Het onbehagen in de digitale cultuur. De opkomst van *digital humanities*

Thomas Crombez, Universiteit Antwerpen / Kon. Academie voor Schone Kunsten Antwerpen / Sint Lucas Antwerpen

De digitalisering van boeken en archiefdocumenten, die in de jaren zeventig een aanvang nam, lijkt op het eerste gezicht een succesverhaal. Bibliotheken en erfgoedinstellingen hebben — mede dankzij Google en Amazon — de verscholen digitale krachten van hun eigen collecties ontdekt. Wat aanvankelijk een betrekkelijk artisanaal proces was van manueel scannen en data invoeren, is nu veranderd in geautomatiseerde massadigitalisering. Maar met welke doelstellingen en attitudes wordt er juist gedigitaliseerd vandaag? Voor wie? En hoe krijgt het publiek toegang tot de nieuwe digitale rijkdommen?

GESCHIEDENIS VAN DE DIGITALE CULTUURWETENSCHAP

Niet in de universiteiten, bibliotheken en archieven moet het begin van de digitaliseringsgolf gezocht worden. Het initiatief kwam van outsiders. Twee belangrijke projecten uit de vroege jaren zeventig liggen aan de bron van zowel het digitale boek als het toepassen van digitale methodes in de cultuurwetenschappen.

Het is niet toevallig dat er omtrent 1970 een breekpunt valt te signaleren. Tot op dat moment was de rekenkracht van computers enkel weggelegd voor bedrijven, overheidsinstanties en wetenschappers. Courante toepassingen waren militaire berekeningen, demografische tellingen en wetenschappelijke dataverwerking. Pas met de komst van de microprocessor worden computers snel en goedkoop genoeg om een veel grotere groep gebruikers te bereiken.

Het allereerste digitaliseringsinitiatief rond tekst is Project Gutenberg. Het ontstaat op 4 juli 1971, wanneer Michael S. Hart, student aan de University of Illinois, de Amerikaanse Onafhankelijkheidsverklaring overtikt, en aanbiedt als download op het (toen nog erg kleine) internet. Later zal hij zelf nog tal van klassieke teksten toevoegen, zoals de Bijbel en de werken van Homeros en Shakespeare. Het ideaal is om zo veel mogelijk teksten die vrij zijn van copyright (op dat ogenblik nog gelimiteerd tot 28 jaar na de dood van de auteur, later onder druk van Amerikaanse lobbyisten opgetrokken tot 50 en daarna 70 jaar) gratis beschikbaar te maken voor een zo groot mogelijke groep. Teksten worden als 'platte tekst' beschikbaar gemaakt, gecodeerd in ASCII, zonder opmaak. Vandaag worden meer dan 42.000 werken aangeboden op www.gutenberg.org, op basis van diezelfde richtlijnen (hoewel ondertussen naast platte tekst ook HTML en EPUB worden aangeboden).

De Apple II computer. Een van de eerste succesvolle home computers (1977-1993).

Ook begin jaren zeventig verschijnt het eerste (gedrukte) volume van de *Index Thomisticus*. Het project had Roberto Busa s.j. al in de jaren veertig bedacht, maar kon pas werkelijkheid worden nadat hij de steun van IBM-stichter Thomas J. Watson had verkregen. Als doctorandus had Busa zelf duizenden fichekaarten gemaakt om een bepaalde filosofische term in Aquino's werk te traceren. Hij begreep dus snel wat het potentieel zou zijn van een geautomatiseerde index. Samen met IBM digitaliseerde Busa het volledige werk van de middeleeuwse filosoof, liet het automatisch lemmatiseren (dat wil zeggen, elk woord tot zijn lemma of stam terugbrengen), en kon zo een index aanmaken die elke vorm van elk Latijns woord uit het werk van Thomas bevat.

STAPPEN NAAR EEN DIGITAAL TEKSTENLANDSCHAP

Busa's werk luidde het begin in van de zogenaamde *humanities computing*, een studieveld waarin men computerprogramma's ging ontwikkelen om taal te analyseren, en die vervolgens ook toe te passen op literaire, wijsgerige en historische bronnen. Die bronnen kwamen steeds vaker uit grootschalige digitaliseringsprojecten zoals Project Gutenberg. Tegelijk gingen ze technisch veel verder dan het initiatief van Hart. Nieuwe consortia en standaarden zagen het licht, zoals het TEI of *Text Encoding Initiative*, dat een XML-variant ontwikkelde waarmee een zo divers mogelijk spectrum aan papieren bronnen (gedrukte boeken, efemera, manuscripten) in digitale vorm kon worden omgezet, met behoud van de structurele kenmerken van het document. Automatische tekstherkenning (OCR of *Optical*

Character Recognition) ging er met rasse schreden op vooruit. Niet langer was het nodig om teksten manueel te laten overtypen. Binnen de *humanities computing* ontwikkelde de computerlinguïstiek zich het snelst, mede onder invloed van de sterke belangstelling voor artificiële intelligentie. Het gaf aanleiding tot nieuwe technieken zoals auteursherkenning, automatische vertaling of het detecteren van eigennamen in een tekst.

De projecten die tijdens de volgende decennia binnen universiteitsbibliotheken en onderzoeksgroepen opgestart werden, focusten zich meestal op één auteur of periode. Ze gingen aan de slag met materialen die legaal zo 'veilig' mogelijk waren, dus uit de 19e eeuw of eerder. Voorbeelden zijn de Perseus Digital Library, een immense collectie van teksten en archeologische objecten uit de Oudheid, het William Blake Archive en het Rossetti Archive. Door deze activiteiten kwam de nadruk hoe langer hoe minder op *computing*, en des te meer op *humanities* te liggen. Het veld bedacht zichzelf met de nieuwe titel *digital humanities*, die ondertussen gemeengoed is geworden, zoals blijkt uit de talrijke gidsen en readers die recent gepubliceerd werden.[1]

Het sleutelmoment in de ommeslag van digitalisering naar massadigitalisering kwam er in 2004 met de lancering van Google Books, een website van het internetbedrijf dat niet enkele honderden of duizenden, maar miljoenen boeken doorzoekbaar aanbood. Na een tijd bleek het Googles voornaamste bedoeling te zijn om digitale boeken te verkopen, en zo de concurrentie aan te gaan met de iets later geïntroduceerde, maar immens populaire Kindle Store (en bijbehorende ebookreader) van Amazon. Maar beide initiatieven deden ook de ogen opengaan van heel wat geesteswetenschappers. Amazon kon bijvoorbeeld op basis van de digitale tekst van elk boek automatisch trefwoorden genereren, die opvallend accuraat waren. Zou het mogelijk zijn om deze reusachtige nieuwe bibliotheken, die vele miljarden woorden uit miljoenen volumes doorzoekbaar maakten, te gebruiken om nieuwe inzichten in de cultuurwetenschap te genereren?

Die vragen werden meer kracht bijgezet door massadigitaliseringsprojecten die niet in de commerciele sector waren ontstaan. Het concept van de digitale bibliotheek Gallica.fr was ontwikkeld samen met dat van de nieuwe fysieke site voor de Bibliothèque nationale de France (BnF). Ondertussen bevat de webbibliotheek meer dan twee miljoen documenten, waaronder ook landkaarten, manuscripten, kranten en bladmuziek. Vrijwel alle boeken uit de Google Bookscollectie die verschenen vóór 1900, werden eveneens opgenomen in de gratis toegankelijke collecties van non-profitorganisaties zoals HathiTrust (een bibliotheekconsortium) en het Internet Archive.

Niet alleen boeken, ook historische documenten werden het onderwerp van grootschalige digitaliseringsprojecten. Strafrechtszaken die voor de Central Criminal Court in Londen verschenen, werden gedocumenteerd via *The Proceedings of the Old Bailey* 1674-1913. Parlementaire debatten werden doorzoekbaar gemaakt via digitale databanken zoals Historical Hansard in Groot-Brittannië of Staten-Generaal Digitaal in Nederland. Wat de plenaire sessies van de Belgische Kamer en Senaat betreft, is de situatie minder gunstig. Anders dan in Nederland, waar de Tweede Kamer de expertise van de Koninklijke Bibliotheek in Den Haag inschakelde, is hier te lande alleen sprake van een academisch project, namelijk Plenum.be, opgestart door historici van de Universiteit Antwerpen. Dit project, dat de auteur van deze tekst technisch hielp verwezenlijken, heeft alle door de Kamer zelf ingescande boekvolumes naar tekst omgezet en doorzoekbaar gemaakt. Bovendien kunnen via 'Advanced Search' ook grafieken worden opgevraagd die de populariteit van een bepaalde zoekterm over de tijd weergeven.
(Zie figuren 1 en 2).

Aan het eind van dit (onvolledige) overzicht moet een recente en belangwekkende ontwikkeling op vlak van massadigitalisering worden gesignaleerd: de aggregatie van talrijke digitale collecties, of het

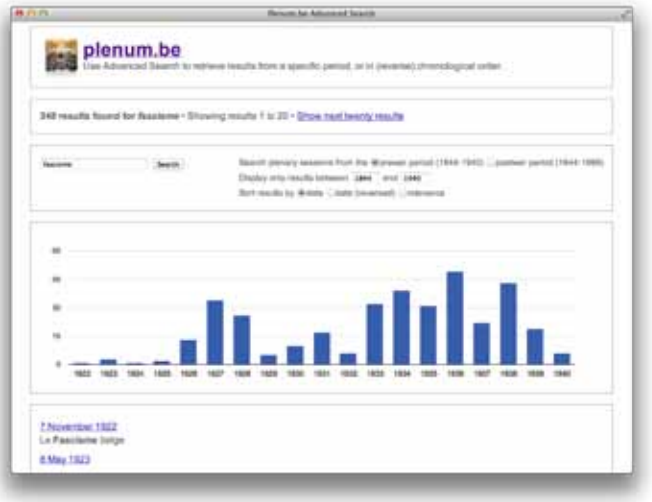

Fig. 1 (boven): Plenum.be, doorzoekbaar archief van de plenaire sessies in de Belgische Kamer voor Volksvertegenwoordigers (1844-1999).
Fig. 2 (onder): Via 'Advanced Search' kunnen grafieken worden opgevraagd die de populariteit van een bepaalde zoekterm over de tijd weergeven. Hier de evolutie van de zoekterm 'fascisme' in de parlementaire debatten (1922-1940).

gekoppeld doorzoeken ervan, bijvoorbeeld via Europeana.eu.

HOE LEES JE EEN MILJOEN BOEKEN?

Welke nieuwe inzichten hebben de nieuwe collecties opgeleverd? Sinds de lancering van Google Books digitaliseerde het bedrijf de voorbije jaren niet minder dan vijftien miljoen boeken en periodieken. Geschat wordt dat er in de hele geschiedenis van het boekbedrijf zo'n 129 miljoen (unieke) boeken gepubliceerd werden. Het corpus vertegenwoordigt circa tien procent van die boekproductie, en kan dus een representatieve steekproef worden genoemd.

Kwantitatief cultuuronderzoek op basis van Google Books werd begin 2011 door Jean-Baptiste Michel en andere onderzoekers met de naam *culturomics* bedacht. In het tijdschrift *Science* publiceerden ze de eerste resultaten van hun analyses. De toepassingsgebieden liepen ver uiteen. Zo konden ze nieuwe en verfijnde schattingen voorleggen over de omvang van het moderne Engelse vocabularium (en de manier waarop het over de laatste twee eeuwen was geëvolueerd). Maar ze toonden ook de effecten van de nazicensuur tijdens de jaren dertig en veertig, door de vermeldingsfrequenties van gecensureerde kunstenaars of schrijvers te vergelijken in het Engelse en het Duitse corpus. De waargenomen verschillen, die heel opvallend zijn voor figuren zoals Marc Chagall, Pablo Picasso of Walter Gropius, kunnen vervolgens gebruikt worden als patroon om andere slachtoffers van de intellectuele repressie te identificeren.[2] Ook de onderzoekers van het Literary Lab aan de universiteit van Stanford (VS) volgen deze benadering. Eén van de stichters, de Amerikaanse literatuurhistoricus Franco Moretti, duidde hun aanpak aan als *distant reading*, een polemisch bedoeld contrast met de algemeen aanvaarde methodologie van de *close reading*. 'Panoramisch lezen' zou een goede vertaling zijn. Hoe vermijd je, zo verdedigt Moretti zijn benadering tegen het klassieke hermeneutische model, dat je steeds weer dezelfde paar honderd canonwerken analyseert? Wanneer je bijv. een geschiedenis van de 19e-eeuwse Engelstalige roman ambieert, moet je dan niet op een of andere manier ook het volledige corpus van de twintig- tot dertigduizend romans gepubliceerd in die periode bij je onderzoek betrekken?[3]

Recent konden onderzoekers van het Literary Lab, op basis van een computerlinguïstische analyse van 2958 Britse romans uit de 18e en de 19e eeuw, aan het licht brengen dat er zich twee duidelijk semantisch samenhangende clusters van woorden aftekenen in de romans. Bovendien worden de woorden van de ene cluster (voornamelijk abstracte begrippen die samenhangen met morele beheersing) minder frequent in de loop van de 19e eeuw, terwijl de cluster van meer concrete woorden toeneemt in frequentie.[4]

In het Nederlandse taalgebied zijn eveneens een beperkt aantal projecten gelanceerd die digitale tekst op grote schaal analyseren om nieuwe inzichten te verkrijgen. Mike Kestemont (Universiteit Antwerpen) kon uit de opvallend afwijkende rijmwoorden in het vijfde deel van de *Spiegel historiael*, met name juist dat deel van de tekst dat onze enige contemporaine bron voor de Guldensporenslag (1302) vormt, afleiden dat de tekst allicht niet door de vermelde auteur (Lodewijk van Velthem) werd geschreven, maar naar alle waarschijnlijkheid uit een andere tekst afkomstig is.[5]

Ook corpora van hedendaagse documenten maken zulke panoramische vormen van onderzoek mogelijk. Het Platform *Digital Humanities* van de Universiteit Antwerpen, dat de auteur van deze tekst in het leven riep, digitaliseerde onlangs de eerste vijfentwintig jaargangen van het podiumtijdschrift *Etcetera* (1983-2008), goed voor een corpus van meer dan 2500 artikels door 657 verschillende auteurs, dat in totaal 5,5 miljoen woorden telt. Een op maat ontwikkelde interface laat toe om termen en namen niet alleen op de klassieke manier op te zoeken, maar ze ook onmiddellijk gevisualiseerd te zien als een grafiek van de evolutie van de vermeldingen over de hele periode. (Zie figuren 3 en 4).

Zo kun je als onderzoeker op zoek gaan naar tendensen en canoniseringsprocessen in de hedendaagse podiumkritiek. Welke namen worden actief gecanoniseerd door de redactie en de auteurs van het tijdschrift? Wie blijft systematisch onvermeld? Erg opvallend is bijvoorbeeld de prominente aanwezigheid van de experimentele, zogenaamd 'postdramatische' theatermakers in de pagina's van *Etcetera*. De generatie van Jan Fabre, Jan Lauwers, Guy Cassiers, Ivo Van Hove en Jan Decorte wordt in niet minder dan vijf tot tien procent van alle artikels tussen 1983 en 1998 vernoemd. Dat alles staat in schril contrast tot de eerder lacunaire aanwezigheid van de regisseurs die vast verbonden zijn aan de stadsschouwburgen tijdens diezelfde periode.

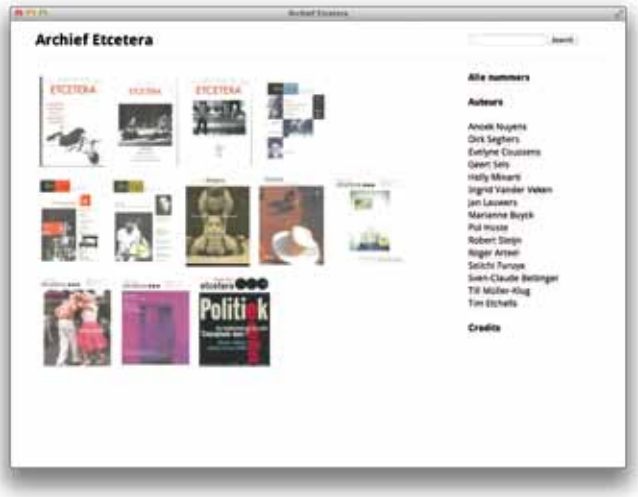

Fig. 3 (boven): Digitaal archief Etcetera.
Fig. 4 (onder): Evolutie van de zoekterm 'participatie' in Etcetera.

"ZOU HET MOGELIJK ZIJN OM DEZE REUSACHTIGE NIEUWE BIBLIOTHEKEN, DIE VELE MILJARDEN WOORDEN UIT MILJOENEN VOLUMES DOORZOEKBAAR MAAKTEN, TE GEBRUIKEN OM NIEUWE INZICHTEN IN DE CULTUURWETENSCHAP TE GENEREREN?"

Die mainstreamregisseurs, waaronder Nand Buyl (KVS), Jean-Pierre De Decker (NTG), Senne Rouffaer (KVS), Dirk Tanghe (NTG) of Hugo van den Berghe (NTG), worden nooit in meer dan drie procent van de artikels vermeld.[6]

Een ander soort projecten zet niet zozeer in op de kwantiteit, maar wel op de andere kwaliteiten van digitale edities. Het Samuel Beckett Digital Manuscript Project, ontwikkeld door Dirk Van Hulle (Universiteit Antwerpen) en Mark Nixon (University of Reading), produceert gesofistikeerde edities van Becketts kladversies, waarin elke doorhaling of kanttekening ook mee gecodeerd wordt. Zulke 'genetische edities' laten toe het ontstaansproces van een tekst nauwkeurig in kaart te brengen. Onderzoekers kunnen op die manier bijvoorbeeld de hele levensloop van een zin doorheen alle kladversies in kaart brengen, een manier van onderzoek doen die bij klassieke historisch-kritische uitgaves vrijwel onmogelijk was.
(Zie figuren 5 en 6).

Het moet benadrukt worden dat de bovenstaande projecten door onderzoekers werden geïnitieerd, niet bibliotheken of archiefinstellingen. Ook wat het eigenlijke aspect digitalisering betreft. Dat staat in scherp contrast met de situatie in de VS of Groot-Brittannië, waar de zogenaamde *research libraries* het voortouw hebben genomen in het oprichten van centra voor *digital humanities* (een overzicht is te vinden op centerNet. In de Lage Landen, en zeker in Vlaanderen, blijft zo'n essentiële interface tussen bibliotheken en erfgoedinstellingen enerzijds en onderzoeksgroepen anderzijds een pijnlijk manco. Mogelijk komt daar een beetje verbetering in via het Nederlab-project, dat ambieert om alle reeds gedigitaliseerde Nederlandstalige teksten en datasets (zoals lexicons en linguïstische corpora) via een eengemaakte collectie aan onderzoekers aan te bieden.

HET ONBEHAGEN IN DE DIGITALE CULTUUR

Het veld van de *digital humanities* mag dan beloftevol zijn, toch lijkt het, te merken aan opiniestukken zoals dat van Stanley Fish in de *New York Times*, ook een trendgevoelig modewoord.[7] Misschien is dat ook wel het grootste gevaar voor dit veld: om nooit meer dan 'beloftevol' te worden. Maar zo nieuw zijn de *digital humanities* nu ook weer niet. Geesteswetenschappers zijn van oudsher op zoek geweest naar nieuwe onderzoeksinstrumenten en nieuwe manieren om hun analyses te visualiseren. Al in de vroegmoderne periode maakte men gretig gebruik van een nieuwe grafische techniek, de gravure, om meer informatie weer te geven op een boekbladzijde dan mogelijk was met de hand. Vergelijkende tijdslijnen, bijvoorbeeld, waren al in omloop sinds de vierde eeuw na Christus. In tabelvorm toonden ze de wereldgeschiedenis door in verschillende kolommen de gebeurtenissen van de Egyptische, Assyrische, Joodse, Griekse en Romeinse tijdsrekening met elkaar te aligneren. Maar via gravures kon men de 'informatiedensiteit' (de term stamt van Edward Tufte) van zulke grafieken sterk doen toenemen, zoals bijvoorbeeld blijkt uit de *Nouvelles tables historiques* die Jean Rou maakte voor de Franse kroonprins in 1675.[8]
(Zie figuur 7 p. 12).

Vandaag zijn er meer mogelijkheden dan ooit om nieuwe databanken en visualiseringen te ontwikkelen. Die laten toe om nieuwe hypotheses te formuleren en te testen. En toch is de situatie vandaag wat betreft de toegankelijkheid van data en de visualiseringsmogelijkheden behoorlijk belabberd. Dat heeft ten

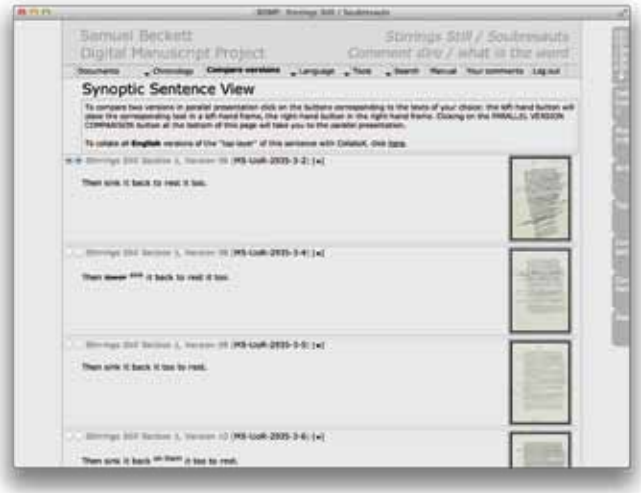

Fig. 5 (links): Levensloop van een zin gevisualiseerd in het Beckett Digital Manuscript Project.
Fig. 6 (rechts): Visualisering van een facsimile in het Samuel Beckett Digital Manuscript Project.

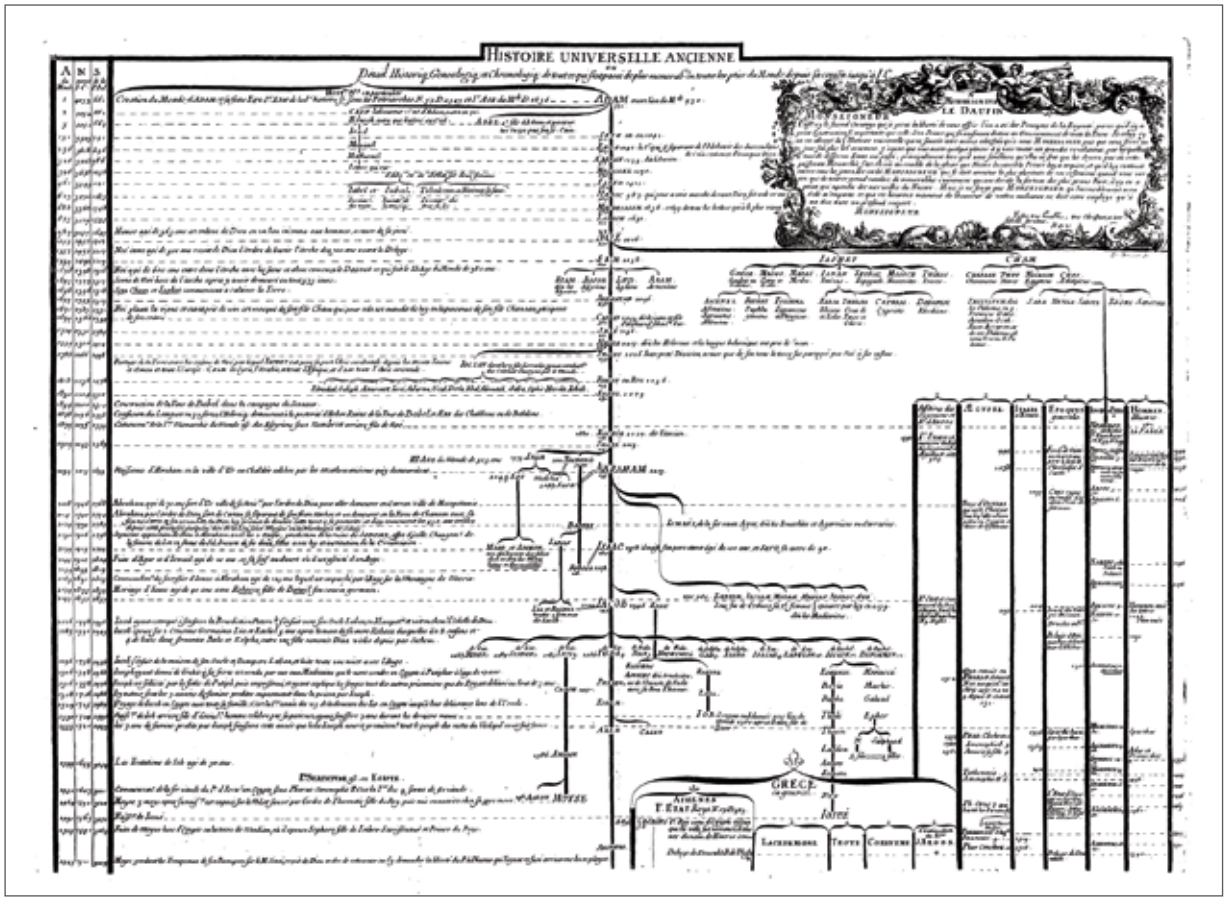

Fig. 7. Jean Rou, *Nouvelles tables historiques* (1675), Histoire universelle ancienne.

"DE UNIVERSITEITS-BIBLIOTHEEK VAN GENT LIET PER VRACHTWAGEN DUIZENDEN BOEKEN DOOR GOOGLE DIGITALISEREN, MAAR WAAR BLIJVEN DE DATA?"

dele met de onderzoekers zelf te maken. Bij de meeste ontbreekt het aan de elementaire expertise over hoe digitale bronnen tot stand komen, en hoe ze gemanipuleerd kunnen worden. Die kennis is nodig vooraleer je nieuwe manieren kunt bedenken om ze te bevragen en in beeld te brengen.

Een andere reden is bij de datavoorzieners te zoeken. Hoewel gestreefd wordt naar standaarden om data op te slaan (zoals TEI-XML of ALTO-XML), is dat veel minder het geval voor de ontsluiting ervan via een website of andere interface. Elke digitaliserende instelling beslist nog steeds autonoom hoe ze hun documenten ontsluiten. Niet altijd is het mogelijk om zoekresultaten volgens verschillende criteria te sorteren, bijvoorbeeld. Bovendien zijn digitale collecties haast nooit als ruwe data toegankelijk, alleen via de 'handige' webinterface. Helemaal schrijnend is het wanneer zelfs die gebrekkige website ontbreekt, zoals bij de krantendigitalisering van de Koninklijke Bibliotheek van België, die een enorme som gemeenschapsgeld investeerde in de digitalisering van 3,2 miljoen krantenpagina's maar er niet in slaagt om die breder te ontsluiten dan op de vijf (!) "speciale pc's in de leeszaal". Zeldzaam is een digitale collectie die in één keer kan gedownload worden, zodat onderzoekers hun eigen interface zouden kunnen ontwikkelen of experimenten bedenken om de data te exploreren.

De oorzaken achterhalen voor het eerste tekort — de ontbrekende technische expertise bij onderzoekers — zou allicht een grondige analyse vragen van de recente ontwikkelingen in de geesteswetenschappen, die in elk geval interdisciplinaire samenwerkingen met informatici of grafische vormgevers niet aanmoedigen.

Het tweede pijnpunt — de haperende toegankelijkheid van veel digitale bronnen — vindt zijn oorzaak in de attitudes die kenmerkend zijn voor de digitaliserende organisaties. Archieven, bibliotheken, musea en andere instellingen die over fantastische collecties beschikken, en druk bezig zijn die te digitaliseren, blijven dat doen volgens een klassieke bibliothecaire of museale logica. Het digitale stuk wordt in een digitale boekenkast of een digitale vitrine gepresenteerd. De fysieke ervaring van de bezoeker probeert men zo goed mogelijk te reproduceren via een website. Enerzijds is dat heel logisch: Marshall McLuhan schreef al dat *"Every new medium begins as a container for the old"*. Bovendien

proberen ze een zo breed mogelijk publiek te bedienen. Anderzijds zijn die keuzes vandaag ook onbegrijpelijk. Ze verraden een diep wantrouwen voor, of tenminste een gebrek aan inzicht in de digitale technologie. Verschillende soorten publiek vragen om verschillende interfaces, en momenteel blijft een aanzienlijk publiek van onderzoekers op z'n honger zitten.

Het besluit van dit beknopte overzicht kan dus moeilijk anders dan signaleren dat er overal opportuniteiten zijn, maar dat die zelden worden aangegrepen. De Universiteitsbibliotheek van Gent liet per vrachtwagen duizenden boeken door Google digitaliseren, maar waar blijven de data? Waar blijft het eerste Vlaamse onderzoeksproject dat deze enorme digitale schatkist kan exploiteren? Tot dusver heb ik, als onderzoeker, de beste kansen wanneer ik langs de sites van het bekende bedrijf in Californië zelf ga (met alle beperkingen van dien). Ander voorbeeld: het project Flandrica.be digitaliseert topstukken uit de Vlaamse erfgoedbibliotheken. Schitterend, maar wat is de omvattende vraagstelling (voor onderzoekers en/of voor een breed publiek) die de selectie heeft gestuurd? Tot nu toe lijkt het een willekeurig samenbrengen van wat op papier het meeste waarde heeft. Ook onze grootste federale erfgoedinstelling, de Koninklijke Bibliotheek van België, lijkt geen kans te laten liggen om zich niet op vlak van digitalisering (laat staan *digital humanities*) te moeten profileren.

De behoefte aan ontmoetingsplaatsen tussen erfgoedwerkers en onderzoekers is dus sterker dan ooit. Het gaat daarbij niet om eenrichtingsverkeer. Deze plekken moeten méér doen dan onderzoekers toelaten hun wensen qua digitalisering kenbaar te maken. Het moeten echte centra voor *digital humanities* worden, die workshops en trainingsessies aanbieden aan studenten, onderzoekers en erfgoedwerkers. Deze plekken zouden daardoor ook kunnen helpen om de collecties van de erfgoedinstellingen voor de erfgoedinstellingen zélf beter inzichtelijk te maken. Wat zit in onze archieven? Wat valt in onze bibliotheken te lezen? Meer dan ooit blijven die vragen open staan, maar meer dan ooit hebben we ook de instrumenten in handen om ze aan te pakken.

> **www.corpusthomisticum.org**
www.perseus.tufts.edu
www.blakearchive.org/blake
www.rossettiarchive.org
www.oldbaileyonline.org
hansard.millbanksystems.com
www.statengeneraaldigitaal.nl
theater.ua.ac.be/etc
digitalhumanities.org/centernet
www.nederlab.nl

1 Enkele noemenswaardige voorbeelden, waarin ook talrijke andere projecten vermeld worden, en waarvan enkele ook online toegankelijk zijn: SIEMENS, Ray en SCHREIBMAN, Susan, eds., *A Companion to Digital Literary Studies*, Londen, Blackwell, 2008 (digitalhumanities.org/companion/view?docId=blackwell/9781405148641/9781405148641.xml); GOLD, Matthew K., *Debates in the Digital Humanities*, University of Minnesota Press, 2012; PRICE, Kenneth M. en SIEMENS, Ray, eds., Literary Studies in the Digital Age, MLA Commons, 2013 (dlsanthology.commons.mla.org).

2 MICHEL, Jean-Baptiste et al., 'Quantitative Analysis of Culture Using Millions of Digitized Books', in: *Science*, 331 (2011), p. 176-182.

3 MORETTI, Franco, *Graphs, Maps, Trees. Abstract Models for Literary History*, Londen, Verso, 2005.

4 HEUSER, Ryan, en LE-KHAC, Long, 'A Quantitative Literary History of 2,958 Nineteenth-Century British Novels: The Semantic Cohort Method', in: *Pamphlets of the Stanford Literary Lab*, 4 (May 2012).

5 KESTEMONT, Mike, 'Velthem et al. A stylometric analysis of the rhyme words in the account of the Battle of the Golden Spurs in the fifth part of the *Spiegel historiael*', in: *Queeste. Journal of Medieval Literature in the Low Countries* 17:1 (2010), p. 1-34.

6 Er zijn evenwel twee belangrijke uitzonderingen op deze observatie: Walter Tillemans (KNS) en Franz Marijnen (KVS). Ik doe een poging om die uitzonderingen te verklaren binnen het kader van een artikel dat binnenkort verschijnt op www.zombrec.be.

7 FISH, Stanley, 'Mind Your P's and B's: The Digital Humanities and Interpretation', in: *The New York Times*, 23-01-2012.

8 ROSENBERG, Daniel, en GRAFTON, Anthony, *Cartographies of Time. A History of the Timeline*, Princeton, Princeton Architectural Press, 2010; TUFTE, Edward. *The Visual Display of Quantitative Information*, 2de ed., Cheshire, CT, Graphics Press, 2001.