



Databank Digitale Dagbladen

Astrid Verheusen

Projectmanager

Afdeling Research & Development

Koninklijke Bibliotheek

Belemmert het auteursrecht de ontsluiting van de 20^{ste} eeuw?
Vereniging voor Auteursrecht - 18 januari 2008

Databank Digitale Dagbladen - Overzicht

- Doelen project
- Achtergronden (kranten-)digitalisering bij de KB
- Aanpak
 - Projectmanagement
 - Selectie
 - Materiaalanalyse
 - Digitalisering & OCR
 - Ontsluiting
 - Opslag & beschikbaarstelling
- Gevolgen auteursrecht voor project

Databank Digitale Dagbladen - Doelen

- Digitalisering, ontsluiting en gratis beschikbaarstelling van belangrijke nationale, koloniale, regionale en lokale dagbladen
- 8 miljoen pagina's (25 miljard woorden)
- 1618 - eind twintigste eeuw
- Doelgroepen:
 - Wetenschap
 - Algemeen publiek
- Doorlooptijd: 2007 - 2011
- Financiering: M€ 12,5 Nationaal Programma Investerings in Grootschalige Onderzoeksfaciliteiten

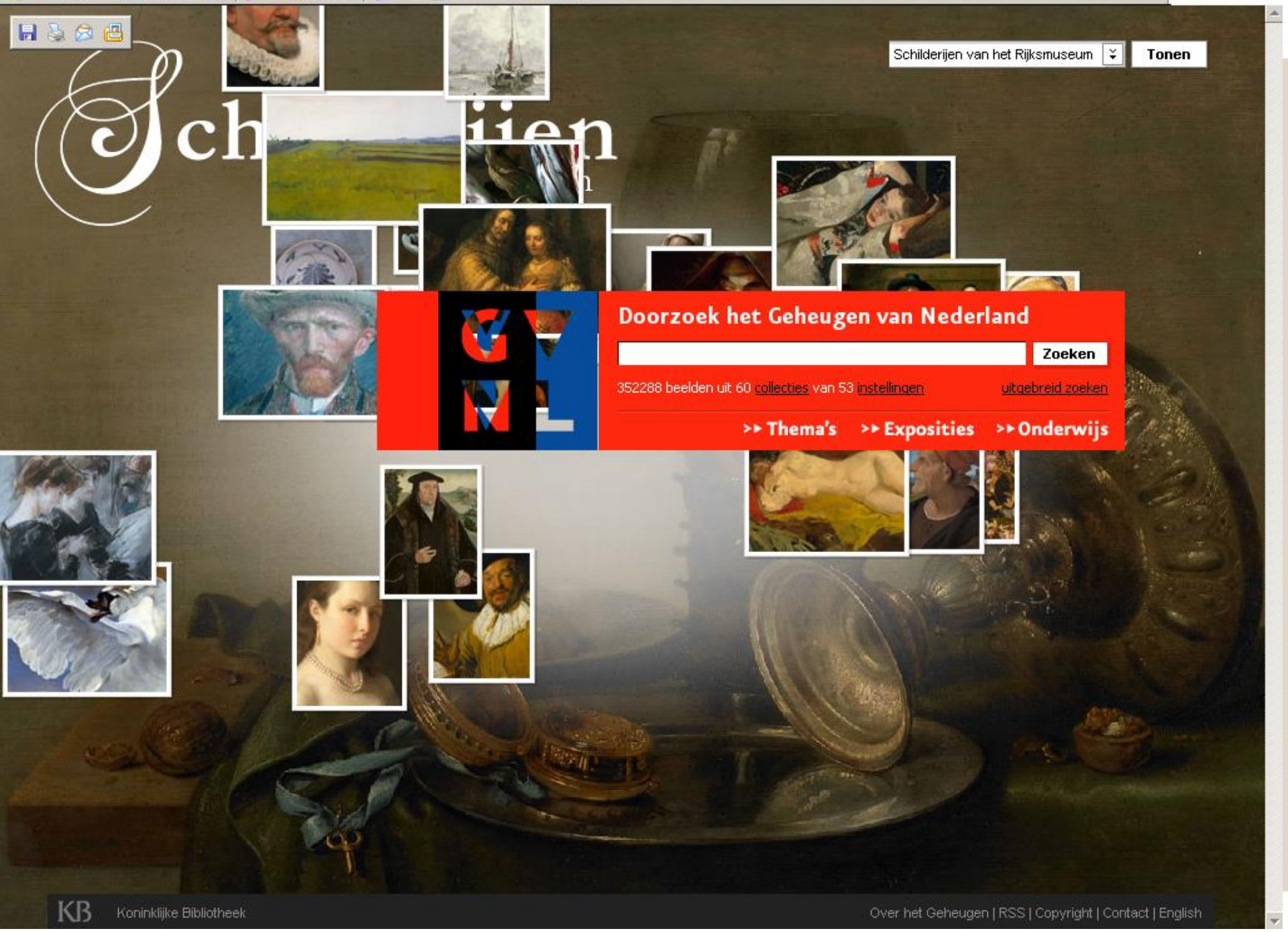
Digitalisering in de KB - verleden

- Sinds midden jaren '90: veel ervaring opgedaan met digitalisering
- Webexposities / hoogtepunten uit collecties
- Voornamelijk beeldmateriaal
- Kleinschalig: kleine en middelgrote collecties
- Uitproberen / grote nadruk op techniek
- Mogelijkheden verkennen
- Samenwerking zoeken



Digitalisering in de KB - 2000-2005

- Grootschalig: grote collecties
- Virtueel bijeenbrengen van verspreide bronnen
- (Inter)nationale samenwerking / sectoroverschrijdend
- Standaarden in methoden en technieken
- Aandacht voor duurzaamheid
- Verkennen van toepassingen
 - Onderwijsmodulen
 - Services



Schilderijen

Schilderijen van het Rijksmuseum ▼ **Tonen**

Doorzoek het Geheugen van Nederland

Zoeken

352288 beelden uit 60 [collecties](#) van 53 [instellingen](#) [uitgebreid zoeken](#)

[>> Thema's](#) [>> Exposities](#) [>> Onderwijs](#)



DE SPOORWEGARBEIDERS

DE STAKING

SUCCES! OF NIET?

STAKINGSOPDRACHT

LOGIN

Naam

Klas



NAAR BRONNENLIJST

 DE SPOORWEGSTAKING VAN
1903

DE SPOORWEGSTAKING VAN 1903

De Nederlandse Spoorwegen (NS) zijn op zoek naar nieuw personeel. Dat wil niet zo lukken. De NS-personeelsleden van nu zijn ontevreden. Soms zo erg, dat ze weigeren verder te werken. Ze gaan dan staken. Door een staking laten ze merken dat het zo niet langer kan. Ze eisen van de directie verbetering van hun werk. Ze sturen vertegenwoordigers uit hun eigen organisaties - de vakbonden - om daarover te praten.

De eerste grote spoorwegstaking was in 1903, zo'n honderd jaar geleden. De arbeiders lieten toen voor het eerst hun macht zien. Ze legden Nederland plat. De regering werd toen erg zenuwachtig. Er kwamen wetten om stakingen voortaan te verbieden. Leger en politie werden de straat opgestuurd. Het leek wel oorlog!



Er zijn veel bronnen over die staking van een eeuw geleden. Daarmee ga je uitzoeken wat er toen gebeurde. Je moet de bronnen dus goed bekijken. Je kunt alleen werken of in een groepje. In [De spoorwegarbeiders](#) bekijk je hoe de spoorwegarbeiders werkten. In [De staking](#) lees en bekijk je het verhaal over de staking zelf. In [Succes! of niet?](#) zie je hoe de staking uitpakte voor de mensen en het land. Als je de thema's bekeken hebt kun je de stakingsopdracht maken.

Digitalisering in de KB - heden en toekomst

Massadigitalisering van (voornamelijk) tekstmateriaal

- Handelingen Staten-Generaal Digitaal 1814-1995
- Digitalisering Bijzondere Collecties
- Nieuwsberichten ANP
- Metamorfoze
- Geheugen van Nederland
- Databank Digitale Dagbladen

Tekstdigitalisering (1)

- Tot nu toe op kleine schaal
- Andere problemen dan bij beeldmateriaal:
 - Structuur/Zoeken/Navigatie
 - Conversie naar full-tekst met optische tekenherkenning (OCR)
 - Bij grote collecties: scannen vanaf microfilm of origineel?
 - Auteursrecht: andere dimensie dan bij kleinschalige projecten

Tekstdigitalisering (2)

Nut

- Beschikbaarstelling en ontsluiting
- Zoeken op ieder woord in de tekst
- Tendens: digitalisering t.b.v. behoud
- Google!
 - Als bibliotheken hun collecties niet zelf digitaal beschikbaar stellen, doet Google het wel

Digitalisering kranten

- Belangrijke bron voor wetenschap
- Krantenpapier bedoeld voor eenmalig gebruik
- Microfilms alleen on site te bekijken
- Microfilms niet gebruiksvriendelijk
- Materiaal is verspreid bewaard
- Verwacht gebruik is groot

Achtergrond project

- **Sinds jaren zeventig** microverfilming van kranten
 - **1998-2001** Microverfilming landelijke kranten (3 miljoen pagina's)
 - **2004-2005** Microverfilming regionale kranten (1 miljoen pagina's)
- **1999** *Roaring twenties*, digitalisering drie landelijke dagbladen
- **2002-2004** *Pilot Oorlog & Revolutie*, digitalisering drie landelijke dagbladen uit 1910-1920 en het Vaderland 1930-1945

Historische kranten in beeld: kranten.kb.nl

Krantencollectie - Microsoft Internet Explorer

Bestand Bewerken Beeld Favorieten Extra Help

Vorige Zoeken Favorieten

Adres <http://kranten.kb.nl/index2.html> Links »

Historische kranten in beeld

Home Introductie **Kranten 1910-1945** Artikelen 1918

Zoeken Bladeren Zoeken Bladeren

Kranten 1910-1945

Zoekterm [help](#)

Editie Alle [help](#)

Datum Alle [help](#)

Binnen zoektermen wordt gezocht op
 Alle woorden [help](#)

Perk uw zoekactie in op krantentitel

☒ het Centrum (1910-1930)

☒ de NRC (1910-1930) [help](#)

☒ het Vaderland (1920-1945)

☒ het Volk (1910-1920)

Afwijking in de spelling:
 geen [help](#)

Koninklijke Bibliotheek - *Nationale bibliotheek van Nederland*

Gereed Internet

Zeg nu maar, dat een winkelier zijn
geld zoo maar verdient.... en wat een
Nederlandsche taalkennis hij wel moet
bezitten.

On geloofelijk, welke eigenschappen en
kennis een winkelier zich moet aan-
kweeken om z'n zaak bloeiend te maken.

Op de eerste plaats: goede inkoop,
goede etaleur- en reclamekunst en ver-
der een reuzen-geduld en een ontzettend
aanpassingsvermogen, om alle klanten
gemakkelijk en vlug te bedienen.

Aanpak Databank Digitale Dagbladen

- Ervaring uit pilot-projecten (selectie, materiaalanalyse, digitalisering en OCR, ontsluiting, website, gebruikte software)
- Aanpak
 - Projectmanagement
 - **Selectie**
 - Materiaalanalyse
 - Digitalisering & OCR
 - Ontsluiting
 - Opslag & beschikbaarstelling

Projectmanagement

- Prince2
- Projectteam (10 fte)
 - Projectleiders voor selectie, materiaalanalyse en digitalisering
- Samenwerking met
 - Onderzoeksinstituten & universiteiten
 - Bibliotheken en archieven
 - Uitgevers
- 2007: voorbereiding
2008: eerste resultaten

Selectie (1)

- Geen complete catalogus voor Nederlandse dagbladen
- Collecties zijn incompleet en verspreid
- Wisselende kwaliteit van microfilms
- Betrekken van wetenschap in het selectieproces
- Auteursrecht
- Lokale initiatieven voor krantendigitalisering

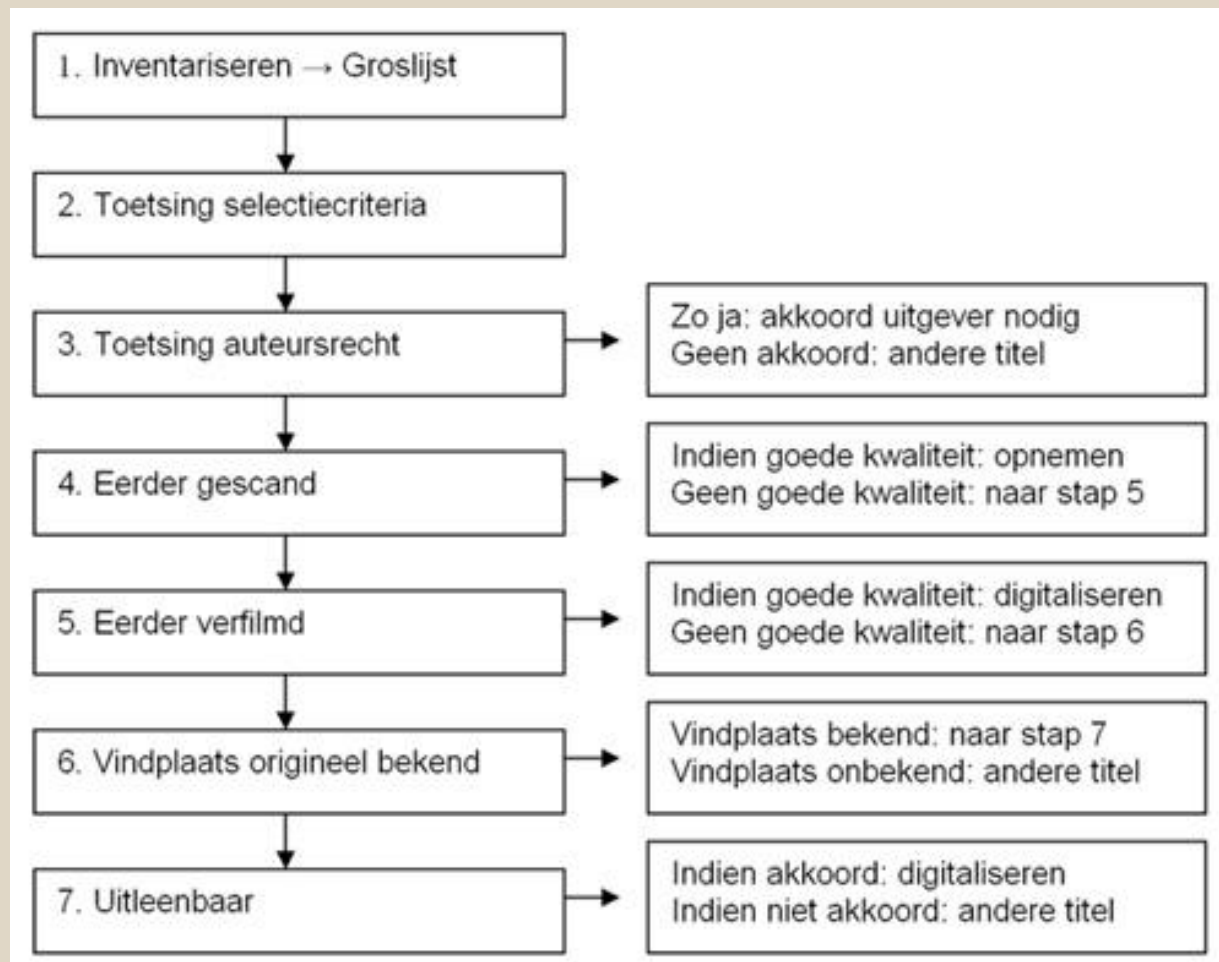
Selectie (2)

- Inventarisatie titels in database
- Wetenschappelijke Adviescommissie
- Stappenplan voor het selectieproces
- Samenwerking met uitgevers

Selectie (3)

- Indeling in tijdvakken
 - 1618-1800
 - 1800-1813
 - 1813-1869
 - 1869-1914
 - 1914-1965
 - 1965-1995
- Selectiecriteria per periode

Selectie (4)



Selectie (5)

- Toetsing auteursrecht
 - Op titelniveau
 - Voorgangers en opvolgers van titels bepalen
 - Eventuele rechthebbenden achterhalen
 - Ca. 1200 titels: ca 1 uur per titel
- Achterhalen rechthebbenden van artikelen: onmogelijk
 - 20^{ste} eeuw: ca. 10 miljoen artikelen en foto's...

Selectie (6)

- Samenwerking met uitgeefsector
 - Gesprekken met de Groep Nederlandse Dagbladpers (NDP)
 - Scenario's voor samenwerking
 - Uitwisselen bestanden (papier en digitaal)
 - Afspraak over een bepaald jaartal
 - Afspraken over uitsluiten van bepaalde titels
- Graag ook samenwerking met auteursorganisaties

Materiaalanalyse

- Overzicht van bewaarplaatsen
- Database met informatie per geselecteerde titel
- Verzamelen materiaal
- Beperkte inzet voor completering
- Repareren originelen
- Beoordelen kwaliteit van microfilms

Digitalisering & OCR

- Scannen voor lange termijn
- Scannen in grijswaarden
- Scannen op artikelniveau (afzonderlijke artikelen kunnen eventueel worden verwijderd)
- Linken van artikels op meerdere pagina's
- Onderzoek naar verbeteren OCR
- Europese aanbesteding

Ontsluiting

- Inventarisatie wensen wetenschap
- Metadata (titel, editie, datum, koppen)
- Onderzoek automatische classificatie en samenvattingen
- Onderzoek spellingsvarianten / fuzzy zoeken
- Navigatie
- Andere zoekmogelijkheden?

Opslag en beschikbaarstelling

- Website: open standaarden
- Aanpassingen workflow en infrastructuur t.b.v. massadigitalisering
- Opslag: 250 Tb
- Duurzame opslag in e-Depot (systeem voor lange termijn opslag)

Belemmert het auteursrecht de ontsluiting van de 20^{ste} eeuw?

Opties

- Afspraken op titelniveau met uitgevers
- Algemene afspraken met NUV/NDP
- Artikelen
 - Opt-out?
 - Diligent search?
 - Afspraken met belangenorganisaties?
 - Eenmalig bedrag, maar hoeveel dan?

Belemmert het auteursrecht de ontsluiting van de 20^{ste} eeuw?

- Onzeker..
- Het vertraagt het in ieder geval
- Mogelijk gevolg: geen toegang tot kranten uit de 20ste eeuw

Vragen?

astrid.verheusen@kb.nl