

Digitale deemstering

Auteursrecht en de digitalisering van boeken in Nederland en Vlaanderen

Digitalisering heeft veel traditionele economieën de markt ontnomen, maar de boekenwereld houdt moedig stand. Waarom is er met het boek niet gebeurd wat er wel gebeurde met muziek, film en fotografie? Filosoof en theaterwetenschapper Thomas Crombez duikt in de wereld van het digitale boek en laat zien waarom de materiële vorm (nog) niet ingehaald is. Hoe dan ook zitten de digitale geesteswetenschappen sinds tien jaar in een stroomversnelling. Digitalisering biedt onderzoekers namelijk veel voordelen, stelt Crombez, maar de hachelijke situatie omtrent auteursrechten gooit vaak roet in het eten. Het is daarom tijd dat onze bibliotheken, musea en archieven de mouwen opstropen en ervoor zorgen dat alle digitale teksten weer tot de digitale lezers zullen behoren.

Wat moeten we aanvangen met miljoenen gedigitaliseerde boeken? Niet bijster veel, zul je voorlopig vaststellen. Het is inmiddels twaalf jaar geleden dat de Kindle Store door Amazon gelanceerd werd, het begin van de wereldwijde opmars van het e-book. Ook het massadigitaliseringsproject Google Books ging rond dezelfde tijd, in 2004, van start. Hoewel ondertussen miljoenen boeken als e-book verkocht worden, is het papieren boek niet dezelfde weg als de dinosaurus opgegaan. Integendeel, want er worden meer boeken gedrukt en verkocht dan ooit tevoren.

Het digitale boek heeft de boekenwereld niet op z'n kop gezet zoals de digitale film of het digitale muziekbestand dat in hun sector wel deden. Boeken werden niet 'opgegeten door software'. Terwijl dit toch de voorspelling was die de rest van de wereld zou ondergaan, volgens internetpionier Marc Andreessen in 2011. In een opiniestuk voor *The Wall Street Journal* betoogde Andreessen dat de succesbedrijven van de toekomst allemaal softwarebedrijven zouden zijn, ongeacht de sector waarin ze actief zijn. Hij verwees naar het enorme succes van vijf grote bedrijven – Facebook, Netflix, Spotify, Amazon en Apple – maar ook naar het toenemende gebruik van software in auto's, in de geneeskunde, in defensie en in het onderwijs. (Andreessen 2011)

De vaststelling van Andreessen was juist. De digitalisering heeft heel wat traditionele economieën 'opgegeten'. In enkele decennia maakten we de overgang van analoge muziekopnames, verkocht op cassette en elpee, naar digitale muziekopnames. Eerst werden ze verkocht op cd-schijfjes, dan als downloadbare bestanden voor mp3-spelers, en vandaag via streamingdiensten als Spotify, Apple Music of YouTube. Denk ook aan foto's. Voor het gros van de gebruikers is dit een puur digitaal product. Ze worden vastgelegd met een digitale sensor, opgeslagen op een digitaal medium, en

bekeken op een digitaal scherm. Waarom is dat met het boek niet gebeurd?

Daarvoor liggen verschillende verklaringen voor de hand. Zo is er veel meer verschil tussen een papieren boek en een digitaal boek, dan tussen analoge en digitale muziek. Je moet een kenner zijn om het verschil te kunnen horen tussen een analoge en een digitale muziekopname. Aan de andere kant is er een manifest verschil tussen een boek op papier en een 'boek' op een iPad. Een papieren boek geeft geen licht en heeft geen elektriciteit nodig. Een digitaal boek kun je (voorlopig toch) niet meenemen in bad; je kunt het niet plooiën; je kunt er niet in bladeren; je kunt er moeilijk in schrijven.

Het ontbreekt digitale boeken dus aan de tastbaarheid van papieren boeken, gebonden in een boekband en beschermd door een mooi vormgegeven omslag. Het ontbreekt ze ook aan duurzaamheid. Dat mijn digitale boeken (en de annotaties die ik bewaar) mogelijks kunnen veranderen of zelfs verdwijnen wanneer ik een nieuwe tablet koop, stemt me niet gelukkig. De annotaties die ik ooit op papier maakte in de boeken die ik als student kocht, staan er vandaag nog altijd in. (Helaas, denk ik nu, want er staat heel wat onzin tussen.)

Die persistentie van papier kan digitale technologie niet bieden. Ze bestaat essentieel in verandering. Elk moment kan ze geüpgraded worden – of voorbijgestreefd raken. En laat nu net die onafgebroken upgrade-cyclus iets zijn wat je in sommige domeinen van het leven niet wil. Analoge informatiedragers – annotaties in boeken, getekende schetsen of dagboeknotities – ontlenen een deel van hun betekenis aan hun persistentie en hun materiële vorm. Die materialiteit brengt nog een kenmerk met zich mee: een ondubbelzinnige bepaling van wie de eigenaar is van een papieren object. Bij heel wat digitale systemen ben je geen eigenaar van een digitale kopie van een werk, maar heb je er slechts toegang toe die je op elk moment weer ontzegd kan worden.

Een derde reden voor het merkwaardige overleven van het boek is esthetisch van aard. Bij e-books bepaalt de gebruiker de typografie. Je kiest zelf bijvoorbeeld het lettertype, de regelbreedte, de interlinie, de corpsgrootte en de grootte van de marges. Dat is heel gebruiksvriendelijk, maar alles ziet er hetzelfde uit. Ongeacht of je nu Franse non-fictie leest, een Amerikaanse thriller, of het laatste boek van Jeroen Olyslaegers. Het is aangenaam en frictieloos, maar ook kleurloos.

De vierde en laatste reden is het belangrijkste: het digitale boek heeft zijn beloftes niet waargemaakt. Laat me hiervoor teruggrijpen op een fundamenteel onderscheid uit de boekwetenschap. Het betreft twee tegengestelde definities van het boek. (Feithert & Sturges 2003) De strekking van de eerste is dat een boek bestaat uit een stapel papier bijeengehouden door een binding. Dat is een materiële definitie van het boek als 'codex'. De tweede definitie, daarentegen, benoemt het boek als een drager van informatie, in welke vorm dan ook. Papyrusrol, middeleeuws handschrift, paperback, cd-rom of e-reader: het kan allemaal in aanmerking komen voor de benaming 'boek'.

De belofte van het digitale boek sloot aan op die tweede definitie. Eindelijk zouden de beperkingen van het materiële boek (zijn gewicht, zijn kwetsbaarheid voor vocht,

vuur en verzuring) verdwijnen. Vanaf nu had je het 'pure' boek in handen - het boek als zuivere informatiedrager. Dat is echter niet helemaal gelukt.

Google Books is een treffend voorbeeld, want de ambitie van het massadigitaliseringsproject was om alle informatie over de wereld opgeslagen in boeken 'op te graven' via een reusachtige scanoperatie. (Auletta 2009: 96) Daarvoor moesten de gescande boekpagina's eerst worden omgezet naar digitale tekst en dan worden 'ontgonnen' met taalsoftware. Die software herkent eigennamen, merknamen, titels, plaatsen en tijdstippen. Bovendien kan de technologie ze ook aan elkaar linken.

Google gebruikt zulke technieken bijvoorbeeld in zijn internetzoekmachine. Sinds 2012 krijg je in een apart venster een zogenaamd 'kennisvenster' (Knowledge Graph) te zien, wanneer je naar de naam van een bekende persoon, organisatie, film of boek zoekt. De software herkent de zoekopdracht 'leonardo da vinci' bijvoorbeeld als een mogelijke verwijzing naar de historische persoon Leonardo da Vinci, die in de achterliggende databank gelinkt is aan een geboortedatum, een overlijdensdatum, een reeks schilderijen, een aantal boeken, enzovoort.

Het kennisvenster is een eenvoudige, maar doeltreffende illustratie van de kracht van Googles taaltechnologie. Het schoentje knelt echter bij de toegankelijkheid ervan. Hoewel het techbedrijf de goed gestructureerde en dagelijks bijgewerkte informatie op de open encyclopedie Wikipedia graag gebruikt en jaarlijks een aardige financiële gift aan de non-profitinstelling overmaakt, is er niettemin geen haar op het hoofd van de CEO dat eraan denkt om ook de eigen databank publiekelijk toegankelijk te maken.

De ambitie waarmee Google Books oorspronkelijk werd gelanceerd, is dus onderweg gefnuikt. De ontgonnen informatie vind je niet terug in de interface van Google Books. Waar blijft de universele databank van alle menselijke kennis geëxtraheerd uit alle boeken ooit gepubliceerd? Die is er niet, of als ze wel bestaat, dan toch niet voor iedereen toegankelijk.

Openbare massadigitalisering

Een gelijkaardig verhaal kun je over de digitale boeken in de Kindle Store vertellen. Amazon moet ondertussen een gigantisch arsenaal aan gegevens over onze leesgewoonten hebben opgebouwd. Hoelang doet de gemiddelde lezer over één pagina van Stephen King? Op welke bladzijde van *The Lord of the Rings* haken de meeste lezers af? Is er een verschil tussen het leesritme van mannelijke en vrouwelijke lezers van Paul Auster? Het zijn het soort vragen die je als onderzoeker van leesgedrag graag zou willen beantwoorden. De marketingafdeling van de Kindle Store zal de kostbare informatie allicht nog een eeuwigheid in haar kluisen opbergen, en hoogstens tegen een hoge vergoeding aan uitgevers verkopen.

Digitale boeken behoren de digitale lezer niet helemaal toe. Dat is de onvervulde belofte. De kennis die ze bevatten – over de wereld, maar ook over hoe wij lezen – is het goed beschermde intellectuele eigendom van enkele puissante en beursgenoteerde bedrijven.

Wil je als onderzoeker zelf die kennis gaan ontginnen, dan kom je bij openbare digitaliseringsprojecten terecht. Daarvan zijn er vele tientallen, maar ze hebben doorgaans heel andere doeleinden dan het digitaliseren van miljoenen boeken. Archieven, musea en bibliotheken digitaliseren bijvoorbeeld heel selectief alleen het drukwerk dat door de verzuring van papier wordt bedreigd. Of ze digitaliseren alleen het materiaal dat afkomstig is uit één bibliotheek, of van één bepaalde auteur, of juist het werk uit één korte historische periode. Het doel is in dat geval niet conservatie, maar wetenschappelijk onderzoek. Het betreft ook dikwijls teksten die legaal zo ‘veilig’ mogelijk zijn. Dat wil zeggen dat ze auteursrechtenvrij zijn, omdat de auteur meer dan 70 jaar geleden is overleden.

Voorbeelden van zulke projecten zijn de Perseus Digital Library, een immense collectie van teksten en archeologische objecten uit de oudheid; het Oxford Text Archive; het William Blake Archive; het Dante Gabriel Rossetti Archive; of het Women Writers Project, dat teksten bevat van vroegmoderne vrouwelijke auteurs.

Weinig digitaliseringsprojecten uit de non-profitsector kun je bestempelen als echte massadigitalisering. Belangrijke uitzonderingen op die regel zijn Project Gutenberg; de boekencollectie van het Internet Archive; de HathiTrust Digital Library; en het Gallica-project van de Bibliothèque nationale de France.

Niettemin zijn er belangrijke restricties verbonden aan de toegang tot de digitale materialen van deze projecten. Gallica, bijvoorbeeld, vraagt om naamsvermelding (niet van de auteur, maar van Gallica zelf) wanneer je een werk uit het publieke domein – dat dus geheel rechtenvrij is – voor een niet-commerciële toepassing wil gebruiken. En voor commercieel gebruik eisen ze licentievergoedingen.

Bovendien maakt geen van deze projecten de bovenvermelde belofte waar. Waar is de kennis die in deze digitale teksten – en in de gegevens over hoe ze digitaal geconsumeerd worden – verborgen zit?

Wat de Lage Landen betreft, is de Digitale Bibliotheek voor de Nederlandse Letteren (DBNL) één van de meest opmerkelijke initiatieven. Met in totaal bijna 17.000 beschikbare ‘werken’ (een generieke term die alles omvat gaande van een pamflet van 1 pagina tot een tijdschriftjaargang van 1.500 pagina’s) is de DBNL gemeten naar de omvang van het taalgebied een grootschalig digitaliseringsproject. Het overstijgt gemakkelijk de beperkingen van zuiver academische of museale digitalisering.

De DBNL omvat namelijk zowel historische als actuele literatuur, en daarnaast ook historische en wetenschappelijke teksten. De interface maakt duidelijk dat de digitale boeken niet alleen voor onderzoekers zijn bestemd. Bovendien zijn veel teksten niet alleen via de website www.dbnl.org te consulteren, maar kun je ze ook als e-book (in epub-formaat) downloaden en lezen.

Digitale bibliotheken en het auteursrecht

De talrijke teksten uit de twintigste en eenentwintigste eeuw brengen onvermijdelijk het probleem van auteursrechten met zich mee. Een tekst digitaliseren en beschikbaar maken op een digitaal platform valt namelijk te beschouwen als een nieuwe uitgave

(‘openbaarmaking’) van de tekst. En daarvoor is uiteraard toestemming van de auteur vereist (wanneer die nog in leven is) of van diens erfgenamen (wanneer de auteur minder dan 70 jaar geleden overleed). Grootschalige digitalisering valt onmogelijk te verenigen met het aftoetsen, en eventueel afkopen, van de toestemming van de rechthebbenden voor elk individueel werk dat men in de collectie zou willen opnemen. Daarom opteerden de academische pioniers van de digitalisering zo vaak voor ‘veilige’ teksten uit de negentiende eeuw of eerder.

Google koos een andere aanpak. Het bedrijf bouwde eerst een omvangrijke collectie van niet-auteursrechtelijk beschermde werken uit, in samenwerking met een aantal gevestigde en prestigieuze academische bibliotheken. Daarna stak het bedrijf zijn tentakels uit in de richting van de wél beschermde werken. Het bood ze aan op zijn platform (onder de vorm van kleine fragmenten of zelfs hele hoofdstukken) op een strikt genomen illegale wijze. Aan auteurs of uitgevers had men geen toestemming gevraagd. Daarmee won het platform wel aan bekendheid en populariteit, wat de onderhandelingspositie van de technologiereus bij latere rechtszaken versterkte. Uiteindelijk werd Google Books een digitale bibliotheek die het midden houdt tussen een onderzoeksbibliotheek (je kunt talrijke werken uit de negentiende eeuw en eerder naar hartelust doorzoeken, en vaak ook als pdf-bestand downloaden) en een boekenwinkel (boeken die vandaag nog in druk zijn, kun je als e-book aankopen).

Problematisch blijven de zogenaamde *orphan works* of ‘verweesde werken’. Dat zijn teksten die strikt genomen nog auteursrechtelijke bescherming genieten, maar waarvan de rechthebbenden niet opgespoord kunnen worden. Deze boeken werden geruime tijd geleden gepubliceerd en de meeste ervan zijn ondertussen uit de handel verdwenen. In de Europese Unie werd in 2012 een ‘Richtlijn inzake bepaalde toegestane gebruikswijzen van verweesde werken’ aangenomen. De voorwaarden waaraan je moet voldoen om verweesde werken online te mogen tonen, zijn echter omslachtig en arbeidsintensief.

Door zijn talrijke overeenkomsten met grote internationale uitgeverijen, kan Google het probleem van de orphan works tot op zekere hoogte omzeilen (tenminste voor de werken waarvan de uitgeverij nog steeds actief is, of door een grotere speler werd opgekocht). Voor een minder grote speler, zoals de DBNL, is dat moeilijker werkbaar. Aangezien de digitale rechten van oudere werken hoogstwaarschijnlijk nog bij de auteurs of hun erfgenamen berusten, zou de DBNL hen individueel moeten opsporen.

Tot mei 2017 koos de DBNL voor de individuele aanpak. Ze vroegen voor elk werk rechtstreeks toestemming aan auteurs en erfgenamen. Sinds die datum sloot de Koninklijke Bibliotheek Den Haag (KB), waar de digitale bibliotheek is ondergebracht, een overeenkomst met de Stichting Lira, de Nederlandse auteursrechtenorganisatie voor schrijvers, vertalers en freelance journalisten. Daarmee worden de auteursrechten van schrijvers van in Nederland uitgegeven tijdschriften en boeken, die niet meer commercieel verkrijgbaar zijn, collectief geregeld. Via Lira betaalt de KB aan de rechthebbende auteurs of hun nazaten een beperkte vergoeding, die vanwege

het niet-commerciële karakter van de bibliotheek beperkt is. Dat geldt niet alleen voor auteurs die aangesloten zijn bij Lira. Ook niet-aangesloten auteurs kunnen aanspraak maken op die vergoeding.

Scheefgroei tussen Nederland en Vlaanderen

De overeenkomst met Lira omvat echter op beperkte schaal de werken die zijn uitgegeven in Vlaanderen. Er zijn wel werken van Vlaamse auteurs beschikbaar op de DBNL, maar dat zijn ofwel werken van vóór 1940, ofwel werken uitgegeven in Nederland, ofwel werken waarvoor een individuele overeenkomst werd gesloten.

De Nederlandse regeling valt namelijk moeilijk te kopiëren naar de Belgische situatie. Terwijl Lira in Nederland de enige collectieve beheersorganisatie voor auteurs is, zijn Vlaamse auteurs bij heel verschillende beheersvennootschappen aangesloten, elk met een eigen specialiteit (de voornaamste zijn SABAM, SACD, Scam, JAM en VEWA). In principe zouden dus onderhandelingen met al die vennootschappen opgestart moeten worden.

Het probleem wordt verergerd door de beperkte slagkracht van de DBNL in Vlaanderen. De DBNL ontstond in 1999 als stichting binnen de schoot van de Maatschappij der Nederlandse Letterkunde, en werd financieel ondersteund door de Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO) en de Nederlandse Taalunie. Vandaag wordt de DBNL gemeenschappelijk gedragen door de Taalunie, de Koninklijke Bibliotheek (KB) en de Vlaamse Erfgoedbibliotheken (VEB).

De Vlaamse Erfgoedbibliotheken is een relatief kleine organisatie. Ze beschikt niet over dezelfde middelen of een groot personeelsbestand als de Nederlandse KB. Wel streeft de VEB samen met de KB en de Taalunie ernaar om ook in België een collectieve regeling te treffen. Tot die tijd werkt de Vlaamse DBNL-medewerker aan het vrijmaken van de rechten op Vlaamse werken.

De Koninklijke Bibliotheek van België (KBR) werkt tot op heden niet mee aan de DBNL. Volgens directeur ad interim Sara Lammens beschikt de KBR evenmin over de budgetten om rechten te kunnen betalen aan de beheersvennootschappen. De KBR trekt ook in twijfel of de Nederlandse oplossing zomaar naar de Belgische situatie gekopieerd kan worden. Zelfs al zouden er van overheidswege budgetten vrijgemaakt worden, dan zou het nog niet eenvoudig zijn gezien het ontbreken van een unieke gesprekspartner in het kluwen van beheersmaatschappijen dat in België bestaat. Het versnipperde landschap maakt het erg ingewikkeld te onderhandelen en contracten af te sluiten.

De KBR gooit het daarom over een andere boeg: ze zal zich inzetten voor een zo breed mogelijke toegang, voor extra middelen én voor een soepeler wettelijk kader. Daarom trad de bibliotheek recent toe tot het Samenwerkingsverband Auteurs en Samenleving (SA&S). Deze lobbygroep van culturele, educatieve en wetenschappelijke organisaties ijvert voor een ingrijpende hervorming van het Europese auteursrecht. (Lammens 2019)

Momenteel maakt de situatie van Vlaamse auteurs op de DBNL dus niet zo'n goede

beurt. Individuele werken worden door een projectmedewerker van de DBNL uitgeklaard door persoonlijk te onderhandelen met de auteurs of hun erfgenamen. Ook worden er afspraken gemaakt met individuele uitgevers en tijdschriftredacties, zodat zij kunnen helpen om de teksten van hun auteurs ‘vrij te krijgen’. Geval per geval behandelen werkt uiteraard tegen de grootschaligheid die een digitale bibliotheek zoals DBNL beoogt.

De consequentie is dat vandaag niet minder dan vierhonderd boeken van (overwegend) Vlaamse auteurs wél gedigitaliseerd zijn door de DBNL, maar door onzekerheden over auteursrechten niet getoond kunnen worden. Evenmin zijn ze beschikbaar voor onderzoekers in de taal- en letterkunde. Hetzelfde geldt voor een vijftigtal jaargangen van tijdschriften. Ook daar blijven honderden teksten onbeschikbaar voor lezers en vorsers. En het probleem wordt erger. DBNL past een bepaalde zelfcensuur toe bij het selecteren van nieuwe teksten voor digitalisering. Werken van auteurs die men niet verwacht beschikbaar te kunnen maken, belanden ook niet onder de scanner. Daaronder dus in hoofdzaak werken van Vlaamse auteurs.

Digitale overvloed – Access denied

De onbeschikbaarheid van honderden werken van Vlaamse auteurs op de DBNL lijkt een gering probleem wanneer dit afgemeten wordt aan de duizenden andere Nederlandstalige titels die wél toegankelijk zijn. In wat volgt wil ik betogen dat er wel degelijk zware consequenties zijn.

Allereerst reproduceert de selectieve digitalisering van literaire teksten op DBNL een bestaande culturele ongelijkheid tussen Nederland en Vlaanderen. Een digitale bibliotheek opgestart in Nederland door Nederlandse organisaties zal allicht prioriteit geven aan Nederlandse auteurs. Juist door de oorspronkelijke samenwerking met de Taalunie was er vanaf het begin de kans om de taalkundige minderheden binnen het Nederlandse taalgebied – de grote Vlaamse minderheid samen met de kleinere Surinaamse, Zuid-Afrikaanse en Friese minderheden – een adequate plek binnen de nieuwe digitale canon te geven. De auteursrechtelijke problemen ondergraven deze kans voor het Vlaams-Nederlands. De cultureel dominante positie van de Nederlands-Nederlandse literatuur wordt zo, eenvoudigweg door de gebrekkige beschikbaarheid van de Vlaamse auteurs, opnieuw bestendig.

Niet enkel voor de doorsnee lezer is dit een slechte zaak. Digitale teksten zijn een zegen voor blinden en slechtzienden, omdat zij door middel van voorleessoftware opeens niet langer afhankelijk zijn van de (dure) omzetting van gedrukte boeken in brailleschrift.

Dat was de reden waarom in de Verenigde Staten ook de National Federation of the Blind betrokken was bij de rechtszaak van auteursvereniging Authors Guild tegen Google Books en de HathiTrust. In die zaak bepleitte de Authors Guild dat de werkwijze van Google een schending was van het auteursrecht, en aanzienlijke financiële schade toebrengt aan auteurs. Het bedrijf – samen met de wetenschappelijke bibliotheken verenigd in de HathiTrust – betoogde daarentegen dat het niet openstellen van

grote digitaliseringsprojecten enorme schade zou toebrengen aan het wetenschappelijk onderzoek en de culturele ontwikkeling; exponentieel groter dan de eventuele financiële schade die de rechthebbenden van oudere werken zouden oplopen. Bovendien zouden blinden en slechthorenden een enorme achterstelling ondergaan, mochten al deze digitale (en dus voor visueel gehandicapten ontsluitbare) teksten opnieuw achter gesloten deuren verdwijnen.

Verder zijn ook onderzoekers de dupe van de hachelijke situatie omtrent auteursrechten. Ik noem de situatie hachelijk omdat het intellectueel recht vandaag beheerst wordt door een zwart-witregime dat de meeste recente teksten zonder pardon – en met uiterst eng afgebakende uitzonderingen, zoals het citaatrecht – achter de muur van het copyright afschermt. Wie baat heeft bij dit regime zijn grote bedrijven die via duur juridisch advies en paginalange contracten (zoals de EULA's, die ellenlange documenten die je moet doorlopen bij een update van de software op je smartphone – ook dat is auteursrecht) de toegang tot hun eigen materialen hermetisch kunnen afsluiten, en op maat toegang kunnen kopen tot materialen van andere bedrijven. Met auteurs of intellectuelen heeft 'auteursrecht' of 'intellectueel recht' al lang niets meer te maken.

Bijgevolg is het leeuwendeel van de twintigste-eeuwse boeken, tijdschriften, kranten, foto's en kunstwerken momenteel – als ze al gedigitaliseerd worden – alleen op de oorspronkelijke analoge drager toegankelijk voor historici. Op het platform Wikimedia Commons, waar digitale afbeeldingen en andere mediabestanden gratis kunnen worden upgeload om ze ter beschikking te stellen van internetgebruikers (het bekendst uiteraard via het gebruik ervan op Wikipedia), is dit ontegenzeggelijk de grootste barrière voor wie met goede bedoelingen iets wil toevoegen. Je kunt dat enkel doen wanneer je een schriftelijke verklaring toevoegt van de rechthebbende dat alle rechten op het werk in kwestie (ook de rechten voor commercieel hergebruik) worden vrijgegeven. Uit mijn eigen ervaring blijkt dat vaak een taai brok voor rechthebbenden. De gevraagde toelating is zó ruim dat zij terugdeinzen om hun werken – zelfs als die meerdere decennia oud zijn, en niet langer commercieel geëxploiteerd worden – ter beschikking te stellen.

Teksten van de DBNL die niet mogen worden vrijgegeven via de website, komen ook niet in de corpora van taalwetenschappers terecht. Precies zulke onderzoekers, die gericht zijn op wat bekend staat als digitale geesteswetenschappen of digital humanities, zitten sinds tien jaar in een stroomversnelling. Dat teksten van talrijke Vlaamse auteurs daarvoor niet kunnen worden gebruikt, is een aanzienlijk nadeel. Onderzoekers als Erik Tjong Kim Sang hielden zich eerder al bezig met linguïstische analyse van teksten afkomstig uit de DBNL. Onder meer de gedigitaliseerde Statenvertaling diende als input voor een onderzoek naar de 'omzetting' van teksten in zeventiende-eeuws Nederlands naar hedendaags Nederlands. (Tjong Kim Sang 2016) Mijn eigen onderzoek focuste op een ander corpus, dat niet in de DBNL zit. In nauw overleg met de redactie werden alle jaargangen van het Vlaamse podiumkunsttijdschrift *Etcetera* gedigitaliseerd en online toegankelijk gemaakt. Vervolgens kon ik

onderzoeken in welke mate het tijdschrift had bijgedragen tot de popularisering van de regisseurs en choreografen van de zogenaamde ‘Vlaamse Golf’ uit de jaren 1980 en 1990. (Crombez 2014)

De mogelijkheden van de DBNL-collectie als corpus voor wetenschappelijk onderzoek zijn nog amper aangesneden. (Het zou onder meer handig zijn als de DBNL daarvoor ook de nodige technische voorzieningen treft, zoals het aanbieden van teksten in XML-formaat of via een API.) Wil je bijvoorbeeld werken aan zo’n universele databank van menselijke kennis, opgeslagen in boeken – de droom waarmee Google Books ooit begon – dan is taalkundige analyse van een grootschalig Nederlandstalig corpus als DBNL een noodzakelijk begin. Op basis daarvan zou je eigennamen, plaatsnamen, merknamen en tijdstippen kunnen extraheren. Die kun je vervolgens met elkaar in verband brengen, om zo het kennisnetwerk versleuteld in het boek te ontsluiten. Dit is enkel mogelijk met teksten uit openbare massadigitaliseringsprojecten, want alleen daar valt met enig recht te zeggen: deze digitale teksten behoren de digitale lezer helemaal toe.

Bekijken we de zaken in een ruimer perspectief, dan wordt duidelijk dat de enige weg voorwaarts ligt in een drievoudige strijd. Het eerste scenario is dat er gewerkt wordt binnen de grenzen van het bestaande auteursrecht, maar met haalbare proviesies voor projecten van cultureel en wetenschappelijk belang. Een tweede piste is de ingrijpende hervorming van het auteursrecht, zoals het Samenwerkingsverband Auteurs en Samenleving (SA&S) bepleit. Zo kun je binnen het bestaande auteursrecht een bredere ‘grijze zone’ afbakenen, waar projecten die het algemeen belang dienen – zoals openbare massadigitaliseringsprojecten of naslagwerken als Wikipedia – hun plek kunnen vinden. Op het vlak van digitale tekstanalyse valt er ook beweging te signaleren: sinds 2016 werkt de Europese Commissie aan een richtlijn die een uitzondering op het auteursrecht voorziet voor *Text and Data Mining*.

Ten slotte zou het ook niet slecht zijn mochten onze erfgoedinstellingen over het muurtje kijken naar de strategieën van hun grote commerciële broertjes. Net als de befaamde Bibliotheek van Alexandrië uit de oudheid, waar alle schepen die aanmerden in de stad hun boeken moesten laten overschrijven, is Google Books niet de grootste bibliotheek van zijn tijd geworden door braaf de regels te volgen. Net als zijn illustere voorganger heeft het de regels herschreven. Het is tijd dat onze bibliotheken, musea en archieven zich iets van de bravoure – ja zelfs de arrogantie – van deze bedrijven aanmeten. Er heerst vandaag te veel angst bij bibliothecarissen en archivariissen om zelfs maar te durven denken aan een digitalisering van auteursrechtelijk beschermde werken. Alsof achter elk boek dat men openslaat, de advocaat van de auteur hen staat op te wachten. De werken van onze recente geschiedenis verdienen beter dan hun huidige lot: de analoge vergeetput.

Wiens rechten worden er vandaag nog geschonden, wanneer men bijvoorbeeld krantenteksten of foto’s uit de eerste helft van de twintigste eeuw online plaatst? Welke commerciële belangen zijn er gemoeid met nieuwsberichten die meer dan een halve eeuw oud zijn? Oud nieuws is geen nieuws. In plaats van hun archieven, net als hun

actuele artikels, achter een betaalmuur op te sluiten, zouden nieuwsbedrijven ze juist makkelijker toegankelijk moeten maken voor onderzoekers en het brede publiek, dat een cultureel en wetenschappelijk belang heeft bij de ontsluiting van deze historische documenten.

Belangrijker is de omgekeerde vraag: hoezeer worden de belangen van onze hedendaagse historici vandaag niet geschaad, omdat zij géén toegang hebben tot deze bronnen onder digitale vorm? Vrienden van de letteren en de historie: er wacht ons een nieuw Alexandrië.

Literatuur

- Andreessen, M.**, 'Why software is eating the world'. In: *The Wall Street Journal*, 20-08-2011.
- Auletta, K.**, *Googled: The end of the world as we know it*, London/New York 2009.
- Crombez, T.**, 'Canoniseringsprocessen in de hedendaagse theaterkritiek: Een frequentie-analyse van de 'Vlaamse golf' in het podiumtijdschrift Etcetera'. In: *Tijdschrift voor Nederlandse Taal- en Letterkunde*, 130 (2014) 1: 45-56.
- Feith, J. & P. Sturges**, 'Lemma Book'. In: *IEILS International Encyclopedia of Information and Library Science*, London/New York 2003.
- Lammens, S.**, persoonlijke communicatie met de auteur (e-mail), 22-02-2019.
- Tjong Kim Sang, E.**, 'Converting Seventeenth-Century Dutch to Modern Dutch', *DHBenelux* 2016, <http://www.dhbenelux.org/wp-content/uploads/2016/05/23_eriktksang_FinalAbstract_DHBenelux2016_long.pdf>, laatst geraadpleegd 16-08-2019.