

Tien stellingen over de ideale digitale bibliotheek

Preprint van Boot, P. 'Tien stellingen over de ideale digitale bibliotheek'. Verschenen in: *De link en de letter. Dossier Nieuwe Media*. Speciaal nummer van *Vooys*, 27 (2010) 4, p. 24-35.
Contact: peter.boot@huygens.knaw.nl

Inleiding

De diepgaande invloed van de komst van het Internet en de digitale revolutie op de beoefening van de geschiedenis en de historische letterkunde is al vaak beschreven, onlangs nog door Anthony Grafton (2008). De elektronische toegankelijkheid van boeken die voorheen een landsgrens of zelfs een oceaan verwijderd konden zijn, kan niet anders dan een ingrijpend effect hebben. Maar Grafton constateert ook dat de 'universele digitale bibliotheek' nog ver weg is. Hij laat zien hoe het visioen van een universele bibliotheek een lange geschiedenis heeft: van de bibliotheek van Alexandrië via de grote microfilmprojecten van de vorige eeuw tot de huidige plannen van Google. Massadigitalisering 'will result not in the infotopia that the prophets conjure up, but in one more in a series of information ecologies, all of them challenging, in which readers, writers and producers of text have learned to survive and flourish' (9). En zeker is dat Van Stipriaans verwachting dat het 'niet lang meer [zal] duren of we hebben nauwelijks nog een academisch klinkend voorwendsel om ons huis te verlaten' (2003: 403) nog lang geen werkelijkheid is.

De Digitale Bibliotheek voor de Nederlandse Letteren (DBNL) bestaat binnenkort tien jaar. De pleidooien van Kuiper (bijv. 1996, 2000) en Van Oostendorp (bijv. 1997) voor een digitale bibliotheek voor de neerlandistiek zijn inmiddels ook rond de tien jaar oud. Er is alle aanleiding de stand van zaken rond de digitalisering van historische letterkunde nog eens te bekijken. Er is uiteindelijk nogal wat veranderd in de afgelopen jaren. De veranderingen zijn gedeeltelijk van technische aard (bijvoorbeeld vrijwel universele beschikbaarheid van breedband internet toegang en spotgoedkope schijfruimte), maar hebben vooral betrekking op de context waarbinnen digitalisering plaats heeft. Het breed gedragen streven naar Open Access voor wetenschappelijke publicaties bijvoorbeeld heeft geleid tot een nationale proefschriftensite,¹ en het heeft iets merkwaardigs dat er geen enkele vorm van integratie bestaat tussen de primaire literatuur in de DBNL en de secundaire literatuur op de proefschriftensite en elders. Een veel grotere verandering in de context wordt gevormd door Google's ambitieuze plannen voor de digitalisering van hele bibliotheken tegelijk: de gedachte dat digitalisering op een dergelijke schaal mogelijk en haalbaar is, kan niet anders dan het denken over digitalisering diepgaand veranderen. Het tempo van digitalisering dat voorheen acceptabel leek, lijkt nu een slakkengang.

Iets van het alom gevoelde ongeduld bleek in het relletje rond de verwijten van Sanders (2007) aan het adres van DBNL. Hij betoogde dat de DBNL verouderde technieken gebruikt en daardoor weinig vooruitgang boekt. De nuchtere reactie van de DBNL (Klapwijk & Van Stipriaan 2007) betoogde dat goed digitaliseren geld kost. Hoewel dat op zich geloofwaardig is, is er toch genoeg gebeurd om de oude vanzelfsprekendheden nog eens tegen het licht te houden.

Ik wil daarom in dit artikel de wetten en praktische bezwaren buiten beschouwing laten, en me afvragen hoe een ideale digitale bibliotheek voor de historische letterkunde er uit zou moeten zien. Ik formuleer daartoe tien stellingen over de ideale digitale bibliotheek. Vanzelfsprekend vraag ik dan meer dan nu gerealiseerd kan worden, maar er zijn toch ook

een paar verlangens waarvan het niet duidelijk is waarom ze nog geen werkelijkheid zijn. Vooraf bespreek ik kort een aantal digitale bibliotheken. Het doel daarvan is enerzijds om te laten zien hoe divers het terrein is, anderzijds om duidelijk te maken dat eisen die extravagant lijken in de ene bibliotheek, staande praktijk zijn in de andere.

Bestaande digitale bibliotheken

Voor het doel van dit artikel is de digitale bibliotheek ‘an intentional collection of digital resources assembled, catalogued, indexed, preserved, and presented to serve the needs of scholarship’ (Unsworth 2000). De digitale bibliotheek is in de eerste plaats een verzameling resources (in ons geval teksten), maar daarnaast ook een presentatiewijze daarvan die ten dienste staat van het wetenschappelijk werk. Digitale bibliotheken kunnen heel veel teksten verzamelen maar toch tekort schieten door de gebrekkige wijze waarop deze teksten toegankelijk worden gemaakt.

Overigens eindigt niet elke digitalisatie-inspanning noodzakelijkerwijs in een digitale bibliotheek. Een digitale editie, waarin één of enkele teksten worden uitgegeven met meer aandacht voor hun bijzondere kenmerken dan in het kader van een digitale bibliotheek mogelijk is, kan heel goed op zich zelf staan. Toch is het in het kader van vindbaarheid en duurzaamheid wenselijk dat zulke digitale teksten ook worden opgenomen in een omvangrijker digitale bibliotheek.

Ik laat de digitale editie in dit artikel verder buiten beschouwing en bespreek een aantal digitale bibliotheken.

Digitale Bibliotheek voor de Nederlandse Letteren (DBNL)

De DBNL² is een digitale bibliotheek die zich vooral richt op de Nederlandse taal- en letterkunde. Uit alle perioden van de geschiedenis van het Nederlands is inmiddels een flinke selectie aan teksten aanwezig. Daaronder bevinden zich ook naslagwerken, letterkundige studies, en een flinke collectie werken die eerder cultuurhistorisch dan letterkundig van belang zijn. Op een enkele uitzondering na bevat de DBNL alleen werken die eerder zijn gedrukt.

Sterke punten van de DBNL zijn de volledige en betrouwbare tekst die van alle werken beschikbaar is, en de auteursthesaurus op basis waarvan alle werken van en over een auteur samen gepresenteerd kunnen worden. Dat laatste kan een vanzelfsprekendheid lijken voor wie de ordelijke wereld van de BNTL of de catalogus van een Nederlandse universiteitsbibliotheek gewend is, maar in de wereld van de digitale collecties op het Internet is het dat allesbehalve. Zwakke punten van de DBNL zijn de ontoereikende zoekmogelijkheden en tekortschietende flexibiliteit in de presentatie van de teksten.³

Google Books

Google Books⁴ bestaat uit twee componenten: een programma dat samenwerkt met uitgevers bij het doorzoekbaar maken van hun nieuwe titels, en een programma waarbij het volledige bezit van grote bibliotheken wordt gescand en doorzoekbaar gemaakt. Van een bibliotheek is natuurlijk pas sprake wanneer de werken ook volledig toegankelijk zijn, en dat is maar in beperkte mate het geval. Google's interpretatie van het auteursrecht buiten de VS is dat alleen boeken van voor (nu) 1869 gegarandeerd rechtenvrij zijn. Oorspronkelijke historische uitgaven zijn dus vaak beschikbaar, maar latere edities niet. Bijvoorbeeld Worps brieven van Huygens (1892-1899) of Brills uitgave van de Rijmkroniek van Melis Stoke (1885) blijven voor ons ontoegankelijk.⁵ De hoeveelheid Nederlandstalig materiaal groeit, maar is nog

tamelijk beperkt. De tekst van de gescande boeken wordt door OCR (Optical Character Recognition, automatische tekenherkenning) verkregen en is van heel wisselende kwaliteit.

Sterke punten van Google Books zijn de schaal waarop men durft te denken en de verbeeldingskracht die het denken over digitalisering en de digitale bibliotheek voor altijd veranderd heeft. Een ander sterk punt is de bereidheid om te experimenteren, bijvoorbeeld door te zoeken naar plaatsnamen en die te lokaliseren op een kaart of door te zoeken naar zinsneden die ook in andere boeken voorkomen. Zwakke punten zijn de abominabel slechte catalogisering, de kennelijke slordigheid waarmee het scannen is uitgevoerd en vooral natuurlijk het feit dat rechtenvrije boeken ons worden onthouden.

Perseus

Perseus⁶ is een project dat begonnen is als een collectie van teksten uit de klassieke oudheid. Inmiddels worden ook collecties aangelegd van teksten uit andere streken en tijdvakken: er is bijvoorbeeld een collectie oud-Noorse teksten, een verzameling topografisch materiaal met betrekking tot het Londen van de Renaissance en een Koran. Het kenmerkende van de aanpak van Perseus is dat men ernaar streeft de computer zoveel mogelijk toelichting en verwijzing te laten genereren. Woorden worden standaard gekoppeld aan een lemma en een vertaling, en er wordt een lijst getoond van mogelijke syntactische analyses. De gebruiker kan in geval van twijfel zijn stem uitbrengen op de z.i. juiste analyse. Dezelfde techniek wordt toegepast bij het herkennen van persoons- en plaatsnamen en van verwijzingen naar passages in klassieke teksten (“Hom. Od. 9.1”). In de teksten wordt dan weer bij elke passage getoond waar die passage wordt becommentarieerd.

Gregory Crane, hoofdredacteur van Perseus, schreef een beroemd artikel met de naam ‘What do you do with a million books’ (2006). Het onderwerp is hoe we nog zinvol onze weg kunnen vinden in de enorme collecties van digitale literatuur die we binnenkort ter beschikking zullen hebben. Classici horen bij de meest actieve denkers op dit terrein.⁷

Internet Archive

Het Internet Archive⁸ is een project van de visionair Brewster Kahle, en voert het motto ‘Universal access to all human knowledge’. Het project archiveert niet alleen teksten, maar ook films en muziek. Het meest opvallende onderdeel van het archief is de Wayback Machine, waarmee oude webpagina’s kunnen worden opgevraagd. Voor wat betreft teksten bevat het archief materiaal uit verschillende bronnen, en de kwaliteit is dan ook wisselend. Opnieuw stellen de metadata weinig voor. Er is een beperkte hoeveelheid Nederlands materiaal, maar het loont de moeite om er te kijken.

Een sterk punt van het archief is dat de gegevens in verschillende formaten beschikbaar zijn: in een online lezer, in pdf, maar ook de originele afbeeldingen in hoge resolutie en de uitvoer van de OCR-bestanden. Als lid van de Open Content Alliance streeft het Internet Archive naar vrije beschikbaarheid van het materiaal. Sympathiek is ook de actieve strijd tegen de excessieve claims van de auteursrechtadvocaten.

e-codices. Virtual Manuscript Library of Switzerland

e-codices⁹ is, anders dan de digitale bibliotheken die hierboven werden besproken, een bibliotheek van manuscripten, die dus geen tekst bevat. De manuscripten worden wel gedetailleerd beschreven, zowel naar fysieke opbouw als naar inhoud, maar er is geen transcriptie.

In Nederland hebben we wel de site Medieval Manuscripts in Dutch Collections (MMDC)¹⁰, maar die site bevat alleen handschriftbeschrijvingen met in een enkel geval een hyperlink naar een manuscript. In de Zwitserse collectie zijn van alle handschriften afbeeldingen aanwezig, en kan ook, zoals dat hoort, vanuit de beschrijvingen worden gesprongen naar de afzonderlijke teksten in het handschrift. De manuscript viewer is een mooie toepassing waarin bladeren, zoomen en kantelen van afbeeldingen op een elegante manier is geïntegreerd.

Tien stellingen

Ik geef tien stellingen over de digitale bibliotheek, met een korte toelichting of argumentatie. In een paar gevallen heb ik één of meer concretiseringën ook in de vorm van een stelling gegoten, zonder daarvoor steeds een afzonderlijke toelichting te geven.

1. De ideale digitale bibliotheek bevat een exemplaar van alles wat in Nederland of in het Nederlands is gedrukt of geschreven.

Het is duidelijk dat elke begrenzing een beperking is. De Nederlandse cultuur is altijd gevoed door de culturen van de ons omringende landen. Het bestuderen van ontwikkelingen in de Nederlandse cultuur is onmogelijk zonder die andere culturen erbij te betrekken. Tegelijkertijd is het niet stellen van grenzen een recept voor mislukking. Ik beperk me hier tot het Nederlands en de Nederlandse situatie, niet omdat de ideale digitale bibliotheek zich veel zou moeten aantrekken van lands- of taalgrenzen, maar omdat de digitale beschikbaarheid van de Nederlandse taal en cultuur nu eenmaal onze verantwoordelijkheid is. Ik zie de stellingen in dit artikel ook als een verlanglijst die Nederlandse onderzoekers kunnen deponeren bij de verschillende instellingen die zich bezig houden of zich bezig zouden moeten of kunnen houden met de inrichting van een digitale bibliotheek: bijvoorbeeld het ministerie van onderwijs, de Koninklijke Bibliotheek en de andere (wetenschappelijke) bibliotheken, de DBNL en de Taalunie.

Men zou de formulering ‘alles (...) wat is gedrukt of geschreven’ een voorbeeld van onvoldoende begrenzing kunnen vinden. Gedrukte werken, dat zijn boeken, tijdschriften, kranten, pamfletten, gelegenhedenuitgaafjes, en het spreekt m.i. voor zich dat deze in het ideale geval digitaal beschikbaar zijn. De ideale digitale bibliotheek haalt zijn neus niet op voor populair of efemer drukwerk. Wat is geschreven behelst natuurlijk in de eerste plaats de handschriften van voor de boekdrukkunst, toen het handschrift nog een vorm van publicatie was. Opnieuw spreekt het voor zich dat zulke handschriften online horen te staan. Voor brieven en dagboeken van publieke figuren ligt de wenselijkheid van digitale beschikbaarheid ook voor de hand. Maar zodra de wetenschappelijke belangstelling zich uitbreidt naar egodocumenten, of naar de ervaringen van ‘gewone’ mensen, wordt ook hun schriftelijke nalatenschap van belang.

Digitale beschikbaarheid is uiteindelijk niet alleen van belang voor de onderzoeker die documenten bestudeert, maar ook voor de verslaglegging en de controleerbaarheid van het onderzoek. Met de mogelijkheden van nu is het onacceptabel dat nog substantiële studies verschijnen over werken die niet digitaal volledig ter beschikking staan (Baßler & Karczewski 2009).

Het ideaal is misschien niet haalbaar. Maar er zijn zeer haalbare tussendoelen, zoals een digitale, doorzoekbare versie van elke titel in de STCN, of een digitale versie van elk middeleeuws handschrift in de MMDC. Gegeven de hoge ambities is een heldere prioritering onmisbaar.

2. De ideale digitale bibliotheek geeft een volledige en vrijwel foutloze tekst.

De ideale digitale bibliotheek geeft een volledige tekst, vanaf de voorkant van het omslag tot en met de achterkant. Titelpagina, impressum, colofon en flaptekst presenteren en introduceren de ‘eigenlijke’ tekst (als die bestaat) bij de lezers. Zonder de paratekst is studie van teksten en hun werkzaamheid onmogelijk.

Of de tekst tot stand komt via OCR of door overtypen in Azië is voor de gebruiker niet van belang. Dat de tekst vrijwel foutloos is, is wel belangrijk, zowel voor het lezen als voor het zoeken. Fuzzy search, waarbij gezocht wordt op woorden die lijken op de zoekterm, kan corrigeren voor fout overgenomen woorden, maar alleen ten koste van grote hoeveelheden irrelevante treffers. De (dure) onderzoeker verliest dan de tijd die de bibliotheek bespaard heeft.

In de praktijk zal niet voor alle tekstdragers en niet in alle omstandigheden een foutloze digitale tekst haalbaar zijn: voor manuscripten is een transcriptie vakwerk, voor kranten is gezien de omvang van het materiaal gebrekkige OCR het hoogst haalbare. Het uitgangspunt zou echter steeds moeten zijn dat een leesbare en doorzoekbare tekst vereist is. Waar dat, om redenen van fasering of de aard van het materiaal niet mogelijk is, moet de status van de verschillende teksten duidelijk worden aangegeven. Voor teksten waar de OCR-techniek tekort schiet kan het publiek een bijdrage leveren.¹¹

2.1. De ideale digitale bibliotheek respecteert originele inhoudsopgaven.

3. De ideale digitale bibliotheek bevat facsimile afbeeldingen van elke pagina, ook van recente werken.

Teksten bestaan niet alleen uit woorden, maar ook uit een wijze waarop die woorden zijn gedistribueerd over het papier. De bibliografische codes die bepalen hoe die distributie er uitziet, beïnvloeden de betekenis van de tekst.¹² Reproductie van alleen de tekstuele inhoud van een tekstdrager is een vervalsing van de historische werkelijkheid. Het is waar dat ook een afbeelding geen volledig beeld van de werkelijkheid geeft: bijvoorbeeld gewicht, papiersoort en geur spelen ook een rol in de bibliografische codes. Vooralsnog zijn deze echter moeilijk te reproduceren op het web.

Als een facsimile wordt gegeven, spreekt het voor zich dat de afbeeldingen een zo goed mogelijk beeld moeten geven van de fysieke aspecten van het boek: ze zijn dus in kleur, op ware grootte, van voldoende scherppte, en geven niet alleen tekstpagina's maar ook omslag, titelpagina, blaadjes met errata, en andere onderdelen van de tekstdrager.

4. De ideale digitale bibliotheek biedt teksten aan in een veelheid aan (aanpasbare) uitvoerformaten, waaronder tenminste een prettig leesbaar formaat (bijvoorbeeld HTML), een gestructureerd formaat (XML), een afdrukbaar formaat (bijvoorbeeld PDF), en platte tekst.

De ideale digitale bibliotheek bedient meerdere gebruikersgroepen met meerdere gebruikswensen. Lezen van het beeldscherm is zo'n wens, maar het is niet de enige. Wie een concordantie wil maken van bijvoorbeeld Multatuli's *Ideën*, wil niet net zoveel bestandjes moeten verzamelen als er ideeën zijn. Voor dergelijke toepassingen is een volledige, 'platte' tekst noodzakelijk. Wie een tekst wil meenemen in de trein wil die tekst misschien netjes kunnen printen. Wie zelf verdere analyses wil uitvoeren op de tekst wil wellicht een XML formaat.

Idealiter – maar we praten ook over een ideale digitale bibliotheek – kan de gebruiker zelf aangeven welke onderdelen van de tekst worden getoond. Wil ik een middeleeuws toneelstuk lezen met of zonder de woordverklaring? Heb ik behoefte aan regelnummers? Wil ik, in de uitgave van een briefwisseling, de brieven van beide correspondenten gezamenlijk of die van elk afzonderlijk? Op basis van de XML codering kunnen dergelijke selecties worden gemaakt.

4.1. De ideale digitale bibliotheek drukt op verzoek een exemplaar van niet meer leverbare boeken in de collectie.

4.2. De ideale bibliotheek levert zijn collectie ook in een formaat dat geschikt is voor e-book lezers en mobiele telefoons.

4.3. De ideale bibliotheek biedt zijn inhoud en diensten aan via een Application Programming Interface (API).

Een API maakt het mogelijk dat de inhoud van een site niet alleen geraadpleegd wordt door een menselijke bezoeker, maar ook door programma's. Externe programmeurs kunnen dan extra functionaliteit bouwen op basis van wat reeds in de digitale bibliotheek ter beschikking is. Een voorbeeld zou kunnen zijn: een betere zoekfunctie, een analyse op intertekstualiteit, een vergelijkfunctie die verschillen tussen tekstversies opspoor, etc.

5. De ideale digitale bibliotheek biedt uitgebreide zoekmogelijkheden in titels en in volledige teksten.

Gebruikers kunnen verschillende redenen hebben om te zoeken, en de zoekfaciliteiten die de digitale bibliotheek biedt, moeten daarmee rekening houden. Gebruikers kunnen op zoek zijn naar een werk (op basis van auteur of titel of een ander kenmerk), en ze kunnen op zoek zijn naar woorden. Bij het zoeken naar woorden kan het zoekcorpus de hele digitale bibliotheek omvatten, het kan zich ook beperken tot een enkel boek, met daartussenin verschillende mogelijke gradaties, zoals het oeuvre van één auteur of alle tekst uit één tijdvak. Eventueel kunnen nog andere criteria worden gehanteerd, zoals geslacht, beroep of regio van de auteur. Verdere verfijningen zijn selecties op basis van aard van de tekst: zoekacties die bijvoorbeeld de noten in een tekst niet meenemen of zich juist daartoe beperken.

5.1. De ideale digitale bibliotheek biedt adequate metadata

Het spreekt helaas niet voor zich dat digitale bibliotheken zorg dragen voor adequate catalogisering. Het komt maar al te vaak voor dat werken wel aanwezig zijn maar volstrekt onvindbaar, misschien doordat een collectie tot stand komt op basis van de collecties van meerdere fysieke bibliotheken, misschien ook doordat de digitale bibliotheek wordt gemaakt door mensen met onvoldoende verstand van of aandacht voor catalogisering.

6. De ideale digitale bibliotheek bevat ook modern materiaal

Veel digitale bibliotheken ontstaan vanuit de wens zeldzame of moeilijk toegankelijke werken uit het verleden bereikbaar te maken voor de moderne onderzoeker of andere geïnteresseerde, en dat is ook de achtergrond van dit artikel. Er is echter geen goede reden waarom dat verleden zich niet zou uitstrekken tot gisteren. Integendeel: met elk modern werk neemt ook de waarde van de oudere werken in de digitale bibliotheek toe. De belangrijkste reden

daarvoor is dat automatische analyses op het vlak van intertekstualiteit en onderlinge verwijzingen (zie stelling 8) meer zullen opleveren naarmate meer materiaal aanwezig is. Wanneer de digitale bibliotheek ook hedendaagse studies over de werken uit het verleden bevat, kunnen bijvoorbeeld bij elk middeleeuws vers de plaatsen in de moderne literatuur worden genoemd die naar dat vers verwijzen.¹³

Een bijzondere klasse van moderne boeken moet zonder meer digitaal ter beschikking komen, en dat zijn moderne tekstuittgaven van oudere literatuur. Als dat niet gebeurt, ontstaat de paradoxale situatie dat oudere uitgaven gedigitaliseerd tot ieders beschikking staan, terwijl de moderne – en waarschijnlijk betere – uitgave ongebruikt op de planken staat te verstoffen.¹⁴

6.1. Verstandige auteurs verlangen directe beschikbaarheid van hun boek in de digitale bibliotheek.

Auteurs die een aanmerkelijk financieel belang hebben bij de verkoop van gedrukte exemplaren van hun boek zullen wellicht aarzelen, maar alle andere auteurs winnen bij de digitale bereikbaarheid van hun werken.

6.2. De ideale digitale bibliotheek zet zich in voor inperking van het auteursrecht.

Het auteursrecht was ooit van belang voor de bescherming van de auteur tegen roefdrukken. In de huidige omstandigheden is het echter verworden tot een rem op de ontwikkeling van digitale bibliotheken. Dat blijkt bijvoorbeeld uit Google's genoemde voorzichtigheid in het ter beschikking stellen van rechtenvrije boeken. Een ander voorbeeld is dat in de Verenigde Staten Google kon worden gedwongen tot substantiële aanpassingen in de presentatie van zoekresultaten van boeken waarvoor het auteursrecht nog niet is verstrekt.¹⁵ Een recent Nederlands voorbeeld van de onwenselijke gevolgen van het auteursrecht in zijn huidige vorm is het feit dat Pictoright, de auteursrechtenorganisatie van fotografen en beeldend kunstenaars, de DBNL kon dwingen afbeeldingen van de site te verwijderen.¹⁶

KB-directeuren Bossenbroek en Jansen bepleiten (2008) opname in de Auteurswet van een speciale internetparagraaf, over niet-commercieel gebruik door erfgoedinstellingen. Daarin zou geregeld moeten worden dat digitale bibliotheken niet verplicht zijn vooraf actief de rechthebbenden op te digitaliseren werken op te sporen. De rechthebbenden zouden het recht houden opname van hun werk te weigeren. Uit oogpunt van tactiek en haalbaarheid is dit misschien een verdedigbaar standpunt. Bezien vanuit het standpunt van (de gebruikers van) de digitale bibliotheek is een veel radicalere aanpassing van de wet wenselijk. Na het verstrijken van een beperkte periode (zeg 20 jaar) zou elk gepubliceerd werk in een niet-commerciële digitale bibliotheek opgenomen moeten kunnen worden, ongeacht de opinie van de rechthebbenden en zonder verdere vergoeding. Het valt niet goed in te zien welk maatschappelijk belang gediend is bij de huidige restrictieve wetgeving. Het is echter duidelijk dat een dergelijke wetswijziging niet binnenkort zal worden gerealiseerd. Tot het zover is zal de digitale bibliotheek de grenzen van de wet moeten opzoeken en actief moeten lobbyen.

7. Van de ideale digitale bibliotheek is er maar één.

Het heeft belangrijke voordelen als alle teksten die voor een bepaald onderzoek of een bepaald veld van belang zijn, beschikbaar zijn in één enkele digitale bibliotheek. De gebruiker kan dan met een enkele zoekopdracht alles wat relevant is voor zijn onderzoek doorzoeken.

De hulpmiddelen voor automatische analyse kunnen het volledige terrein bestrijken. Geautomatiseerde analyses die een gebruiker zelf ontwikkelt kunnen uitgaan van een enkel standaardformaat van de digitale boeken.

Omdat in de praktijk terreinen van studie overlappen (er zijn onderzoekers die geïnteresseerd zijn in periodes, anderen interesseren zich voor culturele thema's, weer anderen voor letterkundige genres) is er voor de digitale bibliotheek geen natuurlijke grens.

Tegelijkertijd zijn er argumenten voor het bestaan van meerdere, kleinere, digitale bibliotheken. Het belangrijkste argument is zonder twijfel dat mensen met aandacht en liefde voor een bepaalde periode, een bepaald genre of een bepaald onderwerp, collecties kunnen maken die aantrekkelijker zijn, overzichtelijker en beter toegesneden op het terrein van studie. Een ander argument is dat de aanwezigheid van concurrentie, of alleen maar het bestaan van vergelijkbare sites, kan bijdragen aan innovatie.

Misschien kan een compromis worden gevonden in een bibliotheek die de collectie beheert en basisontsluiting biedt, en verschillende onafhankelijke instellingen die op (deel-)collecties toegevoegde waarde bieden: inleidingen die de teksten positioneren binnen een terrein van studie, meer specifieke zoekmogelijkheden, extra vormen van indexering, etc. De teksten uit de ideale digitale bibliotheek mogen immers verder worden bewerkt (stelling 10) en de digitale bibliotheek biedt de technische mogelijkheden om teksten nader te analyseren zelf aan (stelling 4.3).

8. De ideale digitale bibliotheek is een centrum voor onderzoek naar (semi-)automatische elektronische verrijking van teksten.

De ideale digitale bibliotheek werkt aan technische ontsluiting van de teksten, bijvoorbeeld door analyse van intertekstualiteit tussen de teksten in de bibliotheek en door het automatisch oplossen van verwijzingen binnen boeken en tussen boeken. Wanneer Busken Huet of Hermans verwijzen naar Multatuli's gebruik van de fictieve Mainzer Beobachter zouden de betreffende passages van Multatuli onmiddellijk aanklikbaar moeten zijn. Een andere vorm van automatische verrijking is named entity recognition, waarbij namen van bijvoorbeeld personen, plaatsen, organisaties worden herkend en vervolgens gekoppeld aan bijvoorbeeld andere publicaties of aan lemmata in naslagwerken. Om zoeken te vereenvoudigen kunnen woorden worden gelemmatiseerd, of kan spelling worden genormaliseerd. Voor al deze semi-automatische processen kan feedback of assistentie van de gebruiker worden gevraagd.

9. De ideale digitale bibliotheek is ook een platform voor werken en samenwerken.

Een digitale bibliotheek hoeft niet alleen een collectie van goed of minder goed ontsloten materiaal te zijn, maar kan daarnaast een omgeving bieden om in te werken. Gebruikers zouden hun bevindingen in de vorm van annotaties moeten kunnen achterlaten, docenten zouden studiepaden door de bibliotheek moeten kunnen definiëren, en groepen zouden moeten kunnen samenwerken aan de studie van deelcollecties binnen de bibliotheek. Zulke door gebruikers toegevoegde gegevens zouden optioneel aan andere gebruikers ter beschikking gesteld kunnen worden.

10. De ideale digitale bibliotheek is een publieke nutsvoorziening.

De ideale digitale bibliotheek is gratis toegankelijk. De bibliotheek biedt de neerslag van ons verleden en moet toegankelijk zijn voor iedereen met een interesse in onze cultuur.

In het Engelse taalgebied bestaat Early English Books Online (EEBO), een online versie van een eerdere microfilmserie, die commercieel verkrijgbaar is. Het resultaat is dat

alleen personen die verbonden zijn aan een instelling die zich EEBO kan veroorloven, toegang hebben tot de oudere Engelse literatuur. Zo'n toestand is onwenselijk. Waar Nederlandse universiteitsbibliotheken waarschijnlijk wel een abonnement op een vergelijkbare Nederlandse collectie zouden nemen, zou de toegang voor onafhankelijk onderzoekers, voor scholieren en voor buitenlandse geïnteresseerden onmogelijk worden.

Deze vrije toegankelijkheid hoort niet alleen een recht om te lezen te zijn, maar ook een recht om door te geven en om in technisch opzicht te bewerken (mits de digitale bibliotheek als bron wordt genoemd). Derden kunnen dan op basis van het materiaal in de digitale bibliotheek bijvoorbeeld een hoogwaardige digitale editie maken, of experimenteren met een intuïtiever interface voor de collecties (zie 4.3).

Conclusie

De ideale digitale bibliotheek bestaat niet, zo zou de conclusie en elfde stelling kunnen luiden, en dat is een waarheid als een koe. Maar ik denk dat we iets verder kunnen gaan dan dat. Belangrijke onderdelen voor een digitale bibliotheek voor de historische letterkunde zijn immers wel degelijk in aanbouw.

De DBNL heeft met een gedegen aanpak de basis gelegd voor zo'n digitale bibliotheek. De aanvankelijke aanpak van de DBNL was sterk selectief, wat voor een project dat van start ging in de vorige eeuw begrijpelijk is.¹⁷ De Koninklijke Bibliotheek is inmiddels begonnen met de digitalisering van een ambitieus volume aan boeken (helaas nog steeds niet volledig) uit de periode 1781-1800, in het project Dutch Prints Online.¹⁸ De voltooiing van de STCN maakt het in principe mogelijk alle drukwerk uit de periode 1540-1800 systematisch af te werken, en levert een onmisbaar instrument voor het structureren van de digitale bibliotheek. De inventarisatie van middeleeuwse handschriften in de MMDC legt de basis voor een uitbreiding van de digitale bibliotheek met manuscript images. Integratie van de nationale proefschriftensite zou een eerste stap zijn naar opname van wetenschappelijke literatuur in de digitale bibliotheek, en op grond van de Verklaring van Berlijn zouden de universitaire repositories moeten volgen.¹⁹

Ondertussen heeft NWO met het CATCH programma aangegeven een open oog te hebben voor de betekenis van informatica voor ontsluiting van cultureel erfgoed en geesteswetenschappelijk onderzoek. Al is helaas nog maar weinig van dat geld terechtgekomen bij de ontwikkeling van de digitale bibliotheek zoals die hier is geschetst, het is duidelijk dat in de integratie van de hier genoemde componenten een aantal interessante informatiekundige uitdagingen verborgen ligt. Hetzelfde geldt voor de (semi-)automatische ontsluiting van het tekstmateriaal (stelling 8) en de mogelijkheden van de digitale bibliotheek als platform voor (samen-)werken (stelling 9). De digitale bibliotheek biedt ruimte voor een hele serie aan CATCH onderzoeksprojecten.

Wat nodig is voor het bouwen van de digitale bibliotheek is, natuurlijk, een beetje overheidsgeld, maar vooral gaat het om verbeeldingskracht, de bereidheid tot samenwerken, en enthousiasme over de zich openende vergezichten.

Literatuur

Band, Jonathan, 'A Guide for the Perplexed: Libraries and the Google Library Project Settlement' (2008). <http://www.arl.org/bm~doc/google-settlement-13nov08.pdf>, 31 augustus 2009.

Baßler, Moritz en Rainer Karczewski, 'Computergestützte Literaturwissenschaft als Kulturwissenschaft. Eine Wunschliste', *Jahrbuch für Computerphilologie* - online, 9 (2009).

<http://computerphilologie.tu-darmstadt.de/jg07/BassKarcz.html>, 31 augustus 2009.

Bossenbroek, Martin and Hans Jansen, 'Auteurswet nekt de digitale bibliotheek', NRC Handelsblad, 2008-04-17.

Crane, Gregory, 'What Do You Do with a Million Books?' D-Lib Magazine, 12-3 (2006).
<http://www.dlib.org/dlib/march06/crane/03crane.html>, 31 augustus 2009.

Crane, Gregory en Melissa Terras (eds.), Changing the Center of Gravity: Transforming Classical Studies through Cyberinfrastructure. In: Digital Humanities Quarterly, 3-1 (2009).

Danowski, P., 'Library 2.0 and user-generated content: What can the users do for us', In: 73rd IFLA General Conference and Council (Durban).
<http://ifla.queenslibrary.org/IV/ifla73/papers/113-Danowski-en.pdf>, 31 augustus 2009.

Grafton, Anthony, Codex in Crisis, New York, 2008.

Holley, Rose. Many Hands Make Light Work: Public Collaborative OCR Text Correction in Australian Historic Newspapers. Z.p., 2009.
http://www.nla.gov.au/ndp/project_details/documents/ANDP_ManyHands.pdf, 31 augustus 2009.

Klapwijk, Cees en René van Stipriaan, 'Alle boeken thuis voor bijna niks: pennywise but poundfoolish'. <http://www.dbnl.org/nieuws/29112007.htm>, 31 augustus 2009.

Kuiper, Willem, 'Een deltaplan voor digitale tekstedities', In: Studiedag van de Stichting Tekstcorpora en Databestanden in de Humaniora: Historische data, modern verpakt, 7 juni 1996, Leeuwarden. Persoonlijke mededeling.

Kuiper, Willem, 'Een digitaal deltaplan voor de literatuur van het verleden'. In: BZZlletin Literair, themanummer Middeleeuwen nu, 277 (2000): 113-21.

McGann, Jerome, The Textual Condition, Princeton, 1991.

Sanders, Ewoud, 'Gun boeken een tweede leven: digitaal'. In: NRC Handelsblad.
<http://weblogs3.nrc.nl/woordhoek/2007/11/28/gun-boeken-een-tweede-leven-digitaal>, 31 augustus 2009.

Unsworth, John, 'The Scholar in the Digital Library', 2000.
<http://www3.isrl.illinois.edu/~unsworth/sdl.html>, 31 augustus 2009.

Van Oostendorp, Marc (1997), 'Het Nederlands op het Net', In: emnet (1997).
<http://www.vanoostendorp.nl/computers/ideaal.html>, 31 augustus 2009.

Van Stipriaan, René (2003), 'Waar zit de winst? Een digitale bibliotheek voor de hele neerlandistiek. Enkele overwegingen bij de inrichting van de DBNL', Spiegel der Letteren, 45-4 (2003): 403-16.

¹ <http://www.narcis.info/index/tab/proef/Language/NL/> (12 oktober 2009).

² <http://dbnl.org/> (12 oktober 2009).

³ In een reactie op een eerdere versie van dit artikel meldt de DBNL dat deze tekortkomingen binnen zeer afzienbare tijd zullen zijn opgeheven. Bovendien zal gewerkt gaan worden volgens een twee-fasenmodel waarbij vooruitlopend op de beschikbaarheid van een betrouwbare tekst alvast scans en ‘vuile’ automatisch herkende tekst worden getoond. Het aantal beschikbare pagina’s zal daardoor snel toenemen.

⁴ <http://books.google.com/> (12 oktober 2009).

⁵ Behalve wanneer de bezoeker d.m.v. een zogenaamde proxy verbergt uit welk land hij afkomstig is.

⁶ <http://www.perseus.tufts.edu/> (12 oktober 2009).

⁷ Zie ook het speciale nummer hierover van *Digital Humanities Quarterly* (Crane & Terras 2009).

⁸ <http://www.archive.org/> (12 oktober 2009).

⁹ <http://www.e-codices.unifr.ch/> (12 oktober 2009).

¹⁰ <http://www.mmdc.nl/> (12 oktober 2009).

¹¹ Zie bijvoorbeeld Holley (2009, voor correctie van OCR bij kranten) en Danowski (2007).

¹² ‘(...) the symbolic and signifying dimensions of the physical medium through which (or rather *as* which) the linguistic text is embodied’ (McGann 1991:56).

¹³ Er zijn nog wat elementairder argumenten voor de opname van modern werk in de digitale bibliotheek: ook moderne literatuur staat niet in ieders boekenkast terwijl bibliotheken sluitingsstijden hebben en niet naast de deur zijn, en ook moderne literatuur moet doorzocht kunnen worden.

¹⁴ Een situatie die nu bijvoorbeeld bestaat bij veel van de gedichten van Huygens.

¹⁵ Zie bijvoorbeeld Band 2008.

¹⁶ Zie <http://dbnl.org/nieuws/20091001.htm> en <http://www.pictoright.nl/nieuws/DBNL.html> (12 oktober 2009).

¹⁷ In het aangekondigde twee-fasenmodel zal het aantal boeken sterk worden opgevoerd.

¹⁸ <http://www.dutchprintsonline.nl/> (12 oktober 2009)

¹⁹ De Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities is een verklaring waarin wetenschappelijke instellingen beloven zich te zullen inspannen voor vrije toegang tot wetenschappelijke publicaties. Deze verklaring is onder andere getekend door KNAW, NWO en alle Nederlandse universiteiten.