

Van Gogh digitaal

MARIANNE PEERBOOM

Eén van de topcollecties uit het Nederlandse papieren erfgoed is de correspondentie van Vincent Van Gogh, die behalve begenadigd kunstenaar ook een enthousiast briefschrijver was. Daarin is een schat aan informatie te vinden, onder andere over zijn artistieke ontwikkeling en de ontstaansgeschiedenis van zijn werken. Het Van Gogh Museum en het Huygens Instituut bereiden een nieuwe wetenschappelijke editie van de brieven voor, het resultaat van langlopend, intensief onderzoek.

In oktober 2009 zullen de uitkomsten van dit onderzoek worden gepresenteerd in een reeks van activiteiten en publicaties. Een rijk geïllustreerde gedrukte editie verschijnt in het Nederlands, Frans en Engels. Daarnaast komt er een wetenschappelijke digitale editie uit, eveneens een gezamenlijk project van het Van Gogh Museum en het Huygens Instituut. In deze web-editie worden naast de geannoteerde tekst ook de digitale facsimile's van alle brieven opgenomen en uitgebreidere annotaties dan in de gedrukte uitgave mogelijk is. Resultaten van het onderzoek zullen ook te zien zijn in een tentoonstelling in het Van Gogh Museum, waarin de kunstenaar zelf aan het woord komt. De bezoeker kan zien wat Vincent



Brief van Vincent van Gogh aan Theo van Gogh (met 1 schets), Van Gogh Museum, Amsterdam

van Gogh over zijn eigen werk dacht en schreef, en kan zijn schilderijen vergelijken met schetsjes ervan die hij in zijn brieven opnam of mees-

tuurde. Dit jaar wordt met subsidie van Metamorfoze de gehele collectie *Brieven en Documenten* van het Van Gogh Museum geconserveerd door middel van *preservation imaging*. Daarmee ontstaat een beeldarchief waaruit voor de gedrukte en digitale editie geput wordt. De collectie *Brieven en Documenten* bevat naast de correspondentie van Van Gogh (samen zo'n 800 brieven, 1200 afzonderlijke

briefbladen) ook circa vierhonderd familiefoto's en overige (familie-) correspondentie, notitieboekjes van Theo van Gogh, broer van de kunstenaar, en dagboeken van diens echtgenote Jo van Gogh-Bonger, lijsten met schilderijen voor tentoonstellingen, uitnodigingen, nota's en kwitanties, programmaboekjes, voordrachten, toespraken en krantenknipsels. De kern van deze collectie komt uit de nalatenschap van Theo van Gogh, Jo van Gogh-Bonger en hun zoon ir. Vincent Willem van Gogh.

Naast de gedrukte en digitale edities wordt uit de gedigitaliseerde *Brieven en Documenten* ook een collectie samengesteld voor het *Geheugen van Nederland* ([www.geheugenvan-](http://www.geheugenvan-nederland.nl)

Vlaanderen: onderzoeksproject multimediale data

'Bewaring & Ontsluiting van Multimedia in Vlaanderen' (BOM-VL) is een onderzoeksproject om strategieën te ontwikkelen voor het op lange termijn digitaal bewaren en ontsluiten van audiovisueel erfgoed in Vlaanderen. Het project loopt van januari 2008 tot juni 2009 en wordt ondersteund door de Vlaamse regering.

In Vlaanderen zijn er duizenden uren aan audio- en videomateriaal, of 'time-based media', opgeslagen op diverse analoge dragers. Deze collecties bevinden zich niet alleen bij de omroepen maar ook bijvoorbeeld bij culturele organisaties, erfgoedinstellingen, overheidsinstellingen. Volgens UNESCO behoort dit au-

diovisueel materiaal tot het meest kwetsbare cultureel erfgoed.

Het BOM-VL-project wil een langetermijnstrategie ontwikkelen voor het digitaal bewaren en ontsluiten van deze collecties in Vlaanderen. Aan culturele zijde neemt een brede waaier van organisaties deel: de Boekentoren (Universiteit Gent), Instituut voor Beeldende, Audiovisuele en Mediakunst (BAM), Vlaams Theater Instituut (Vti), Muziekcentrum Vlaanderen en FARO (Vlaams steunpunt voor cultureel erfgoed). Uit de mediasector participeren de Vlaamse openbare omroep VRT, en ook de commerciële en regionale omroepen. Het IBBT (Instituut voor Breedband Technologie) en het VRT Medialab zorgen voor de wetenschappelijke omkadering.

Binnen het BOM-VL project wordt een gemeenschappelijk innovatieplatform opgezet dat toelaat om stelselmatig bestaande kennis over archiveren en ontsluiten van multimediale data te toetsen aan de realiteit, samen met de onderzoekresultaten uit de verschillende werkpakketten van dit project. Het is verder de bedoeling van de partners om alle resultaten van het project beschikbaar te stellen aan alle spelers in het Vlaamse en Europese cultuur-, erfgoed- en mediaveld. (Jeroen Walterus)

www.ibbt.be
www.faronet.be



nederland.nl). Daarmee wordt het materiaal voor een breed publiek ontsloten als onderdeel van het educatieve programma dat rond de brieven wordt ontwikkeld.

Via deze verschillende activiteiten en publicaties komt een unieke en kwetsbare collectie uit het Nederlandse erfgoed wereldwijd ter beschikking van verschillende geïnteresseerden: van kunsthistorici en andere wetenschappers tot iedereen die belangstelling heeft voor het leven en werk van Vincent van Gogh.

IMPACT: betere OCR voor de digitale bibliotheek

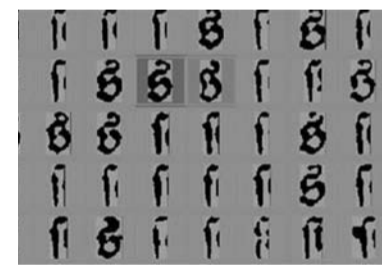
HILDELIUS BALK

In januari 2008 ging het vierjarige Europese project Improving Access to Text (IMPACT) van start. In dit project werken vijftien partners uit bibliotheken, research en bedrijfsleven onder leiding van de Koninklijke Bibliotheek samen.

Doel van het project is om innovatieve oplossingen te vinden voor de verbetering van optische tekenherkenning (OCR), en methodes om het proces van massadigitalisering van gedrukte historische tekst in Europa te versnellen. De eerste resultaten zijn veelbelovend. In juni werd de eerste (interne) versie van het *overall design* van de IMPACT OCR software opgeleverd. Deze software is speciaal toegesneden op het digitaliseren van bibliotheekmateriaal. De kern is een krachtige OCR machine, ontwikkeld onder leiding van IBM Research in Haifa, die zich tijdens het OCR-proces kan aanpassen aan de tekst hij aantreft. Hij hoeft niet, zoals de bestaande OCR software, voorafgaand aan de digitalisering van ouder materiaal met moeilijk leesbare lettertypen en spellingsvarianten te worden 'getraind'. Daardoor verloopt het digitaliseringsproces sneller en goedkoper. De machine wordt geïntegreerd met een andere module van het IMPACT-project: de *Collaborative Correction*. Dit is een

uitgekiend gereedschap waarmee vrijwilligers zonder speciale kennis online kunnen meehelpen aan de correctie van OCR resultaten. Tekens waar de machine over twijfelt bij de OCR – bijvoorbeeld de oude schrijfwijze van de letter s, die er vaak uitziet als een letter f – worden automatisch naar de correctiemodule gestuurd voor beoordeling door een menselijke lezer. De feedback uit de correctie gaat direct terug naar de OCR machine, die zich aanpast aan de nieuwe informatie.

Binnen het uitgebreide verrijgingsprogramma dat door taalkundige partners in IMPACT wordt ontwikkeld, is ook voorzien in hulp van, nu wel specialistische, vrijwilligers. Het Instituut voor Nederlandse Lexicologie uit Leiden presenteerde dit voorjaar binnen het project het ontwerp van een databank met *Named Entities* (historische namen van plaatsen, personen enzovoort) die zal dienen voor de verbetering van de OCR resultaten. Hieraan gekoppeld is een online *Collaborative*



OCR herkenning van gotische s die op f lijkt

Environment, waar lijsten met eigenamen door specialisten kunnen worden toegevoegd.

De bibliotheken in het project werken hard aan de voorbereiding van de demonstratieprojecten, die begin 2009 moeten gaan draaien. Zo zal in de KB een proefopstelling komen van de IMPACT-tools met materiaal uit het KB-project *Dagbladen Digitaal*.

Op 6 en 7 april 2009 worden de eerste resultaten gepresenteerd op de internationale IMPACT conferentie in de Koninklijke Bibliotheek.

impact@KB.nl
www.impact-project.eu



Brief van Vincent van Gogh aan Theo van Gogh (met 3 schetsen), Van Gogh Museum, Amsterdam

Taal- en Tekstdatabanken in kaart gebracht

DANS (Data Archiving and Networked Services) is dit voorjaar begonnen met het project 'Inventarisatie Taal- en Tekstdatabanken in Nederland'. Doel is om aanbevelingen op te stellen voor het opzetten, gebruiken en duurzaam bewaren van databanken op het gebied van taal- en tekst in Nederland. Daartoe worden de bestaande databanken op dit terrein geïnventariseerd. Ook worden de wensen van de Nederlandse taal- en letterkundigen ten aanzien van mogelijke verbeteringen in de data-infrastructuur van dit onderzoeksgebied in kaart gebracht. Zijn er leemtes en zo ja welke? Hoe is de situatie, gewenst en feitelijk, ten aanzien van

de langetermijn-bewaring van deze databanken?

Deze inventarisatie richt zich op alle in Nederland bestudeerde talen, maar beperkt zich wel tot bestanden die gevormd zijn in een *academische* onderzoeksomgeving. De belangrijkste werkzaamheden zijn deze inventarisatie, in samenwerking met het Europese project CLARIN (www.clarin.eu), en het afnemen van een beperkt aantal diepte-interviews met een representatieve groep onderzoekers 'in het veld'. Het project wordt in het najaar van 2008 afgesloten met een publicatie en een workshop met onderzoekers in het veld. Informatie: heiko.tjalsma@dans.knaw.nl