

Een databankmodel voor fotocollecties

Bij het Digitaal Productiecentrum (DPC) van de Universiteitsbibliotheek van de Universiteit van Amsterdam is een algemeen toepasbaar model ontwikkeld voor het creëren van beelddatabanken. Afbeeldingen uit allerlei collecties en hun beschrijvingen worden op efficiënte wijze ontsloten, doorzoekbaar gemaakt en getoond. Het model bevat naast de webinterface en middleware voor de databank ook een beschrijving voor het digitaliseringproces en de workflow. Dublin Core, als overkoepelende metadatastructuur, maakt integratie van verschillende beelddatabanken mogelijk. Renze Brandsma beschrijft het UB-project en het uiteindelijke databankmodel.

BINNEN DE COLLECTIES van de Universiteit van Amsterdam (UvA) is veel beeldmateriaal aanwezig dat uitermate geschikt is om in digitale vorm via internet toegankelijk te maken. Denk bijvoorbeeld alleen al aan de kaarten en atlanten in de cartografische collectie van de Universiteitsbibliotheek (UB). Door deze afbeeldingen te digitaliseren en te koppelen aan de tekstuele gegevens over deze afbeeldingen (in sgml/xml) is het mogelijk om via een webinterface een beelddatabank te creëren. Het algemeen toepasbaar model voor het creëren van beelddatabanken is ontwikkeld aan de hand van de collectie portretten (schilderijen en beelden) van het Universiteitsmuseum. Deze collectie bevat onder meer 325 schilderijen, 22 tekeningen, 75 borstbeelden en 36 plaquettes.

Wat is er mogelijk in de beelddatabank

De beelddatabank is te vinden via de digitale bibliotheek van de UB, of door direct naar de webadressen van de bestanden te gaan (zie kader op pagina 20).

In de beelddatabank is er voor iedere enkele collectie, zoals de portrettencollectie, een apart zoekscherm. Er kan dan gezocht worden op de specifieke velden van deze collectie. De vormgeving van de webinterface kan per collectie anders zijn.

Voor het zoeken in meerdere collecties tegelijkertijd gaat men naar een speciaal ingericht zoekscherm, waar alle collecties staan vermeld met een optie om deze wel of niet mee te nemen in de zoekactie. Standaard worden alle collecties meegenomen. Bepaalde collecties kunnen uitgesloten worden door deze 'uit te vinken'. Zo kan men bijvoorbeeld maar in drie van de tien collecties zoeken als men dat wenst. Er wordt alleen op de Dublin Core-velden gezocht. Er is voor het zoeken in meerdere collecties tegelijkertijd maar één zoekscherm waarvan de vormgeving vaststaat.

Bij het zoeken in beide opties kan gekozen worden om in alle velden te zoeken of de zoekactie te limiteren tot bepaalde velden, zoals auteur of titel. Er kan eventueel een alfabetische woordenlijst opgeroepen worden, waaruit men een woord kan kiezen om verder de zoekactie mee te vervolgen.

Het zoekresultaat wordt naar keuze getoond in een lijst met titels of in een lijst met thumbnails en titels. Door op een item te klikken in het zoekresultaat verkrijgt men de volledige beschrijving en een afbeelding van het portret in klein formaat. In de volledige beschrijving worden de woorden waarop gezocht is benadrukt. In een beperkt aantal gevallen wordt geen afbeelding getoond, omdat de betreffende rechten niet verworven zijn.

Men kan vanuit de volledige beschrijving een nieuwe zoekactie opstarten door te klikken op de gekleurde woorden in de velden vervaardiger, vakgebied of opmerkingen. Door bijvoorbeeld bij vakgebied op wijsbegeerte te klikken wordt een lijst verkregen van alle beschrijvingen waarbij het woord wijsbegeerte in vakgebied voorkomt. Tevens kan bij sommige beschrijvingen een link gevolgd worden naar meer uitgebreide thematische informatie.

Op het portret kan men inzoomen door een zoomfactor te kiezen en daarna op het gewenste detail in de afbeelding te klikken. Met de grootste zoomfactor wordt vaak de craquelé van het schilderij duidelijk zichtbaar. Daarnaast is het mogelijk om de afmetingen van de getoonde afbeelding op het scherm te vergroten tot een middelgroot formaat (circa 500 pixels breed) of een groot formaat (circa 1000 pixels breed).

In de beelddatabank is een optie om het zoekresultaat als een 'slide show' te tonen. Daarnaast is er ook de moge-

lijkheid om portretten in twee aparte windows naast elkaar te vergelijken. Verder is er een formulier om informatie aan te vragen over het bestellen van reproducties.

Het digitaliseren van de afbeeldingen

Van de portretten zijn door een professionele fotograaf 4 x 5 inch kleurendia's vervaardigd, die voor het digitaliseren gebruikt worden. Van elke kleurendia is een master of archiefbeeld gemaakt. Van de master zijn weer afgeleiden gemaakt in het MrSID (Multi Resolution Seamless Image Database) formaat voor presentatie in de beeldbank en jpeg-thumbnails voor het snel tonen van het zoekresultaat.

Het vervaardigen van kwalitatief goede digitale afbeeldingen is tijdrovend, duur en specialistisch werk. Voorkomen moet worden dat de kleurendia's op een later

moment opnieuw gedigitaliseerd dienen te worden. Daarom is het verantwoord tijdens het digitaliseringsproces een master of archiefbeeld te maken. De masters worden op hoge resolutie gescand en ongecomprimeerd opgeslagen in tiff-formaat. De belangrijkste functie van een master image is om als archiefexemplaar te dienen en als bron voor de afgeleide beelden in de beeldbank. De master kan ook gebruikt worden voor reproductie of drukwerk.

Het scannen gebeurt met een hoge kwaliteit flatbed scanner (Saphir Ultra 2), gekoppeld aan een zeer krachtige Macintosh-computer met als kleurenmanagementsysteem LinoColor. Deze hoogwaardige en ultramoderne apparatuur is recentelijk aangeschaft door de UB. Het Rijksmuseum in Amsterdam gebruikt een vergelijkbaar systeem. De diapositieven zijn gescand op 1080 dots per



FOTO: EGON VIEBRE

De collectieve portretten

Zoals vele universiteiten bezit de UvA een grote collectie portretten. Deze collectie werd begonnen in 1743, toen de Amsterdamse koopman Gerard van Papenbroeck ruim vijftig portretten van 'Geleerde en Vermaarde mannen' naliet aan het Athenaeum Illustre, de voorloper van de UvA. In de collectie bevonden zich portretten van de eerste hoogleraren van het Athenaeum, Gerard Johannes Vossius, Caspar Barleus en Simon Episcopus. Vanaf de tweede helft van de achttiende eeuw werd de collectie uitgebreid, aanvankelijk met enkele portretten van grote geleerden, later vooral met portretten van hoogleraren van het Athenaeum Illustre.

In 1877 werd het Athenaeum omgevormd tot universiteit. De belangstelling voor de collectie nam toe en op aandringen van professor J.W.R. Tilanus werden alle nabestaanden van hoogleraren aangeschreven met het verzoek een portret te schenken. Dit leidde tot de aanwinst van enkele tientallen kunstwerken. Schilders als J.A. Kruseman, N. Pieneman en Jozef Israëls werden in de collectie vertegenwoordigd.

De bloei van de universiteit rond het jaar 1900 bracht vele beroemde hoogleraren voort die allen geportretteerd werden.

Er ontstond een traditie waarin hoogleraren hun eigen portret nalieten aan de universiteit, wat leidde tot grote toename van de collectie. In de eerste helft van de twintigste eeuw werden meer dan honderd werken vervaardigd, onder anderen door belangrijke Nederlandse kunstenaars als Isaac Israëls, Jan Veth, Thérèse Schwartze, Jan Toorop, W. Schuhmacher, Edgar Fernhout en Georg Rueter. Ook in de tweede helft van de twintigste eeuw werd veel werk vervaardigd door contemporaine kunstenaars, onder wie Willem Gerard Hofker, Kees Verwey en Johan Wertheim. De traditie wordt nog steeds voortgezet. Tegenwoordig worden er naast hoogleraren ook voorzitters van het College van Bestuur geportretteerd.

De groei van de universiteit en de verspreiding over vele gebouwen in de stad heeft geleid tot versnippering van de collectie. Daarnaast is de omvang van de collectie zo toegenomen (momenteel 490 portretten) dat plaatsing in één gebouw, zoals voorheen, onmogelijk is geworden. In de beeldbank is de volledige collectie samengebracht. Deze databank biedt de unieke mogelijkheid alle portretten, waarvan een groot deel niet openbaar toegankelijk is, bij elkaar te zien.

inch (dpi). De tiff-bestanden (master/archief) van circa 50 Mb worden tijdelijk op de Mac opgeslagen. Van de tiff-beelden zijn MrSID-afgeleiden gemaakt met behulp van compressiesoftware van LizardTech. De tiff-bestanden worden gecomprimeerd met een factor 20 zonder dat verlies van informatie door het menselijk oog merkbaar is. Verder integreert deze compressiesoftware verschillende resoluties van een digitaal beeld in één enkele file, waardoor internetgebruikers steeds verder kunnen inzoomen en meer en meer detail te zien kunnen krijgen. Tijdens het scannen zijn ook de technische gegevens en de bestandsnamen opgeslagen in een databank voor intern gebruik met behulp van het commercieel verkrijgbare softwarepakket Cumulus. Cumulus is een archiefsysteem voor het beheren van grote aantallen digitale afbeeldingen en multimedia. Bepaalde gegevens van digitale afbeeldingen, zoals resolutie, type compressie, worden automatisch in het systeem opgeslagen. De te volgen workflow kan ingesteld worden en er wordt bijgehouden waar de afbeeldingen op een cd-rom of ander medium staan. Daarnaast kunnen eigen velden worden gedefinieerd. Bij de UB worden onder andere de gebruikte scanapparatuur, de operator, de scandatum en een inventarisnummer van het object bijgehouden. De richtlijnen van technische metadata van digitale afbeeldingen van de Research Libraries Group (RLG) worden daarbij gevolgd.

Nadat het scannen is afgerond, zijn de MrSID-bestanden en de thumbnails geplaatst op de server voor de web-interface. De tiff-bestanden zijn als archief op cd-rom opgeslagen.

Auteursrecht

Een belangrijk aspect bij het creëren van de beeldbank is het auteursrecht. De rechthebbenden voor de portretcollectie zijn:

- maker origineel;
- nabestaanden maker, tot zeventig jaar na het overlijden van de maker;
- maker of opdrachtgever reproductie, tot zeventig jaar na overlijden van de maker;
- afgebeelde personen;
- nabestaanden afgebeelde personen, tot tien jaar na overlijden afgebeelde personen.



Het Universiteitsmuseum heeft in een contract met de fotograaf de rechten van hem overgenomen. Voor meer dan tweehonderd portretten was het benodigd om toestemming te verkrijgen van de andere rechthebbenden, om de afbeeldingen op het internet te tonen. Het mag duidelijk zijn dat dit een enorm inspanning heeft gevergd van de medewerkers van het Universiteitsmuseum. Indien geen toestemming is verleend of de rechthebbenden niet zijn achterhaald wordt het portret alleen getoond via het intranet (UvA-domein) en niet wereldwijd.

Technische infrastructuur

De basis voor de databank is een *sgml/xml*-zoekmachine. Met deze zoekmachine kan zeer snel tekstuele informatie doorzocht worden. De grote kracht van deze zoekmachine ligt in het feit dat specifiek gezocht kan worden op de inhoud van elementen en/of attributen uit de *sgml/xml*-documentstructuur.

SGML/XML is gericht op het benoemen van inhoudelijke elementen in een tekst door middel van markeringslabels of 'tags'. Zo kan de titel in een tekst als volgt gemarkeerd worden: `<titel> Portretten in beeld </titel>`. De zoekmachine kan, als de gebruiker dit aangeeft, het zoekresultaat limiteren tot de titel. SGML/XML is een formaatkeuze die toekomstgericht is en al jaren bestaat als een internationale standaard. De voordelen zijn platformafhankelijkheid, flexibiliteit, scheiding van inhoud, structuur en presentatie en makkelijke uitwisselbaarheid. De elementen in een document worden gedefinieerd door de gebruikte *document type definition (dtd)*. Naast de verworven zoekmachine is ook middleware aanwezig, die de communicatie met de zoekmachine en de webinterface verzorgt. De middleware zendt de zoekvragen naar de zoekmachine, ontvangt de gevonden resultaten en vertaalt deze naar een webpagina. Per klasse digitale gegevens of documenten is specifieke middleware en een *dtd* aanwezig. De middleware en *dtd* vormen samen een zogenaemde 'Class'. Voor full-text bestanden is er bijvoorbeeld een 'Text Class', waarbij de *dtd* gebaseerd is op *tei (text encoding initiative)*. Voor de beeldbank wordt gebruik gemaakt van de 'Image Class', met een *dtd* die speciaal ontwikkeld is voor de vereiste functionaliteiten van een beeldbank.

Conversie naar *sgml/xml*

De beschrijvingen van de portretten staan in het museumformaat van het bibliotheekstelsel van Pica, dat

gebruikt wordt voor het catalogiseren van collecties. Deze beschrijvingen van de portretten zijn uit dit systeem gehaald en geconverteerd naar een *sgml*-formaat volgens de 'Image Class'. Het zo verkregen *sgml*-bestand wordt gevalideerd en op de server geplaatst, waar ook de zoekmachine en middleware staat.

De conversie vindt plaats in een Access-database, waarin de portretbeschrijvingen uit Pica en de gegevens voor de bestandsnamen uit Cumulus worden geïmporteerd en samengevoegd. Na deze samenvoeging worden de specifieke velden uit de beschrijvingen in een bepaalde collectie gekoppeld aan de overeenkomende *sgml*-elementen. Op deze wijze kan een *sgml*-bestand worden gegenereerd.



Naast het zoeken in een enkele collectie is het ook gewenst om over meerdere collecties tegelijkertijd te zoeken. Er zijn echter soms grote verschillen in de beschrijvingen van collecties. Een kaartcollectie heeft andere velden dan een collectie portretten. Dublin Core is een algemeen model dat voor beide collecties gebruikt kan worden. Een van de elementen is bijvoorbeeld 'Creator'. Zowel de schilder van een portret als de maker van een kaart wordt in het Dublin Core-element 'Creator' geplaatst.

Voor het tegelijkertijd zoeken in meerdere collecties worden aan het *sgml*-record Dublin Core (DC) metadata toegevoegd, om op deze standaardvelden te kunnen zoeken. De toekenning vindt in de conversie automatisch plaats, nadat van tevoren is vastgelegd welke specifieke velden overeenkomen met de vijftien DC-elementen.

Het proces van de conversie is zo opgezet dat het in principe voor iedere collectie met afbeeldingen op relatief eenvoudige technische wijze plaats kan vinden.

Specifiek aspect voor musea

Wat voor veel beheerders in de museumwereld een interessant aspect is, is dat men kan blijven invoeren in de eigen database (indien deze van voldoende kwaliteit, uitwisselbaar enzovoort is). In de museumwereld is decennia lang gepoogd iedereen op eenzelfde database over te laten gaan (vergelijkbaar met Pica voor de bibliotheekwereld), met als doel de gehele 'Collectie Nederland' centraal doorzoekbaar te laten maken (zoals in de NCC). Dit is echter onhaalbaar gebleken; eenieder heeft zo zijn eigen systeem en is niet bereid of in staat aan te sluiten bij een groot systeem.

De conversie van de inhoud uit willekeurige databases naar een *sgml*-formaat met gebruikmaking van Dublin

Core zorgt voor de oplossing van dit probleem; zo kan je beschikken over een eigen beheerssysteem én een centrale databank voor het publiek.

Vervolg

Bij de ontwikkeling is gebruikgemaakt van infrastructuur, methoden en hulpmiddelen die het DPCstandaard wil gebruiken voor tekstuele, bibliografische en beeldbestanden, waaronder sgml/xml, Dublin Core-metadata en een sgml/xml search engine. Deze search engine vormt het hart van de verkozen technische infrastructuur. Er is in het project ruime ervaring opgedaan met de implementatie en mogelijkheden van deze krachtige zoekmachine. Daarnaast is veel praktische ervaring en kennis opgedaan in het digitaliseren van dia's en foto's. Het ontwikkelde model voor beelddatabanken zal gebruikt worden voor diverse andere interessante collecties van de UvA. Na de portrettencollectie zal vervolgd worden met een selectie van kaarten en atlanten uit de cartografische collectie van de UB en de collectie botanische platen van het Universiteitsmuseum. Andere collecties zullen volgen. Ook organisaties buiten de UvA kunnen voor een beelddatabank bij het DPC terecht

Voor de ontwikkeling van het model is 15.730 euro eenmalig geïnvesteerd (exclusief kosten scannen dia's en aanschaf/support zoekmachine). Door deze eenmalige investering is het mogelijk om op een snelle en efficiënte wijze nieuwe beelddatabanken te creëren. Indien de beschrijvingen in Pica of een willekeurig ander beheerssysteem aanwezig zijn, zal het maximaal vijf dagen kosten om een beelddatabank op het web beschikbaar te stellen. De grootste kostenpost is dan het digitaliseren (en eventueel fotograferen) van het materiaal als het om een groot aantal items gaat. In het geval van de relatief

kleine portrettencollectie waren de kosten voor het maken van de master en de afgeleiden circa 5.700 euro (11,40 euro per dia).

Deze kosten maken het noodzakelijk om zorgvuldige keuzes te maken van de collecties die worden gedigitaliseerd. Deze keuzes moeten gemaakt worden volgens een goed overdacht en van tevoren bepaald beleid. Stapsgevoel kan dan toegewerkt worden naar omvangrijke en waardevolle digitale collecties.

De kosten zijn te rechtvaardigen als bedacht wordt dat beelddatabanken een uitstekend middel zijn om belangrijke doelstellingen van musea en bibliotheken vorm te geven. Ten eerste maakt de beschikbaarstelling van digitale collecties op het web de collecties veel meer zichtbaar en wereldwijd makkelijker en beter toegankelijk. Zoeken naar beeldobjecten in de collecties wordt dan ook mogelijk. Ten tweede zijn door een goede zichtbaarheid en heldere positionering van het cultureel erfgoed via het internet de erfgoedbeheerders in staat om – naast het vaste publiek – nieuwe publieksgroepen aan te spreken. Als derde kunnen virtuele tentoonstellingen of educatieve programma's vervaardigd worden voor gebruik bij voorlichting en cultuureducatie op het internet.

Webadressen

Portrettencollectie: <http://logos.uba.uva.nl/portrettencollectie>

Zoeken in meerdere beelddatabanken: <http://logos.uba.uva.nl/beelddatabank>

Digitale Bibliotheek: <http://www.uba.uva.nl/nl/digitalebib/>

Digitaal Productiecentrum: <http://www.uba.uva.nl/dpc>

Renze Brandsma is hoofd Digitaal Productiecentrum van de Universiteitsbibliotheek, Universiteit van Amsterdam.

