



UvA-DARE (Digital Academic Repository)

Virtuele collectie 'openbare gemakelikheden'

Brandsma, R.; Bouissaghouane, M.; Brussee, R.; Smit, S.; Wartena, C.

Publication date

2009

Document Version

Final published version

Published in

Cultuur in context: erfgoeddata in nieuwe samenhang

[Link to publication](#)

Citation for published version (APA):

Brandsma, R., Bouissaghouane, M., Brussee, R., Smit, S., & Wartena, C. (2009). Virtuele collectie 'openbare gemakelikheden'. In N. van Asseldonk, P. van Mensch, & H. van Vliet (Eds.), *Cultuur in context: erfgoeddata in nieuwe samenhang* (pp. 54-63). Reinwardt Academie.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.



3. virtuele collectie 'openbare vermakelijkheden'

**Renze Brandsma, Mohamed Bouissaghouane,
Rogier Brussee, Saskia Smit, Christian Wartena**

het koppelen van digitale collecties en bestanden

Gedigitaliseerde collecties bieden een uitstekende mogelijkheid om een brede kijk op cultureel erfgoed te operationaliseren door een rijke en geïntegreerde digitale infrastructuur op te bouwen, die ook openstaat voor aansluiting bij andere bronnen. Het tegelijkertijd kunnen zoeken en vinden van informatie over bijvoorbeeld een auteur of persoon, of over een bepaald thema in verschillende digitale collecties en bronnen is van belang voor onderwijsdoeleinden en onderzoek maar ook voor een willekeurige archiefbezoeker: Wat is er bijvoorbeeld bekend over dezelfde plaats of tijd of over hetzelfde onderwerp? Welke literatuur, archiefrecords, beelden, geluidsobjecten of documenten zijn specifiek aan het object gerelateerd?

Het programma *Cultuur in Context* heeft dit koppelen van digitale bestanden aangepakt met de aanwezige partners in het programma voor het domein van 'openbare vermakelijkheden' (theater en film) met een nadruk op de periode van circa 1800 tot 1950. Als uitwerking van deze ambitie is gewerkt aan een virtuele collectie door koppeling van subsets van data uit bestaande digitale bestanden bij de diverse programmadeelnemers uit de bibliotheek-, museum- en archiefwereld. Er is een testomgeving ingericht om ervaring op te doen om bestaande digitale bestanden technisch te koppelen en beschikbaar te stellen. Die ervaring is ook nadrukkelijk vastgelegd en gedeeld met deze community. Ter ondersteuning hiervan is een begrippenkader (ontologie) rondom het domein van 'openbare vermakelijkheden' ontworpen. Deze ontologie is bedoeld als inhoudelijk referentiekader voor de feitelijke koppeling van collecties over openbare vermakelijkheden. Het spreekt voor zich dat bij het maken van een dergelijke ontologie veel domeinkennis nodig is, wat een direct beroep doet op de expertise van professionals.

methoden voor het koppelen

Het koppelen van digitale collecties of zoeken in meerdere digitale collecties kan op verschillende manieren worden bewerkstelligd. Het begrip interoperabiliteit is daarbij een kernbegrip. Interoperabiliteit kent technische, semantische en organisatorische aspecten. De organisatorische aspecten zullen we hier niet behandelen.

technische interoperabiliteit

De voornaamste methoden om tot technische interoperabiliteit te komen zijn de volgende:

1. Linken. Het linken van objecten uit verschillende digitale collecties is een simpele wijze om collecties te verbinden. Een dergelijke link kan idealiter op basis van stabiele identifiers (of persistente identifiers) plaats vinden. Let wel, er is een verschil tussen een identifier en een link. Dat wil zeggen dat objecten en/of metadata records een stabiele identifier moeten hebben. De links kunnen dan het beste opgebouwd worden op basis van deze identifiers. Bijvoorbeeld door een combinatie van de URL van een collectie (www.cinemacontext.nl/) met een stabiele identifier (bijv. B000209 voor de bioscoop Luxor in Alphen). De link wordt dan: www.cinemacontext.nl/id/B000209.
2. Crawling. Zoekmachines zoals Yahoo en Google gebruiken webcrawlers om webpagina's op te halen, te verzamelen en doorzoekbaar te maken. De gegevens in databanken worden daarbij veelal niet meegenomen, omdat deze niet direct als webpagina's beschikbaar zijn, maar pas op het moment dat een gebruiker er naar vraagt als een webpagina worden gepresenteerd.
3. Metadata harvesten en centrale index. Metadata harvesten is een manier om metadata records te verzamelen uit databanken en deze doorzoekbaar te maken in een centrale index of databank. Het resultaat van een zoekactie in deze centrale index toont dan de metadata en veelal een link/URL naar de oorspronkelijke online databank met behulp van een stabiele identifier. Het Open Archive Index Protocol voor Metadata Harvesting (OAI-PMH) is een gangbaar protocol om batchgewijs de metadata automatisch uit online databanken te harvesten. De digitale collecties of systemen dienen dan wel technisch aan het OAI-PMH protocol voldoen. Als deze systemen niet aan een dergelijk protocol kunnen voldoen, dan dient op regelmatige basis een update van de metadata records op CD's (of via FTP) opgestuurd te worden.
4. Tegelijkertijd zoeken in digitale collecties (gedistribueerde zoeken). Het is ook mogelijk met bepaalde technieken tegelijkertijd te zoeken in diverse digitale collecties of databanken. In de bibliotheekwereld is het SRU protocol (of voorheen Z39.50 protocol) een gangbare techniek. Een zoekvraag kan via dit internetprotocol naar de diverse online collecties gestuurd worden en de records die als respons worden gegeven kunnen verzameld worden. Bij deze methode wordt dus geen centrale index gemaakt. Ook hiervoor geldt echter weer dat systemen moeten voldoen aan het protocol.

semantische interoperabiliteit

Semantische interoperabiliteit betreft de koppeling van de betekenis van de data. Er zijn mechanismen ontworpen om trefwoorden of andere inhoud en hun betekenis in bestanden of digitale collecties aan elkaar te verbinden, bijvoorbeeld door het geza-

menlijke gebruik van dezelfde thesauri of trefwoordensystemen; of door het maken van gedetailleerde mappings tussen thesauri. Een andere aanpak is het maken van ontologieën. In de informatica is een ontologie het product van een poging een strikt begrippenkader te formuleren over een bepaald domein. In een ontologie gaat het om de begrippen en de relaties daartussen, terwijl in een thesaurus in beginsel het meer om gebruikte termen gaat. Ontologieën bevorderen de uitwisseling tussen verschillende applicaties en systemen.

stand van zaken bij instellingen in project

De deelnemende instellingen aan *Cultuur in Context* hebben veelal databanken met gegevens (metadata) over objecten van hun collecties online toegankelijk gemaakt. Het betreft dan catalogi, beeldbanken, digitale archiefbeschrijvingen of specifiek onderzoeks-databanken, zoals de databank Cinema Context, www.cinemacontext.nl bij de Universiteit van Amsterdam. Voor een deel zijn de instellingen nog bezig om hun collecties te beschrijven, te digitaliseren en online te brengen. De online systemen die daarbij gebruikt zijn voldeden ten tijde van het project niet aan protocollen om de metadata geautomatiseerd te harvesten of tegelijkertijd te zoeken in meerdere online collecties. De data zijn daarom eerst verzameld en daarna geconverteerd naar een gemeenschappelijk datamodel op basis van de ontwikkelde ontologie. Na de conversie werden de data in een enkele, in dit geval een door het Digitaal Productiecentrum (DPC) van de Universiteit van Amsterdam beheerde, database gestopt. Daarmee kunnen allereerst de problemen die de unificatie van verschillende data met zich meebrengt onderzocht worden. Het verzamelen heeft via het opsturen van dvd's met metadata van objecten en data uit de collecties van de instellingen plaats gevonden. De verzamelde metadata en data zijn zodoende in een centrale index (of centrale databank) bij het DPC opgenomen. Waar mogelijk is in de verzamelde metadata ook een identifier van het record of object meegeleverd, waarmee een link gemaakt kan worden naar de digitale collectie of databank van de betreffende instelling.

ontologieën als bindmiddel

Data-integratie is pas nuttig als er op betekenisniveau overeenkomsten en verbanden zijn. Bij openbare gemaklijkheden is dat bijvoorbeeld zo omdat in films en voorstellingen door dezelfde acteur wordt gespeeld of in dezelfde gebouwen zowel film- als theatervoorstellingen werden geprogrammeerd. De verschillende databases zijn echter verder verschillend en heel anders gestructureerd. Om ze te kunnen integreren is het nuttig een begrippenapparaat te hebben dat door iedereen gedeeld wordt en dat concreet genoeg is om bestaande data, en de relaties ertussen te kunnen uitdrukken. Hiertoe is een ontologie ontwikkeld.

Een ontologie is een formalisatie van de werkelijkheid. Ontologieën geven dus niet zo zeer betekenis aan begrippen, maar leggen een begrippenkader vast en documenteren eigenschappen en betekenis van die begrippen. De eigenschappen van een begrip die we daadwerkelijk kunnen formaliseren zijn meestal erg eenvoudig. De belangrijkste is de zogenaamde 'is een'-relatie die aangeeft dat elk object van type A *a fortiori* een object van type B is. Een typisch voorbeeld is: elke koe is een dier. Welke eigenschappen

horen bij welke begrippen is een ander belangrijk deel van de formele structuur van een ontologie. We kunnen bijvoorbeeld zeggen dat een stad in een land ligt, of dat een persoon een vader en moeder heeft. Dit soort eigenschappen zijn precies de dingen die zich gemakkelijk laten uitdrukken in Web Ontology Language (OWL) en Resource Description Framework (RDF), twee op elkaar voortbordurende talen van het W3C-consortium, bedoeld om informatie op het web te koppelen met behulp van ontologieën.

Het domein 'openbare gemakkelikheden' is veelomvattend. Er zijn verschillende wetenschappelijke perspectieven op. Bovendien beschrijven wetenschappers en archivariissen dezelfde dingen op een verschillende manier. Geen van de in het project deelnemende historici had ervaring met het ontwerpen en schrijven van formele ontologieën. Voor een eerste opzet zijn we daarom begonnen met een inventarisatie van de belangrijkste begrippen uit het domein. Als bronnen dienden naast een brainstormsessie met deskundigen vooral databases die door een aantal wetenschappers wordt gebruikt, in het bijzonder de beschrijvingen van theaterproducties van het Theater Instituut Nederland, de database Cinema Context over bioscoopvoorstellingen van de afdeling Media Studies van de Universiteit van Amsterdam en de bestanden met data over zitplaatsen, abonnementen, directies en programmering van theatervoorstellingen in Rotterdam van de Universiteit Utrecht.

Ordering in het begrippenapparaat bleek de moeilijkste fase in het project. Het gedeelde uitgangspunt was om 'wie, wat, waar, wanneer' vragen te kunnen beantwoorden. Al snel bleek echter dat de grote verscheidenheid van elkaar gedeeltelijk overlappende begrippen ogenschijnlijk eenvoudige begrippen onduidelijk maakte. Functies van gebouwen liggen niet vast, theaters worden als bioscoop gebruikt, en de aard van de voorstelling of de auteur van een werk is onduidelijk. De meeste problemen zijn opgelost na uitvoerige discussies. Hierbij werden de wensen van de verschillende wetenschappers regelmatig geconfronteerd met de beperkingen van een formele taal als OWL, waarin geen ruimte is voor nuances, kanttekeningen en voorbehouden. Het formaliseren van de uitkomsten in de vorm van een ontologie werkte hierbij uitermate vruchtbaar omdat het dwingt tot helder geformuleerde definities. Hierdoor kon bovendien voortdurend de consistentie met eerder genomen beslissingen worden getoetst. Na het vastleggen van een hoofdstructuur was de belangrijkste opgave het uitwerken van details voor specifieke deelgebieden. Daarvoor moesten verschillende wetenschappers gelijktijdig aan de ontologie kunnen werken. Het grootste obstakel was dat omgevingen voor het bewerken van ontologieën zoals Ontostudio of Protégé te complex zijn om gebruikt te worden zonder specifieke kennis. Grafische representaties van de concepten en relaties bleken te verworden tot een indrukwekkende kluwen van ballen, pijlen en kleuren. Het meest geschikt voor onze doeleinden bleek een plugin voor Protégé, OWLDoc, die een website genereert waarin de structuur via de 'is-a'-relatie duidelijk zichtbaar wordt gemaakt, en waarmee alle concepten afzonderlijk eenvoudig kunnen worden bekeken (zie figuur 1). Een website kan bovendien makkelijk online toegankelijk gemaakt worden voor alle betrokkenen.

Terugkoppeling door de inhoudelijk deskundigen zowel via e-mail als mondeling zorgde voor veel verbeteringen van de ontologie. Een wiki-pagina van elk concept voor het plaatsen van suggesties en het vasthouden van discussies werd niet erg veel gebruikt, maar die werd ook pas in een laat stadium van het project ingevoerd. Hoewel de interesse van verschillende wetenschappers in het project eerder uitgaat



Saskia Smidt (Universiteitsbibliotheek Amsterdam)



Presentatie Renze Brandsma (Universiteitsbibliotheek Amsterdam)

naar het perspectief van de bezoeker van een voorstelling dan naar de kunstenaar als creatieve oorsprong, is als centraal ordeningsprincipe toch gekozen voor de klassieke indeling in werk, manifestatie en exemplaar. Deze ordening laat de beschrijving van de complexe verhoudingen tussen werken, producties en uitvoeringen toe en wordt ook gehanteerd in de *Functional Requirements for Bibliographic Records* (FRBR) van de *International Federation of Library Associations and Institutions* (IFLA). De begrippen werk, manifestatie en in mindere mate exemplaar worden in belangrijke mate bepaald door het perspectief van de beschrijver, dat in verschillende instellingen en collecties vaak anders is. Een theaterproductie, bijvoorbeeld de encensering van Hedda Gabler door Toneelgroep Amsterdam onder regie van Ivo van Hove, kan gezien worden als de realisatie van het toneelstuk Hedda Gabler van Henrik Ibsen als geschreven werk. De afzonderlijke uitvoeringen in het seizoen 2006/2007 zijn de bijbehorende exemplaren. Dit is het perspectief waarin we de nadruk leggen op Ibsen als auteur van het werk en de theaterproductie hiervan als afgeleide zien. We zouden echter ook de nadruk kunnen leggen op de creatieve prestatie van de regisseur of producent en de theaterproductie zien als een werk met als makers de regisseur Van Hove, en de producenten Toneelgroep Amsterdam en het Holland Festival. In de ontologie zijn beide perspectieven toegestaan: een theaterproductie is zowel een werk als een manifestatie en heeft eigenschappen van beide categorieën. Een theaterproductie kan exemplaren (uitvoeringen) hebben, het kan een werk realiseren, maar heeft ook een maker en een voltooiingsdatum. Hierdoor ontstaat een zeer flexibel beschrijvingsysteem.

van metadata naar ontologie, naar XML en een webinterface

Voor de verzamelde metadata van de instellingen zijn mapping-overzichten gemaakt voor het relateren van de velden van de aangeleverde metadata naar de begrippen (of klassen) van de ontologie. De eerste stap in het maken van deze overzichten is het mappen naar de hoofdelementen van de ontologie zoals 'Werk', 'Manifestatie' en 'Exemplaar'. In de volgende stappen zijn de velden van de aangeleverde metadata zo gedetailleerd mogelijk naar de essentiële en dieperliggende begrippen van de ontologie doorgezet. Bijvoorbeeld bij Personen naar Voornaam en Achternaam. In figuur 2 wordt een indruk gegeven hoe een dergelijk overzicht voor de mapping van elementen van Cinema Context naar de ontologie er uit ziet. Ook hier geldt dat het slechts een summier overzicht is met twee elementen, en niet het werkelijke complexere schema, terwille van de duidelijkheid en leesbaarheid van de tabel.

Cinema Context elementen	Overeenkomende <i>Cultuur in Context</i> ontologie begrippen, van detail (links) naar hoger niveau (rechts)		
Film (NL) =	Filmversie →	Uitgave →	Manifestatie
Persoon =	Man of vrouw →	Persoon →	Agens

Figuur 1 Summier overzicht van Cinema Context naar ontologie

Verder is getracht om de relaties, die in de diverse systemen van de instellingen aanwezig zijn, te bewaren en te converteren naar de centrale databank bij het DPC. Bijvoorbeeld als er een relatie is tussen personen en een gebouw in de aangeleverde

metadata is, dan zal deze relatie ook terugkomen in de centrale databank. Door het mappen van de metadata van de aangeleverde bestanden naar de ontwikkelde ontologie is de ontologie aan een grondige toets onderworpen en kon wederom een groot aantal aanvullingen en verbeteringen aangebracht worden.

De aangeleverde metadata werd na conversie naar de ontologie in een centrale databank bij het DPC via het web doorzoekbaar en beschikbaar gesteld. De technische infrastructuur van het DPC is mede gebaseerd op XML-bestanden en een zoekmachine die kan zoeken op elementen en attributen van XML. Om gebruik te maken van deze technische infrastructuur werd de metadata voor de webinterface verzameld in een aantal algemene eenheden (XML-containers) geïnspireerd op het principe van wie, wat, wanneer en waar. Deze indeling is nog verder gerealiseerd met behulp van de ontologie hoofdbegrippen. Per hoofdbegrip van de ontologie zijn er XML-containers gedefinieerd, waaronder containers voor locatie, persoon, rechtspersoon, vertoonbaar werk, etcetera. De XML-containers zijn zo gedetailleerd mogelijk gedefinieerd in overeenstemming met de begrippen in de ontologie, waarbij de gebruikte benamingen voor de elementen en attributen in de XML afkomstig zijn uit deze ontologie. Gegevens die niet essentieel genoeg waren of niet generiek voorkomen in de diverse databanken van de instellingen zijn in een soort verzamelrubriek in de XML-containers terecht gekomen. Bijvoorbeeld de titulatuur of gevolgde opleiding komt bij personen in overig. Bij het maken van de XML-containers moest helaas wel rekening gehouden worden met een aantal beperkingen in de architectuur. Vooral het gebruik van complexe RDF-achtige notaties in de ontologie en de hoeveelheid verwijzingen vormen een probleem om terug te laten komen in de XML-infrastructuur en architectuur van de webapplicatie. Uiteindelijk is een simpele zoekinterface voor de centrale databank gebouwd (cultuur-context.dpc.uba.uva.nl/). In deze webinterface is een zoekopdracht mogelijk door te zoeken in alle verzamelde metadata (en daarmee alle containers of hoofdbegrippen) of de zoekopdracht te beperken tot een bepaalde container (bijvoorbeeld alleen zoeken in 'persoon en rechtspersoon'). Het resultaat van de zoekopdracht is een lijst met de gevonden records. Hierbij wordt het volledige record getoond. Een volledig record geeft alle gegevens weer, inclusief de naam van de container en ander begrippen van de ontologie waar de gegevens toe behoren. Indien aanwezig worden eveneens links weergegeven naar andere gerelateerde gegevens (zoals bij filmvertoning naar locatie). Bij de weergave van de gegevens wordt eveneens de naam van de instelling getoond, om aan te geven waar de metadata vandaan komt. In de weergave is ook de functie 'Context ?' opgenomen. Als men hierop klikt dan krijgt men de context te zien (indien aanwezig). Bijvoorbeeld: zoek op 'Chris van Dinteren' in 'persoon en rechtspersoon'. In de context zie je zowel zijn werk voor bioscopen als zijn activiteiten als muziekschrijver van theaterproducties. Men kan kiezen voor de weergave in XML of in HTML. In de weergave in HTML wordt een versimpelde naamgeving voor de begrippen van de ontologie gebruikt ten bate van de leesbaarheid en duidelijkheid voor de eindgebruiker.

relaties en context

Een nadeel van bovenstaand methode is dat de metadata wel gemapt worden naar een uniform datamodel (de ontologie), maar dat er geen relaties gelegd worden als die niet al in de metadata aanwezig zijn. Als bijvoorbeeld het Stadarchief Breda metadata van personen aanlevert, en daar zit de naam van een schouwburgdirecteur bij, dan

wordt niet automatisch de relatie gelegd tussen de schouwburggegevens die het TIN heeft aangeleverd. In Cinema Context is een uitgebreide redactie bezig geweest om de gegevens te selecteren, te analyseren en de relaties te leggen. In het *Cultuur in Context* programma was niet voorzien in redactionele werkzaamheden op die grote schaal. Wel bleek het mogelijk om op kleine schaal als demonstratie een paar van dergelijke relaties handmatig te leggen. Dit is gedaan voor een aantal personen of rechtspersonen, gebouwen, werken of producties. De meest overeenkomende databases, die van Cinema Context en het TIN zijn gebruikt om de meeste relaties te leggen tussen personen en gebouwen. Een aantal producties van het Theaterinstituut zijn vergeleken met de toneelvoorstellingen die onderdeel vormden van het filmprogramma in Cinema Context. Zoek bijvoorbeeld op Luxor in de collectie "bioscoop en theater". Er zijn 81 Luxors. Men moet de Luxor in Rotterdam hebben (B000924). Als men daar de context van opvraagt, ziet men dat er 1289 voorstellingen zijn geweest waarvan de laatste theatervoorstellingen zijn. Hieruit kan men concluderen dat er dus niet alleen maar filmvoorstellingen in een bioscoop werden gegeven.

Ook zijn een aantal theaterproducties vergeleken met de bestanden van de Universiteit Utrecht. Hier ontstond echter de moeilijkheid dat de data die Universiteit Utrecht in de loop der jaren heeft verzameld vooral de periode tot 1900 beslaat.

Het productiebestand van het TIN bevat weinig materiaal van voor 1900. De data geleverd door de archieven bestaan voornamelijk uit archiefbeschrijvingen of inventarissen. Deze kunnen gekoppeld worden aan de data in de *Cultuur in Context* database als bron. Dit verbreedt niet zo zeer de directe context van de 'openbare gemaklijkheden' maar geeft de gebruiker een makkelijk overzicht waar nog meer informatie te vinden is over wat gezocht en gevonden is.

Bij het leggen van relaties tussen bestanden moet goed in acht worden genomen dat de interpretatie van de waarde van de relatie bij de gebruiker/wetenschapper ligt. Daarom is het element uitspraak opgenomen, waarmee wordt gezegd dat volgens een met naam genoemde bron deze twee records naar grote waarschijnlijkheid naar dezelfde instantie verwijzen. Zo wordt ruimte voor interpretatie opengelaten.

discussie en lessons learned

De ICT-systemen van de partners voldeden vaak niet aan open standaarden en standaarduitwisselingsprotocollen om op een geavanceerde technische manier een virtuele collectie te creëren. Koppeling van collecties is wel mogelijk gebleken door middel van het toepassen van meer basale technieken. Door de samenwerking is men zich wel bewust geworden van het belang van standaard uitwisselingsprotocollen (OAI-PMH en SRU) waar ICT-systemen aan moeten voldoen en men zal daar rekening mee houden bij het aanschaffen van software.

In een paar gevallen was het lastig om metadata te verkrijgen van de partners in het project. Dit kan meerdere oorzaken hebben, maar had waarschijnlijk te maken met de beperkte personele inzetbaarheid of het feit dat instellingen bezig waren hun eigen systemen te vernieuwen, waardoor er maar weinig tijd over bleef voor het *Cultuur in Context* project. Een duidelijk vastgelegd commitment van en ondersteuning door het management van deze partners aan het begin van het programma *Cultuur in Context* was wellicht beter geweest. Kortom, de organisatorische interoperabiliteit verdient meer aandacht.

De aangeleverde metadata waren niet altijd even consistent en van goede kwaliteit.

Dit komt onder andere doordat de betreffende databanken waaruit de metadata is geleverd over langere periodes (decennia) zijn opgebouwd. Bovendien is er ook sprake van een cultuurverschil tussen wetenschappelijke onderzoekers en erfgoedprofessionals van niet wetenschappelijke instellingen. Andere doelstellingen leveren andere databestanden op. Dit maakt koppeling gecompliceerder.

Een ontologie als die voor *Cultuur in Context* staat niet op zichzelf maar raakt aan andere gebieden van cultureel erfgoed. Een integratie van data binnen dit specifieke gebied is dus uiteindelijk een begin voor een verdere integratie die toegang geeft tot andere historische data. Het CIDOC initiatief (cidoc.ics.forth.gr/) is een poging om voor dit veel omvattender gebied een ontologie te ontwerpen. Een verdere integratie van de *Cultuur in Context* ontologie in CIDOC zou zeker nuttig zijn. Een ander punt zijn de hulpmiddelen voor het maken en gebruiken van de ontologie door onderzoekers of inhoudsdeskundigen. De huidige software is eigenlijk niet te gebruiken voor mensen die wel inhoudelijke kennis bezitten maar kennis ontberen van ontologietechnologie. Met de nodige beperkingen en enig kunst-en-vliegwerk is dit in het project geregeld met bestaande hulpmiddelen.

Op een ander niveau bleek het volledige gebruik van de ontologie niet haalbaar. Er is uiteindelijk gekozen voor een op de ontologie gebaseerd XML-containermodel dat makkelijker ondersteund kon worden met de bestaande XML-database infrastructuur van de Universiteitsbibliotheek van Amsterdam. De verwachting is dat met het toenemende belang van RDF en ontologieën de architectuur hiervoor geschikter wordt gemaakt.

In het project is de koppeling tussen databestanden handmatig op kleine schaal gedaan. Bij vervolgprijekten waarin context van belang is zullen op grotere schaal deze redactionele werkzaamheden plaats moeten vinden. Ook dienen er authority lijsten voor persoonsnamen, gebouwen et cetera gemaakt te worden. Er kan dan gekozen worden voor een uitgebreide redactie die bij projecten gevormd wordt en de context aanbrengt in de centrale databank. Anderzijds kunnen werkverbanden worden aangegaan, waarbij instellingen van de diverse sectoren samenwerken bij het aanbrengen van context. Als voorbeeld kan de samenwerking bij de universiteitsbibliotheken gelden die een gezamenlijke authority lijst aanleggen voor auteurs en onderzoekers bij hun instellingen. Iedere auteur of onderzoeker werkzaam bij een universiteit krijgt daarbij een uniek onderscheidbaar nummer: de Digitale Auteurs Identifier (DAI). Deze DAI wordt opgenomen als element van de metadata in diverse bestanden (waaronder bestanden met publicaties en bestanden met onderzoeksgegevens). Met de DAI kan men de context tussen de online bestanden automatisch verzorgen.

conclusie

De koppeling van collecties is mogelijk gebleken door middel van het toepassen van basale technieken. Voor een geautomatiseerde koppeling op grote schaal is het echter van belang dat systemen van culturele erfgoedinstellingen moeten voldoen aan standaard uitwisselingsprotocollen (OAI-PMH en SRU). De ontologie heeft meer helderheid verschaft over wat data in een database nu eigenlijk betekenen en in het creëren van een begrippenapparaat waarin dat uitgedrukt kan worden. Als laatste is gebleken dat het aanbrengen van context tussen de data van verschillende instellingen voorlopig nog handmatig door een redactie uitgevoerd moet worden.