

---

## Geïntegreerde toegang tot erfgoeddatabanken



Henk Vanstappen  
Rubenianum  
Kolveniersstraat 20  
2000 Antwerpen  
Tel.: 03-201 15 77  
E-mail:  
henkvanstappen@hotmail.com

### ARTIKEL

De rol van musea, archieven en bibliotheken als bewaarders van het culturele erfgoed staat de laatste jaren prominent op de politieke agenda. Steeds meer wordt de nadruk gelegd op netwerking en samenwerking over de grens van instellingen heen. Bovendien worden zij niet alleen gewaardeerd op basis van de intrinsieke waarde van hun collecties, maar ook om de manier waarmee zij die ter beschikking stellen van wetenschappers en van het brede publiek. Dit publiek verwacht snelle en eenvoudige toegang tot de informatiebronnen die door erfgoedinstellingen bewaard worden. Binnen die optiek groeit de interesse voor een geïntegreerde toegang tot digitale gegevens, ongeacht of ze betrekking hebben op tekstuele, picturale of fysieke bronnen. In dit artikel willen we onderzoeken hoe het simultaan ondervragen van geïntegreerde bibliografische en museale databanken mogelijk is, en welke aspecten daarbij komen kijken<sup>1</sup>.

Het registreren en ontsluiten van kunsthistorische objecten vertoont een aantal overeenkomsten met registratie in bibliotheekcatalogi: elk document of object draagt een naam of titel, kan toegeschreven worden aan een persoon of instelling die verantwoordelijk was voor het ontstaan ervan, heeft kenmerkende fysische eigenschappen en bevindt zich op een bepaalde plaats. Inhoudelijke ontsluiting gebeurt door de bron met bepaalde concepten te verbinden, door het toekennen van trefwoorden of het onderbrengen in een classificatieschema. Gemeenschappelijk is dus ook dat hiermee telkens informatiebronnen ontsloten worden. Wanneer een kunsthistoricus bijvoorbeeld een bepaald kunstwerk

als onderzoeksobject heeft, zal hij geïnteresseerd zijn in de picturale en de technische kenmerken, maar ook in de wetenschappelijke kunsthistorische literatuur waar dat kunstwerk wordt besproken (catalogues raisonnés, kunstenaarsmonografieën, tentoonstellingscatalogi, veilingcatalogi...) Publicaties en kunstwerken zijn in het kunsthistorisch onderzoek noodzakelijke bronnen die in elkaars verlengde liggen.

Vanuit deze vaststelling ligt het voor de hand te onderzoeken hoe deze informatie op een geïntegreerde wijze ondervraagd kan worden. Twee elementen spelen hierbij een sleutelrol: standaarden die een uniforme terminologie voor de inhoudelijke ontsluiting bieden, en standaarden voor de structuur van bronbeschrijvingen. Verder bespreken we een aantal modellen volgens welke integratie van erfgoeddatabanken geïmplementeerd kan worden. We verwijzen daarbij naar praktijkvoorbeelden om aan te tonen hoe de modellen er in realiteit kunnen uitzien.

Integratie van verschillende datatypes kan op uiteenlopende manieren gebeuren en er zijn in de literatuur dan ook diverse definities te vinden die min of meer gelijkaardige concepten beschrijven.

Bij Lampens staat integratie voor *“het gezamenlijk in een computer opslaan en ontsluiten van verschillende documenttypes, waarbij één zoekvraag*

resulteert in beschrijvingen van verschillende documenttypes." (Lampens, 1992). Greenstein en Miller hanteren de term 'resource discovery', "[which] entails searching on line catalogue databases of differently described resources" (Greenstein en Miller, 1999). Verder wordt het concept 'cross-domain searching' gehanteerd, waarmee de nadruk gelegd wordt op het feit dat gebruikers databanken willen doorzoeken die tot uiteenlopende vakgebieden behoren. Andere varianten zijn 'cross-database searching' en 'federated searching': in beide gevallen gaat het om het simultaan doorzoeken van meerdere databanken.

Een vaak gehoorde term in dit verband is 'interoperability', wat betrekking heeft op de uitwisselbaarheid van data met uiteenlopende formaten of de technieken die dit mogelijk maken<sup>2</sup>.

Wat wij onder het begrip 'geïntegreerde databanken' verstaan, mag na het voorgaande al duidelijk zijn: *een systeem waarbij gegevens die verschillende bronnen beschrijven en/of in verschillende databanken zijn opgeslagen, door de gebruiker middels één zoekopdracht geconsulteerd kunnen worden.*

## WAAROM INTEGREREN?

De eerste vraag die opduikt bij het onderzoek naar de mogelijkheden van integratie van databanken, is die naar het nut van zo'n onderneming. We geven een opsomming van de belangrijkste argumenten die we in de literatuur rond dit onderwerp hebben ontmoet (Morel-Deckers, 1989; Phelan, 1999; Greenstein en Miller, 1999; Lampens, 1992):

- *Ondersteunen van wetenschappelijk onderzoek.* Door gegevens over

objecten te linken met gegevens over documenten die hiermee in verband staan, kan zowel het beheer van, als het onderzoek over deze objecten vereenvoudigd worden.

- *Gebruiksvriendelijkheid.* Gebruikers weten vaak niet in welke databank ze hun informatie moeten halen en/of op welke manier de verschillende databanken doorzocht kunnen worden. Een verkennende zoekopdracht doorheen meerdere gegevensbestanden kan hier een uitkomst bieden.
- *Serendipiteit.* Geïntegreerde databanken brengen bronnen samen die voordien gescheiden waren. Niet zelden ontdekt men bij zoeken doorheen verschillende datatypes onvermoede verbanden.
- *Kostenbesparing.* Wanneer meerdere datatypes in één databank, en dus ook in één softwarepakket opgenomen kunnen worden, kan een instelling een besparing realiseren op de aankoop en het onderhoud van een databank.
- *Uitwisselbaarheid van gegevens (interoperabiliteit).* Een belangrijke randvoorwaarde voor integratie is het gebruik van standaarden voor formele en inhoudelijke ontsluiting. Een neveneffect hiervan is dat instellingen hun collectiegegevens (hetzij van documenten of objecten) vlotter kunnen uitwisselen.

## STANDAARDEN

Zoals al eerder gesteld, spelen standaarden een sleutelrol. Naargelang van de functie kunnen we drie groepen informatie-standaarden onderscheiden:

- *Data value standards* hebben te maken met het vocabularium dat in de onderscheiden velden gebruikt kan worden, met andere woorden: standaarden voor inhoudelijke ontsluiting<sup>3</sup>.
- *Data structure standards* hebben betrekking op de formele structuur van records, zoals de definiëring van verschillende velden en de relaties die ertussen bestaan.

- *Information interchange standards* omschrijven het technische kader voor de uitwisseling van informatie tussen verschillende systemen. Bekende voorbeelden zijn SGML en Z39.50. Deze 'uitwisselingsstandaarden' maken het mogelijk om systemen te ontwikkelen waarin integratie van databanken daadwerkelijk gerealiseerd wordt. Men spreekt daarom ook van *system standards*.

## Data value standards

Voor de inhoudelijke ontsluiting van bronnen werden in de loop der jaren diverse gestandaardiseerde systemen ontwikkeld. Doorgaans onderscheidt men drie groepen: classificaties, thesauri en trefwoorden. Een aantal daarvan werd ontwikkeld voor het beschrijven van kunsthistorische verzamelingen, zoals de Art & Architecture Thesaurus (AAT), Uniform List of Artist Names (ULAN), Thesaurus of Geographic Names (TGN), Thesaurus of Graphic Materials (TGM-I en TGM-II) of Iconclass. Andere indexeertalen zijn universeel opgezet. Meest bekend zijn de Library of Congress Subject Headings (LCSH), de Library of Congress Classification (LCC), UDC of SISO.

Diverse onderzoeken toonden aan dat kunsthistorische instellingen er erg uiteenlopende gewoonten op nahouden wanneer ze indexeertalen gebruiken om hun collecties te beschrijven (Graham, 1999, Bierbaum, 1996). Zij maken zowel gebruik van universele indexeertalen als van vakgebonden thesauri, of een combinatie van beide. Andere instellingen ontwikkelen bovendien een eigen ontsluitingsstelsel, veelal in de vorm van een trefwoordenlijst. Ook de praktijk van het ad hoc toekennen van trefwoorden (ongecontroleerde trefwoordenlijsten) is nog steeds gangbaar.

Verscheidene auteurs wijzen echter op het belang en de bruikbaarheid van specifiek voor het kunsthistorisch vakgebied ontwikkelde indexeertalen, zowel voor de ontsluiting van de muse-



### INTEGRATIE?

ale collectie als voor het gebruik in de museumbibliotheek (Greenberg, 1993; Molholt en Petersen, 1993; Rogiest, 2000). In het bijzonder de AAT lijkt hierbij een grote kanshebber. De Nederlandse vertaling van deze Amerikaanse thesaurus kan een belangrijke impuls betekenen voor de toepassing van de AAT in ons taalgebied<sup>4</sup>.

Het is evident dat een grotere verspreiding van een algemeen aanvaarde indexertaal de mogelijkheden voor cross-domain searching en interoperabiliteit tussen museale collecties positief kan beïnvloeden.

### Data structure standards

Historisch gezien hebben vooral bibliotheken veel ervaring in de formele beschrijving van documenten. Het gebruik van standaarden hiervoor is dan ook al lang ingeburgerd: ISBD en MARC zijn bij uitstek voorbeelden van algemeen aanvaarde data structure standards. Objectregistratie in musea kan niet steunen op dergelijke universele standaarden (o.a. Ouwerkerk, 1989; Koot, 2000; Giral, 1988). Als belangrijkste reden hiervoor wordt vaak aangehaald dat de aard van de objecten in een typische museumcollectie veel diverser is en de objectregistratie daardoor veel complexer wordt. Bovendien kent een museum behoefte

aan een groter aantal beheersgegevens, zoals de oorsprong (provenance), bruikleeninformatie, restauratie- en holdinggegevens, detailafbeeldingen...

Bovendien zijn objecten in een museum bijna per definitie uniek. Gestandaardiseerde beschrijving met het oog op het opsporen van meerdere exemplaren van eenzelfde object is dan ook geen punt.

Toch groeide gedurende de jongste decennia de aandacht voor gestandaardiseerde beschrijvingsformaten voor museale objecten, doorgaans in het kader van de verhoogde aandacht voor het volgen van procedures bij het beheer van de collectie. Enkele voorbeelden zijn de *Historische voorwerpskaart* en de daarvan afgeleide *Basisregistratie*, *Categories for the Description of Works of Art* (CDWA), *Core Categories for Visual Resources* (VRA), *Van Eyck...*<sup>5</sup>.

### Dublin Core

De voorvermelde data structure standards werden ontwikkeld om een bron van een bepaald type te beschrijven. Een in dit opzicht erg verschillende standaard is Dublin Core (DC)<sup>6</sup>. Hoewel hij oorspronkelijk opgesteld was als een standaard voor het

beschrijven van elektronische (web)documenten, blijkt DC bij uitstek geschikt om zowel tekstuele bronnen als objecten te beschrijven<sup>7</sup>.

Kenmerkend voor DC is de ruime interpretatie die aan de betekenis van een veld (in DC-terminen: element) gegeven kan worden. Zo kan in het element `dc.creator` zowel de auteur van een tekst als de maker van een kunstwerk opgenomen worden. Om nauwkeuriger beschrijving mogelijk te maken, laat de standaard toe om de basiselementen van het format te verfijnen door zogenaamde qualifiers. Een qualifier geeft een nadere aanduiding van de inhoud van het gebruikte element, bijvoorbeeld de rol van de creator (in het voorbeeld `dc.creator.author` of `dc.creator.painter`). Daarmee kan de zoekfunctionaliteit verbeterd worden. Indien nodig kunnen de qualifiers echter ook genegeerd worden: wanneer een zoekmachine de gebruikte qualifier niet herkent of wanneer men een generieke zoekvraag wil stellen, wordt teruggegrepen naar 'dc-simple'. Door dit 'dumb-down-principle' kan interoperabiliteit tussen databanken verbeterd worden.

### Metadata mapping en crosswalks

Wanneer databanken die volgens verschillende recordstructuren zijn opgebouwd, met elkaar geïntegreerd worden, moet worden nagegaan op welke manier de velden van verschillende formats inhoudelijk (semantisch) met elkaar overeenstemmen. Dit proces wordt vaak omschreven als (metadata) mapping of semantic mapping. Met metadata mapping bedoelt men het correleren van de onderscheiden velden van verschillende databanken die gelijke of gelijksoortige informatie bevatten. Door mapping kan een

zoekmachine velden met vergelijkbare inhoud doorheen verschillende databanken simultaan doorzoeken (semantic interoperability). In een crosswalk worden deze correlaties schematisch voorgesteld.

Zelfs een ideale crosswalk kan niet alle problemen van een simultane search in meerdere databanken oplossen. We verklaren dit aan de hand van een voorbeeld: wanneer in verschillende datastructuren de gewone beschrijvingsregels worden gevolgd (bv. ISBD, resp. Basisregistratie), zal de naam van een kunstenaar in het eerste geval in het onderwerpsveld ondergebracht worden, maar in de objectbeschrijving in het veld voor de vervaardiger, auteur, creator of iets dergelijks. Zoeken op kunstenaarsnaam zou in dat geval enkel de bibliografische referentie weergeven, omdat enkel op de geïndexeerde onderwerpsontsluiting wordt gezocht. Lampens stelt voor dit probleem op te lossen door redundante gegevensopslag: bij de objectbeschrijving wordt de naam van de kunstenaar zowel in het auteursveld als in het trefwoordenveld ingevoerd (Lampens, 1992).

### Geïntegreerde databanken: problemen en oplossingen

Tot nu toe hebben we een aantal randvoorwaarden behandeld voor het integreren van databanken. Vervolgens rijst de vraag hoe cross-domain searching in de praktijk gerealiseerd kan worden.

#### *De minimalistische aanpak*

Objectdatabanken in musea bevatten vaak een veld waarin bibliografische gegevens kunnen worden opgeslagen. Deze optie kan nuttig zijn wanneer

het gaat om stukken die uitsluitend betrekking hebben op het beschreven object, zoals rapporten over de restauratie van een welbepaald kunstwerk. Meestal zijn deze velden echter niet (of moeilijk) opzoekbaar, doordat de gegevens niet volgens een bibliografische standaard opgeslagen werden. Nog moeilijker wordt het wanneer men een document aan meerdere objecten wil koppelen: men is dan verplicht de gegevens bij elke voorwerpsbeschrijving in te voeren.

Wanneer bibliografische referenties in een afzonderlijk bestand werden opgeslagen, worden die vaak in een eenvoudig menu aan de gebruiker gepresenteerd. Dit houdt in dat de gebruiker steeds opnieuw moet wennen aan de interface van de databank: de vormgeving, de beschikbare operatoren, de benaming van de velden..., en dus een langer leerproces moet ondergaan. Bovendien moet hij zijn zoekopdracht telkens herhalen wanneer hij meerdere databanken wil ondervragen. Van een echte integratie is er in dit model dus geen sprake<sup>8</sup>.

Door databanken in een menu aan te bieden blijven de zoekfunctionaliteiten van elke databank optimaal behouden. Een hieraan bijna volledig tegengestelde oplossing bestaat erin de gebruiker de mogelijkheid te bieden om in de gekozen databanken vrij te zoeken naar een gegeven term. Er wordt dan geen rekening gehouden met de recordstructuren van de ondervraagde databanken: de plaats (veld) waarin de zoekterm voorkomt wordt door de zoeksoftware als het ware genegeerd.

Een dergelijke oplossing biedt het voordeel dat de gebruiker geen rekening moet houden met de specificiteit van de databank die hij ondervraagt. De leercurve wordt daarmee op het laagst mogelijke niveau gehouden. Ook de zoekinterface kan zeer eenvoudig gehouden worden: in haar meest rudimentaire vorm volstaat één invulveld om de zoektermen in te geven en een knop op het scherm om de zoek-

opdracht te versturen. Er bestaan een aantal commerciële softwarepakketten die vooral in de bedrijfswereld door kennismangers gepromoot worden<sup>9</sup>. Op deze wijze geïntegreerde databanken kunnen in hun oorspronkelijke vorm blijven bestaan en gebruikt worden. Conversie naar andere standaarden of het uitwerken van complexe crosswalks is overbodig. Hierdoor blijft ook het onderhoud van het zoekstelsel minimaal.

Free text oplossingen lijden echter aan het euvel dat elke gebruiker van Google c.s. kent: de mogelijkheid om verfijnde zoekacties te verrichten wordt opgeofferd in het voordeel van de gebruiksvriendelijkheid.

### Integratie op basis van information exchange standards

In de aangehaalde voorbeelden ontbreekt het dus telkens ofwel aan functionaliteit, ofwel aan gebruiksvriendelijkheid, of zelfs aan beide. Vanuit deze vaststelling werden verschillende oplossingen ontwikkeld, die ondergebracht kunnen worden in de volgende typologie:

- gegevensconversie;
- linken;
- virtuele integratie.

Elk type steunt op verschillende technologieën, die we hier echter niet uitgebreid zullen behandelen. We gaan wel in op de information interchange standards, die met name bij virtuele integratie van belang zijn.

#### *Conversiemodel*

In het conversiemodel worden gegevens uit één of meerdere databanken naar een ander formaat omgezet, vaak ook naar een andere software geconverteerd. De data worden dus op basis van een crosswalk in een nieuw gegevensbestand ondergebracht, waarbij elementen uit de 'oude' datastructuren naar een nieuwe worden gemapped. Men kan daarbij gebruikmaken van

een domein-specifiek formaat, zoals ISBD of MARC. In dit geval moet getracht worden gegevens die een object beschrijven, in een semantisch verwant veld van de bibliografische recordstructuur onder te brengen<sup>10</sup>. Een niet onbelangrijk voordeel is de eenvoud van deze oplossing: de interface moet slechts één databank ondersteunen en de gebruiker kan alle zoekopdrachten via hetzelfde zoekscherm uitvoeren.

Verscheidene auteurs hebben de haalbaarheid van een dergelijk model in een museumomgeving onderzocht. Hun conclusies lopen uiteen, maar in principe is men het erover eens dat er bij deze oplossing nogal wat kunst- en vliegwerk te pas komt om een object in een bibliografisch formaat accuraat te beschrijven (Lampens, 1992; Ouwkerk, 1989; Baron, 1995; Kempe, 1999; Blackaby en Sandore, 1997; Woodley, 2000). Een alternatieve aanpak is het gebruik van een generiek formaat, zoals DC, dat per definitie universeel opgezet is<sup>11</sup>.

#### *Linking-model*

Producenten van databanken – commerciële uitgevers, bibliotheken en intermediaire organisaties – trachten steeds vaker hun producten een toegevoegde waarde te geven door verwante informatie-eenheden aan elkaar te koppelen of te linken. Typische links in een bibliotheekcontext zijn koppelingen tussen secundaire databanken en catalogi: de eindgebruiker zoekt in een databank naar bibliografische referenties met betrekking tot zijn onderzoeksdomein, waarna een link met de lokale catalogus hem toont waar het tijdschrift met het gewenste artikel kan worden geraadpleegd<sup>12</sup>.

Dit linken op recordniveau onderscheidt zich van het geïntegreerd of simultaan zoeken in verschillende databanken doordat de integratie van de gegevens niet gebeurt op het niveau van de zoekopdracht, maar vertrekt vanuit een gevonden resultaat. Aan

de hand van een antwoord worden dus nieuwe antwoorden gevonden. De basisidee waarop deze koppeling van informatie steunt, is in wezen dezelfde als die van de hyperlinks op het world wide web<sup>13</sup>.

In de meeste gevallen worden links gecreëerd door gebruik te maken van gemeenschappelijke data values, zoals de authority files met onderwerpstermen of persoonsnamen.

Een bekend voorbeeld is Agrippa, de databank van het AMVC<sup>14</sup>. Centraal in Agrippa staat de klapper, een encyclopedisch apparaat van 55.000 namen van personen, instellingen... Hoewel de objectbeschrijvingen in Agrippa ook op andere manieren opzoekbaar zijn, vormt de klapper de centrale as van het systeem: de integratie van meerdere objecttypes wordt gerealiseerd door gebruik te maken van gemeenschappelijke data values.

Het Maritiem Museum<sup>15</sup> biedt een zoekstelsel waarmee literatuur- en objectcollecties van vijf Nederlandse scheepvaartmusea simultaan doorzocht kunnen worden. De databanken van de deelnemende instellingen zijn geautomatiseerd met een toepassing uit de AdLib familie, AdLib Bibliotheek en AdLib Museum. Een webapplicatie, die de bovenbouw van het systeem vormt, zorgt voor de 'cross-linking' van records in de verschillende bestanden door het gezamenlijk gebruik van authority files en thesauri. Daardoor kan de gebruiker doorheen verschillende bestanden zoeken, ongeacht het fysieke formaat of de gebruikte data structure standard.

#### *Virtuele integratie*

Toepassingen als Agrippa en het Maritiem Museum gaan al veel verder dan het louter linken van bronnen: vanuit het perspectief van de gebruiker kan met één zoekscherm naar verschillende datatypes gezocht worden. Een belangrijke beperking is echter dat deze systemen binnen de grenzen van

een bepaalde software werken. De controle op de integratie ligt dus volledig bij de beheerders van de databanken.

Virtueel geïntegreerde databanken functioneren daarentegen als een open systeem. De rol van de beheerders is beperkt tot het openstellen van hun systeem, meestal door dit compatibel te maken met een bepaalde information interchange standard (protocol).

Toepassingen die deze virtuele integratie beheren noemt men middleware. Een middleware is verantwoordelijk voor het transport van zoekopdrachten en resultaten tussen verschillende databanken op verschillende platformen<sup>16</sup>.

Middleware steunt op protocols zoals Z39.50, of het Open Archives Initiative Protocol<sup>17</sup>.

#### *Z39.50*

Z39.50 is opgebouwd volgens het client/server model, waarbij een computer (de client, of in Z39.50 terminologie: de origin) een verzoek naar een andere computer stuurt (de server, ofwel target) die de vraag verwerkt. De target zorgt voor de communicatie met de ondervraagde databank door de zoekopdracht te vertalen naar een voor die databank begrijpbare taal.

Z39.50 biedt het voordeel dat de communicatie tussen client en server platform-onafhankelijk is. Een andere troef is dat de origin simultaan verschillende targets kan ondervragen en de resultaten samenvoegen, zodat het voor de gebruiker lijkt alsof hij met slechts één (virtuele) databank werkt.

De manier waarop een vraag door de target wordt behandeld, wordt in sterke mate bepaald door de attribute sets. Een attribute set kan omschreven worden als een verzameling van waarden, die bepalen welke velden doorzocht moeten worden, hoe de zoektermen zich verhouden tot de waar-

den in de databank (gelijk aan, groter dan...), of en hoe er getrunceerd moet worden... Erg populair is het Bib-1 attribute set, dat voor bibliografische toepassingen werd ontwikkeld maar ook in andere gemeenschappen toegepast wordt.

De USE-attributes geven aan welke velden gebruikt kunnen worden in een zoekopdracht. De server zorgt voor de vertaling (mapping) van de attributen naar de semantisch verwante velden van de onderliggende databank (de access points). Use-attributes kunnen overeenstemmen met één bepaald veld uit de databank (bv. author-name personal bij een MARC-record), maar ook combinaties van velden zijn mogelijk, zoals de combinatie author-title-subject (keyword-searching). Het meest brede use-attribute is 'any', waarbij gezocht wordt in elk opzoekbaar veld van de databank.

Om communicatie mogelijk te maken tussen origins en targets die tot verschillende vakgebieden behoren, werd eind jaren 1990 het Bath Profile ontwikkeld. Doel was om tegemoet te komen aan de behoefte aan een relatief eenvoudig profiel waarin de waarden van alle attribute sets duidelijk omschreven zijn. Elk 'Bath-compliant' systeem kan een minimum aantal zoekopdrachten op eenduidige wijze interpreteren en behandelen. Voor eenvoudige cross-domain searches wordt het Bib-1 attribute set aanbevolen (Bath profile, 2000).

Belangrijk onderzoek in verband met de toepassing van Z39.50 werd gedaan door het CIMI, dat in 1996 het project Cultural Heritage Information on line (CHIO) startte<sup>18</sup>. Doel was te onderzoeken hoe Z39.50 gebruikt kan worden in een toepassing voor het gedistribueerd zoeken in culturele erfgoeddatabanken<sup>19</sup>.

Een praktische toepassing van de bevindingen van het CIMI was het Interoperability Testbed (1997-1998), een proefproject waaraan dertien verschillende organisaties deelnamen (o.a. musea en universiteiten, maar ook commerciële softwareontwikkelaars). Aan het einde van het project hadden de deelnemende organisaties acht Z39.50 servers en drie clients operationeel gemaakt. Een client die de CIMI Z39.50 standaard ondersteunt, biedt de eindgebruikers een uniforme interface waarmee erfgoeddatabanken ondervraagd kunnen worden. Deze bronnen kunnen zowel bibliografische beschrijvingen, afbeeldingen met bijbehorende tekst of objectbeschrijvingen met afbeelding bevatten. De clients zijn in staat zoekresultaten uit verschillende bronbestanden samen te voegen en te sorteren. Tests wezen uit dat er voldoende overeenstemming bereikt was over de implementatie van de standaard bij de verschillende deelnemers om efficiënt gedistribueerd zoeken mogelijk te maken (Moen, 1998).

#### *Drill-down principle*

Een tweede voorbeeld van het gebruik van Z39.50 voor geïntegreerde toegang tot erfgoeddatabanken is de Arts and Humanities Data Service (AHDS), een gateway service die toegang biedt tot een breed scala van culturele databanken. Elke deelnemende instelling heeft eigen, expliciete eisen met betrekking tot de elektronische opslag en ontsluiting, die moeilijk in te passen zijn in een generieke datastructuur. Anderzijds bevat elk beschrijvingsformaat een aantal velden die min of meer identiek zijn. De AHDS koos daarom voor een drill-down model: bij een brede query ontvangt de gebruiker rudimentaire informatie over de beschikbare bronnen, met als enig doel het bestaan ervan op te sporen. In dit stadium is een generiek formaat als DC-simple zeer bruikbaar. In een volgende fase zal de gebruiker rijkere informatie willen om de gevonden gegevens te evalueren, en zijn de meer

domein-specifieke beschrijvingsformaten noodzakelijk.

Een belangrijk minpunt van dit 'drill-down' model is dat de softwarematige uitwerking er niet eenvoudiger op wordt. Deze opmerking geldt overigens voor Z39.50 in haar geheel: het uitwerken van een goed functionerende applicatie vereist een aanzienlijke investering in tijd, middelen en kennis.

#### *Open Archives Initiative*

Een conceptueel met Z39.50 verwante, maar veel minder complexe oplossing werd ontwikkeld binnen het Open Archives Initiative (OAI)<sup>20</sup>. Het OAI werkte een set van information system standards uit, dat de toegang tot wetenschappelijke informatie op het web moet vereenvoudigen. Het OA-framework bestaat uit twee 'klassen' van deelnemers: data providers beheeren een web server die het OA-protocol ondersteunt, waardoor metadata over de inhoud van de eigen databanken opgevraagd kunnen worden door een metadata harvester. Dit 'oogsten' van metadata wordt beheerd door een organisatie uit de tweede klasse: de service providers. De aldus verworven metadata kunnen door de service provider verder bewerkt en aan een specifieke doelgroep aangeboden worden. Hoewel dit OA-framework initieel werd ontwikkeld om e-prints te ontsluiten, blijkt het ook een geschikt instrument voor resource discovery in andere domeinen. Zo onderzoekt het CIMI momenteel het nut van OA voor musea en educatieve of commerciële instellingen in de culturele sector. Eén van de aandachtspunten daarbij is de bruikbaarheid van Dublin Core, die integraal deel uitmaakt van het OA protocol (Perkins, 2001).

#### **BESLUIT**

De nadruk op standaardisering kan gelden als de bottom line van dit artikel: projecten om cultureel erfgoed in

digitale vorm beschikbaar te stellen, kunnen enkel slagen wanneer ze rekening houden met de ruime context van ervaringen, procedures en (nogmaals) standaarden die wereldwijd worden ontwikkeld.

We willen afsluiten met een citaat uit de afscheidsrede van Eleanor Fink, voormalig hoofd van het Getty Information Institute: *"Met de opkomst van het internet als een universeel communicatiemedium, werd informatie over cultureel erfgoed in de vorm van beeld, geluid en tekst aan een steeds sneller tempo gedigitaliseerd en op het Web beschikbaar gesteld. Toch geeft deze universele toegang in zijn huidige vorm de indruk van een lappendeken die geen enkele mogelijkheid biedt om op een efficiënte manier een gegeven te lokaliseren. (...) Standaarden zijn essentieel als bindmiddel om digitale bronnen over kunst te ontwikkelen en met elkaar te verbinden, en om wereldwijde toegang ertoe op eenvoudige wijze mogelijk te maken"* (Fink, 1999).

## NOTEN

1. Dit artikel is een bewerking van de scriptie *Integratie van erfgoeddatabanken: aspecten en modellen van resource discovery in museale en bibliografische gegevensbestanden*, ingediend tot het behalen van het Graduaat Bibliotheekwezen en Documentaire Informatiekunde (Antwerpen, 2001). Met dank aan Saskia Scheljtens voor de waardevolle suggesties.
2. Borgman geeft als definitie: *"The ability of information systems to interact in a useful way on a real-time basis over communication works."* (Borgman, 1997).
3. Data content standards houden verband met de regels en conventies die bepalen hoe data ingevoerd worden (bv. regels voor catalografie).
4. Het AAT-vertaalproject loopt nog tot september 2002 ([www.rkd-db.nl/aat/](http://www.rkd-db.nl/aat/).)
5. Voor een overzicht: Bower en Roberts, 1996.
6. url: [www.dublincore.org](http://www.dublincore.org).
7. Over het gebruik van DC in een museumcontext: zie CIMI, 1999 en Woodley, 2000.
8. Zie bijvoorbeeld het Hogarth-project: [www.courtauld.ac.uk/hogarth/](http://www.courtauld.ac.uk/hogarth/).
9. Bekende producten zijn onder andere Verity, Excalibur of OpenText, die overi-

gens ook meer complexe zoekacties dan louter free-text searches aankunnen.

10. Een variëteit is van bij de aanvang data in één domeinspecifieke datastructuur in te voeren, waarbij geen conversie in strikte zin te pas komt. Voor- en nadelen blijven echter dezelfde.
11. Zie onder andere het project [www.deventergeschiedenis.nl](http://www.deventergeschiedenis.nl) (Zandhuis en Dijksterhuis, 2002).
12. Een van de geavanceerdere technieken is SFX van Ex Libris (Van de Sompel, 1999).
13. Een potentieel interessante toepassing hiervan vonden we in het Koninklijk Museum voor Schone Kunsten te Antwerpen. In het bibliotheekstelsel wordt een veld gebruikt om het inventarisnummer op te nemen van kunstwerken uit de eigen museumcollectie, waarvan een afbeelding in de beschreven publicatie is opgenomen. De eindgebruiker kan een inventarisnummer ingeven om alle publicaties te vinden die een afbeelding (en vaak ook uitgebreidere informatie) van het kunstwerk bevatten. Deze 'integratie' wordt door het systeem echter niet ondersteund: de gebruiker moet zelf eerst het inventarisnummer opzoeken, waarmee hij in de bibliotheekcatalogus de relevante publicaties kan terugvinden.
14. Archief en Museum van het Vlaamse Cultuurleven ([www.dma.be/cultuur/amvc/agrippa.html](http://www.dma.be/cultuur/amvc/agrippa.html)).
15. [www.maritiemmuseum.nl](http://www.maritiemmuseum.nl).
16. Een meer tot de verbeelding sprekende definitie vonden we bij Klingenstein: *"Middleware is the intersection of the stuff that network engineers don't want to do with the stuff that application developers don't want to do."* (Klingenstein, 1999).
17. Een derde voorbeeld is het Simple Digital Library Interoperability Protocol (SDLIP) (Paepcke, 2000).
18. Consortium for the Computer Interchange of Museum Information ([www.cimi.org](http://www.cimi.org)).
19. The CIMI Profile: Z39.50 application for Cultural Heritage Information. Dit profiel geeft richtlijnen voor het gebruik van de Z39.50-1995 standaard voor de ontsluiting van cultureel erfgoed. De informatiebronnen kunnen variëren van digitale afbeeldingen, fysische artefacten en hun beschrijving tot bibliografische data en full-text documenten.
20. [www.openarchives.org](http://www.openarchives.org).

## BIBLIOGRAFIE

- Baron, Robert A., *Using MARC for collection management: the arguments against*, in: *Spectra*, 22 (1995) 4.
- Bath Profile: internationale Z39.50 profiel voor bibliotheekapplicaties [HTML]. - Release 1.0. - 2000. [http://www.ukoln.ac.uk/interop-focus/activities/z3950/int\\_profile/](http://www.ukoln.ac.uk/interop-focus/activities/z3950/int_profile/)

[bath/draft/stable1.html](http://bath/draft/stable1.html)

- Bierbaum, Esther Green, *Museum libraries: the more things change...*, in: *Special Libraries*, 87 (1996) 2: p. 74-88.
- Blackaby, Jim en Sandore, Beth, *Building integrated museum information systems: practical approaches to data organization and access* [HTML]. - 1997. <http://images.grainger.uiuc.edu/papers/mweb215.htm>.
- Borgman, Christine, *From acting locally to thinking globally: a brief history of library automation*, in: *Library quarterly*, 67 (1997): p. 215-249.
- Bower, Jim en Roberts, Andrew, *Museum and cultural heritage information standards and organisations* [HTML]. - CIDOC, 2001. <http://www.cidoc.icom.org/stand2.htm>.
- CIMI Dublin Core Metadata Working Group, *Guide to best practice: Dublin Core (DC1.0=RFC2413)* [PDF]. - CIMI, 1999. - 104 p. [http://www.cimi.org/documents/meta\\_bestprac\\_final\\_0899.pdf](http://www.cimi.org/documents/meta_bestprac_final_0899.pdf).
- Fink, Eleanor E., *The Getty Information Institute: a retrospective* [HTML], in: *D-Lib magazine*, 5 (1999) 3. <http://www.dlib.org/dlib/march99/fink/03fink.html>.
- Giral, Angela, *At the confluence of three traditions: architectural drawings at the Avery Library*, in: *Library trends*, 37 (1988) 2: p. 232-242.
- Graham, Margaret E., *The description and indexing of images: report of a survey of ARLIS members. 1998/99*. - Newcastle upon Tyne: Institute for Image Data Research, 1999.
- Greenberg, Jane, *Intellectual control of visual archives: a comparison between the Art and Architecture Thesaurus and the Library of Congress Thesaurus for Graphic Materials*, in: *Cataloging and classification quarterly*, 16 (1993) 1: p. 85-101.
- Greenstein, Daniel en Miller, Paul, *Discovering online resources across the humanities: an application of the Dublin Core and the Z39.50 network application protocol*, in: *Information landscapes for a learning society: networking and the future of libraries*. - London: UK Office for library and information networking, 1999.
- Kempe, Deborah, *SCIPIO art and rare books catalogue file: perspectives from a valued user and contributor*, in: *RLIN Focus*, (1999) 40.
- Klingenstein, Kenneth J., *Middleware: the second layer of IT infrastructure*, in: *Cause/effect journal*, 24 (1999) 4.
- Koot, Geert-Jan, *De museumbibliothecaris als informatiestrateg*, in: *Bibliotheek- & archiefgids*, 76 (2000) 9: p. 12-19.
- Lampens, Dieter, *Geïntegreerde informatie-opslag en -ontsluiting: een case-study*, in: *Bibliotheek- & archiefgids*, 68 (1992) 2: p. 15-27.
- LeVan, Ralph, *Dublin core and Z39.50*

- [HTML]. - (1998). <http://www.oclc.org/oclc/research/projects/core/documents/notes-levan-19980202.htm>.
- Miller, Paul, *Developing the Bath Profile* [HTML] in: *Ariadne*, (1999) 21. <http://www.ariadne.ac.uk/issue22/at-the-event/bath-profile.html>.
  - Moen, William E., *Distributed access to cultural heritage information: the CIMI Z39.50 Profile* in: *Library technology*, 3 (1998) 4.
  - Molholt, Pat en Petersen, Toni, *The role of the Art and Architecture Thesaurus in communicating about visual art*, in: *Knowledge organization*, 20 (1993) 1; 30-34.
  - Morel-Deckers, Yolande, *Geïntegreerde ontsluiting in musea*, in: *Museumleven*, (1989); p. 49-50.
  - Ouwerkerk, Annemiek, *Standaard en structuur: de ontsluiting van museumdocumenten*, in: *Open*, 21 (1989) 3; p. 89-95.
  - Paepcke, Andreas, *Search middleware and the Simple Digital Library Interoperability Protocol* [HTML], in: *D-Lib magazine*, 6 (2000) 3. <http://www.dlib.org/dlib/march00/paepcke/03paepcke.html>.
  - Perkins, John, *A new way of making cultural information resources visible on the web: museums and the Open Archives Initiative* [HTML], 2001. - Paper presented at Museums and the Web 2001. <http://www.archimuse.com/mw2001/papers/perkins/perkins.html>.
  - Phelan, Colleen en Beaulieu, Micheline, *A model to support literary research collections on the world wide web* [HTML] - Paper presented at Museums and the Web, 1999. <http://www.archimuse.com/mw99/papers/phelan/phelan.html>.
  - Rogiest, Peter, *De kunstbibliotheek van het KMSKA en haar inhoudelijke ontsluiting*, in: *Bibliotheek- & archiefgids*, 76 (2000) 4; p. 8-17.
  - Van de Sompel, Herbert en Hochstenbach, Patrick, *Reference linking in a hybrid library environment. Part 1: Frameworks for linking* [HTML] in: *D-Lib magazine*, 5 (1999) 4. [http://www.dlib.org/dlib/april99/van\\_de\\_sompel/04van\\_de\\_sompel-pt1.html](http://www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt1.html).
  - Woodley, Mary, *Crosswalks: the path to universal access?* in: *Introduction to meta-data: pathways to digital information* / ed. Tony Gill, Anne Gilliland-Swetland and Murtha Baca. - 2de ed., (2000). - S.l.: Getty Information Institute. - ISBN 0-89236-533-1.
  - Zandhuis, Ivo en Dijksterhuis, Ton, *Dublin Core bij het Deventer Iselfront-project* in: *Informatie professional*, 6 (2002) 3; p. 28-31.

## SAMENVATTING

Integratie van databanken in de sector van het cultureel erfgoed maakt het mogelijk dat onderzoekers een aantal bestanden simultaan ondervragen en bronnen opsporen die tot verschillende objecttypes behoren, zoals kunstwerken, artikels of monografieën. Dit artikel onderzoekt het bestaan en het gebruik van standaarden waarmee bronnen in bibliotheekcatalogi en/of museale databanken beschreven kunnen worden. Data value standards werden ontwikkeld waarmee zowel bibliografische als objectbronnen inhoudelijk kunnen worden ontsloten. Bovendien zijn er data structure standards beschikbaar die voor beide objecttypes geschikt zijn (o.a. MARC en Dublin Core). Vervolgens wordt een typologie gegeven van systemen die deze integratie verwezenlijken, door gebruik te maken van information exchange standards. De modellen worden geïllustreerd aan de hand van voorbeelden uit de praktijk.

## ABSTRACT

This article treats the issue of cross-database searching in the field of culturale heritage, which allows users to search a set of databases simultaneously on a given subject, and retrieve descriptions of different datatypes, such as paintings, monographs, articles.

The availability and use of standards for the description of sources in libraries and/or art collections are examined. It is stated that universal data value standards have been developed that can be applied for the description of bibliographic, as well as object-like documents or sources. Also, methods are available for the integration of different record structures (i.e. data structure standards such as MARC or Dublin Core).

Next, a typology is given of different systems that can realize a cross-domain integration, based on information exchange standards. These models are being illustrated with examples of current practice and research.