstellen dat alle aanbieders, nationaal en internationaal, zich zullen aanpassen aan één standaard en zich zullen schikken naar de wensen van gebruikers. ‘Deze onderzoekers zullen andersom moeten werken, zij zullen zich moeten aanpassen aan de nieuwe vormen waarin digitaal bronnenmateriaal zich nu eenmaal aandient. Die aanpassing zal de geesteswetenschappen meer vooruit helpen dan welke standaard ook.’ De Kruif wijst erop dat software voor *data-* en *textmining* al ruimschoots beschikbaar is. Hiermee kan ongeordende digitale informatie worden ontleed, waarna er een statistische analyse op kan worden losgelaten.⁵ Historici blijken echter huiverig voor computers en afkerig van alles wat riekt naar cijfers, statistiek en kwantitatief onderzoek. Volgens Mijnhardt raakten de humaniora rond 1900 ‘onder de invloed van Duitse filosofen met hun “begrijpend observeren” en lieten zo kansen liggen’. Optimistisch vervolgt hij dat digitalisering ‘de kloof tussen alfa- en bètawetenschappen’ kan helpen dichten.⁶ Dat lijkt een wensdroom waarvan de verwezenlijking nog ver weg is – en die naar ik vrees voor veel geesteswetenschappers ook ver weg zal moeten blijven –, maar dat digitalisering invloed zal hebben op de methodologie en technieken waarvan onderzoek zich bedient, lijkt buiten kijf te staan.

De bijdragen in dit themanummer cirkelen rond de vraag welke mogelijkheden digitalisering van bronnenmateriaal de mediageschiedenis biedt en voor welke methodologische problemen en uitdagingen dit haar stelt. Door het omzetten van analoge bronnen in digitale ontstaan feitelijk nieuwe bronnen die qua aard en functie fundamenteel verschillen van het ‘origineel’. Er is hierbij dus veeleer sprake van transformatie dan van migratie. Dit betekent dat oude onderzoeksstrategieën aan revisie onderhevig moeten zijn om op zinvolle wijze de nieuw ontstane archieven te exploreren.

Sonja de Leeuw beziet in haar bijdrage televisiearchieven als netwerk van connecties. Wat voorheen een fysieke plaats was waar audiovisueel materiaal werd bewaard, verandert nu in een reeks virtuele platforms waarop allerlei actoren, of het nu curatoren, onderzoekers, televisiemakers of ‘gewone’ kijkers zijn, betekenis toevoegen aan het beschikbare bronnenmateriaal. Daardoor ontstaat een nieuw ‘cultureel geheugen’ dat medieert tussen geschiedenis en herinnering. In Europese projecten wordt daaraan de dimensie van transnationaliteit toegevoegd. Materiaal dat een vluchtig karakter had, want vaak slechts eenmaal werd uitgezonden en aan nationale contexten gebonden was, is nu structureel beschikbaar voor een internationaal publiek. De Leeuw laat zien dat ook hier beperkingen zijn waar het gaat om beschikbaarheid en standaardisatie van metadata, maar wijst tegelijkertijd op de mogelijkheden die zich nu aandienen voor een werkelijk transnationale geschiedenis van Europese televisie.

In een bijdrage over digitale krantenarchieven constateer ikzelf dat deze nieuwe bronnen de mogelijkheid bieden om de inhoud van kranten systematisch te bestuderen. Nog steeds is onderzoek naar de berichtgeving een bijkans witte vlek in de historiografie. De krant zelf is nauwelijks object van studie; zij is vaak slechts een bron voor het smakelijke citaat en het illustratieve voorbeeld in (pers-)historische analyses. Met de overdaad aan materiaal die de krant als seriële bron biedt, weet de geschiedwetenschap zich echter slecht raad. Historici die onderzoek doen in digitale krantenarchieven moeten zich realiseren dat zij te maken hebben met een nieuwe bron met een eigen logica. Het is daarom van groot belang dat de onderzoeker weet heeft van de methodologische valkuilen die zich vooral centreren rond het zoeken op trefwoorden. Niettemin opent zich, zo betoog ik, nu een nieuwe scala aan mogelijkheden voor onderzoek naar journalistiek als proces van betekenisgeving.

Edwin Klijn, die als projectleider nauw betrokken was bij het digitaliseren van kranten door de KB, schetst in zijn artikel welke technische problemen optreden bij massadigitalisering. De kwaliteit van het scannen, de metadata, de OCR-software en de zoekmachine bepalen hoe goed een collectie doorzoekbaar is en hoe bruikbaar zij daarom is voor wetenschappelijk onderzoek. De financiële mogelijkheden en juridische beperkingen (auteursrechten) bepalen daarbij de grenzen waarbinnen erfgoedinstellingen moeten werken. Die zullen zich volgens Klijn in toenemende mate moeten gaan toeleggen op contextualisering van het materiaal om zo hun traditionele taak als hoeders van het culturele kapitaal opnieuw inhoud te geven.

Het gebruik van digitale bronnencollecties staat centraal in de bijdrage van Olivier Nyirubugara. Aan de hand van empirisch onderzoek op twee middelbare scholen beschrijft hij hoe scholieren dit materiaal integreren in hun opdrachten voor het vak geschiedenis. Ten grondslag aan het digitaliseringsbeleid van de overheid en erfgoedinstellingen lagen verwachtingen over het cognitieve nut van dit type bronnen in het onderwijs, maar in toenemende mate ook ideologische overwegingen: het

behoud van nationale identiteit werd steeds vaker gebruikt als argument voor digitalisering. Uit het onderzoek van Nyirubugara blijkt echter dat de zorgvuldig samengestelde collecties van erfgoedinstellingen het afleggen tegen allerlei andere online bronnen van informatie die geen ‘kwaliteitsstempel’ hebben. Dat komt vooral doordat zij slecht vindbaar zijn via zoekmachines. Ook zien zij samenwerking met andere instellingen en *linken* naar websites van ‘concurrenten’ – wat hun *ranking* bij bijvoorbeeld Google zou verbeteren – als bedreiging voor hun eigen identiteit en positie.

Martijn Kleppe laat zien welke mogelijkheden en problemen optreden wanneer onderzoekers zelf een database met gedigitaliseerd materiaal aanleggen; in zijn geval een bestand met foto’s die zijn gebruikt in geschiedenismethodes voor het middelbaar onderwijs. Hij gaat daarbij voornamelijk in op de methodologische vragen die optreden bij het verrijken van foto’s met metadata. Wat staat eigenlijk op de afbeelding? Dezelfde foto heeft meerdere betekenissen en wordt anders geïnterpreteerd vanuit het doel van het onderzoek. Dat leidt volgens Kleppe tot de bijna onvermijdelijke conclusie dat het zinloos is te zoeken naar één systeem voor beschrijvingen. Zolang woorden nodig zijn om foto’s te beschrijven, blijft een subjectief element aanwezig. Wellicht dat beeldherkenning in de toekomst dit probleem kan oplossen.

Ten slotte onderzoekt Julia Noordegraaf aan de hand van recente experimenten met *social tagging* hoe in de beschrijving van archiefmateriaal een nieuwe vorm van *participatory knowledge* wordt geïntroduceerd. Door het publiek metadata aan televisiemateriaal te laten toevoegen, verandert de aard van het archief. Collectiebeschrijving had het karakter van een statische, bijna encyclopedische omschrijving van vaststaande inhoud, maar wordt nu dynamisch en interactief. ‘The classical, file-oriented archival practices yield “to the use-oriented” (“to be completed”) *dynarchive*’, zoals mediatheoreticus Wolfgang Ernst het omschrijft. In de praktijk blijkt *social tagging* de toegankelijkheid vanarchieven te vergroten, maar leidt (nog) niet tot een werkelijke democratisering – tot ‘knowledge by consensus’, constateert Noordegraaf. Archieven houden nog sterk vast aan de traditionele vormen van ‘authoritative knowledge production’.

De artikelen in dit themanummer laten zien dat naarmate steeds meer archiefmateriaal wordt gedigitaliseerd, ook de mediageschiedenis zal moeten veranderen. Het archief als fysieke ruimte waar de onderzoeker als individu een uniek object bestudeert dat fysiek en temporeel begrensd is, maakt plaats voor een virtueel platform waarin meerdere gebruikers hetzelfde object kunnen bekijken en interactief kunnen manipuleren. Dit theoretisch inzicht heeft vergaande praktische implicaties. Digitalisering provoceert het ontwikkelen van nieuwe methodes en onderzoekstechnieken evenals nieuwe vormen van presentatie van wetenschappelijke analyses. Een nieuw veld opent zich met onderzoeksvragen die eerder wel gesteld, maar niet of nauwe-

lijks beantwoord konden worden en met nieuwe vragen die voorheen niet aan de orde waren. Deze nieuwe wegen voor de mediageschiedenis bevinden zich op het niveau van de inhoud van media. Dat lijkt erg voor de hand liggend, maar is het toch niet. Mediageschiedenis heeft zich tot nu toe in meerderheid gericht op de bestudering van de institutionele context en de productie van media. De berichtgeving in kranten en tijdschriften, de inhoud van radio- en televisieprogramma's of de beeldtaal op foto's zijn in veel mindere mate object van onderzoek geweest. Een zeer wenselijk effect van digitalisering zou zijn dat het mediahistorisch onderzoek zich nu ook hierop zal gaan richten.

Noten

1 *NRC Handelsblad*, 10/11-9-2011.

2 *Idem*.

3 *NRC Handelsblad*, 8/9-10-2011.

4 René van Stipriaan, 'Zoeken naar romantiek. Geavanceerd onderzoeksinstrumentarium in de DBNL', *TS. Tijdschrift voor Tijdschriftstudies* nr. 27 (2010), 5-17, aldaar 13.

5 *NRC Handelsblad*, 17-9-2011.

6 *NRC Handelsblad*, 10/11-9-2011.