

KB en Google plaatsen tachtigduizend boeken uit bibliotheek online

# ‘Digitaal boek is godsgeschenk’



Natuurkundige afbeelding, één van de extra gedigitaliseerde beelden credits KB, Den Haag

**Zoekgigant Google heeft met een aantal bibliotheken contracten gesloten om (een deel van) hun collectie online doorzoekbaar te maken. Zo ook met de KB. Peter Boot**

Op 3 april presenteerde de Koninklijke Bibliotheek de eerste resultaten van haar samenwerking met Google: via boeken.kb.nl zijn de eerste tachtigduizend boeken beschikbaar. Het is de helft van de collectie van 1700 tot 1870 die de KB door Google laat digitaliseren. Tijdens deze presentatie maakte KB-directeur Bas Savenije duidelijk dat zonder de samenwerking met Google dit tempo nooit zou worden gehaald.

De boeken waren al beschikbaar op de site van Google Books. De presentatie van de boeken op de KB-site is een eerste stap naar een geïntegreerd aanbod van digitaal materiaal, waarbij verschillende boekencollecties, samen met de gedigitaliseerde kranten en tijdschriften, zullen worden aangeboden. Eerder dit jaar presenteerde de KB al een digitale collectie van tachtig gedigitaliseerde tijdschriften uit de periode 1850-1940.

**Nieuwe vormen van onderzoek**  
Feestredenaar Inger Leemans (VU) benadrukte hoezeer het historisch onderzoek verandert door digitale beschikbaarheid van het onderzoeksmateriaal, en schetste hoe elektronische

beschikbaarheid geheel nieuwe vormen van onderzoek mogelijk maakt. Overigens is daarvoor vaak wel een betrouwbare elektronische tekst nodig, terwijl digitalisering à la Google niet verder gaat dan ongecorrigeerde OCR (*optical character recognition*). Wat dat betreft blijft er nog het nodige te wensen over.

Onderzoeker Willem Kuiper (UvA) is ook blij met Google Books en andere digitale bibliotheken. “Voor het REMLT (Repertorium van Eigennamen in Middel nederlandse Literaire Teksten) zoeken wij wekelijks het internet af en zien tot onze verbazing dat opeens een boek digitaal beschikbaar is. Je kunt zo’n boek online lezen, maar veel beter is de .pdf die je mag downloaden, waarop je vervolgens OCR loslaat. Vaak is het resultaat verbluffend. Het is een godsgeschenk om zo je eigen digitale handbibliotheek te kunnen opbouwen”, aldus Kuiper.

## Boodschap voor de minister

Een schrijnend probleem, dat door de publicatie van deze collectie nog eens wordt benadrukt, is de tekortschietende digitale beschikbaarheid van modern materiaal, waar in veel gevallen nog auteursrecht op rust. In zijn toespraak beklemtoonde Bas Savenije nog eens hoe belangrijk het is de auteurswet zo aan te passen, dat digitale bibliotheken ook materiaal waarop nog rechten rust kunnen aanbieden. De staatssecretaris had zich bij de bijeenkomst laten excuseren, maar laten we hopen dat die boodschap bij het ministerie is aangekomen.

[boeken.kb.nl](http://boeken.kb.nl)

[books.google.com](http://books.google.com)

CEDAR-project onderwerpt Volkstellingen aan nieuwe techniek

## Linked Data zorgen voor meer dan brood alleen

**Het CEDAR-project probeert met Linked Data de historische volkstellingen beter bruikbaar voor onderzoek te maken. René van Horik**

Hoeveel bakkers telde Nederland in 1795? Een interessante vraag voor sociaal-historisch onderzoek. Het antwoord ligt in de data van de Nederlandse Volkstellingen. Deze tellingen, van bevolkingsgrootte, leeftijden, geslacht en beroepsgroepen, zijn van 1795 tot 1971 gehou-

den. Inmiddels zijn de tabellen van deze zeventien tellingen grotendeels gedigitaliseerd met het project ‘Digitalisering Nederlandse Volkstellingen 1795-1971’.

De tellingen zijn een fantastische bron voor onderzoek. Maar de oorspronkelijke tabelstructuur kent ook beperkingen. Door de jaren heen veranderen bijvoorbeeld plaatsnamen, fuseren gemeenten en worden beroepen anders ingedeeld. Handmatig zijn datasets uit tabellen nog wel te analyseren, maar dit kost

veel tijd. Bieden Linked Data-technieken een uitkomst?

### Gewenste koppelingen

Linked Data beschrijven de relatie tussen objecten. Zo kun je met Linked Data-technieken beschrijven dat het beroep ‘bakker’ in de ene telling een relatie heeft met het beroep ‘broodbakker’ uit een andere telling. Als alle variabelen van de Volkstellingen als Linked Data kunnen worden uitgedrukt (in een bestandsformaat waar een computer mee overweg kan), dan zijn in theorie de gewenste koppelingen te maken.

Het CEDAR-project onderzoekt de mogelijkheden om dit beschrijven van variabelen automatisch te harmoniseren en zodoende de tabellen van de tellingen beschikbaar te stellen als Linked Data. Belangrijk is dat de oorspronkelijke gegevens en tabelstructuur niet verloren mogen gaan. Bij de harmonisatie wordt gebruik gemaakt van externe classificaties, bijvoorbeeld voor histori-



**Bakker met bakkershoorn door Adriaan de Lelie (1755 - 1820) credits Nederlands Bakkerijmuseum**

sche beroepen. De eerste resultaten van deze omzetting van de Volkstellingen naar Linked Data-structuren staan op de CEDAR-website.

### Fantastische methode

Het CEDAR-project laat zien dat de vele uitzonderingen, definitieveranderingen en vaagheden in de oorspronkelijke data verder onderzoek noodzakelijk maakt. Maar dat met

Linked Data voor onderzoekers een fantastische methode voor onderzoek beschikbaar komt, is wel bewezen. Handmatige voorbewerking van data wordt tot een minimum beperkt. Gegevens zijn veel gemakkelijker te verbinden met andere Linked Data-bronnen. En bovendien kunnen gegevens sneller worden gevisualiseerd.

[cedar-project.nl](http://cedar-project.nl)

### CEDAR (Census data open linked)

CEDAR staat voor ‘Census data open linked’, en is een samenwerkingsverband tussen de Vrije Universiteit en de KNAW-instituten IISG (Internationaal Instituut voor Sociale Geschiedenis) en DANS. CEDAR werkt aan het semantisch ontsluiten van de data van de Nederlandse Volkstellingen uit de laatste twee eeuwen. Deze data worden met behulp van RDF (*Resource Description Framework*) interoperabel gemaakt met andere hubs van historische socio-economische en demografische gegevens. Het hierdoor opgebouwde semantisch web-based datamodel maakt het mogelijk om vragen te beantwoorden op het gebied van regionale identiteit, technologische vooruitgang en geografische migratie.