

Een ontologie van openbare gemakkelijkheden

In het novembernummer van InformatieProfessional werd in het eerste deel van een tweeluik over het project Cultuur in Context de balans opgemaakt. In dit project worden databestanden van een aantal erfgoedinstellingen en wetenschappelijke instituten aan elkaar gekoppeld. Om die data te kunnen koppelen, ontwikkelden de samenwerkende partijen een ontologie. Rogier Brussee, Harry van Vliet en Christian Wartena doen hiervan verslag.

Rogier Brussee, Harry van Vliet en Christian Wartena

Het collectieve geheugen van Nederland is 'gestold' in ons cultureel erfgoed: een enorme hoeveelheid schilderijen, boeken, archieven, archeologische overblijfselen en audiovisuele objecten is beland in een groot aantal verschillende gebouwen en collecties. Deze fysieke gebondenheid legt beperkingen op aan bezoekers en onderzoekers. Zo zijn gerelateerde objecten bijvoorbeeld vaak opgeslagen op verschillende locaties, zodat óf de bezoeker/onderzoeker, óf het object moet reizen. De digitalisering golf die eind vorige eeuw is ingezet, leek het breekijzer voor verandering in de toegang tot het cultureel erfgoed.

Digitale toegang krijgt pas meerwaarde als het toegang is tot interessante informatie. Een digitale versie van een cultureel object dient daarom zoveel mogelijk het isolement te doorbreken. Dat kan door het in te weven in een web van relaties met andere culturele objecten én informatie over zijn context. Een dergelijke verrijking kan de ervaring van de bezoeker verdiepen, de kans op herkenning vergroten en aanzetten tot reflectie. Dit geldt zeker ook voor de professionele bezoeker van het culturele erfgoed: de wetenschapper, voor wie gerelateerde gegevens vaak de primaire onderzoeksinteresse. Een meer omvangrijke 'data-

pool' met rijkere informatie draagt bij aan analyses die opvattingen gebaseerd op gefragmenteerde, incoherente en geïsoleerde informatie, aanscherpen dan wel weerleggen.

Digitale collecties zijn heel geschikt voor zo'n brede kijk op cultureel erfgoed, omdat ze gerelateerd kunnen worden aan andere digitale informatie. De digitale infrastructuur moet dan wel óver grenzen van instituten, collecties en informatiesystemen heen worden versterkt. Maar digitalisering is veelal net zo verkokerd binnen de muren van de culturele instellingen als de fysieke collecties, zodat de cultuurliefhebber nog steeds langs ver-

schillende (digitale) toegangspoorten moet. Bovendien is de digitaliseringslag vooral gericht geweest op het digitaliseren van zoveel mogelijk objecten. De beschrijving van de individuele objecten is meestal minimaal vanwege de grote hoeveelheden materiaal en de benodigde arbeid om er uitgebreide en juiste metadata aan toe te voegen. Alles wijst daarom op een noodzakelijke volgende stap in het digitaliseringsdenken: minder vanuit het eigen collectiebeheer en meer gestuurd op de verbondenheid van collecties en het overkoepelende verhaal.

Cultuur in context plaatsen

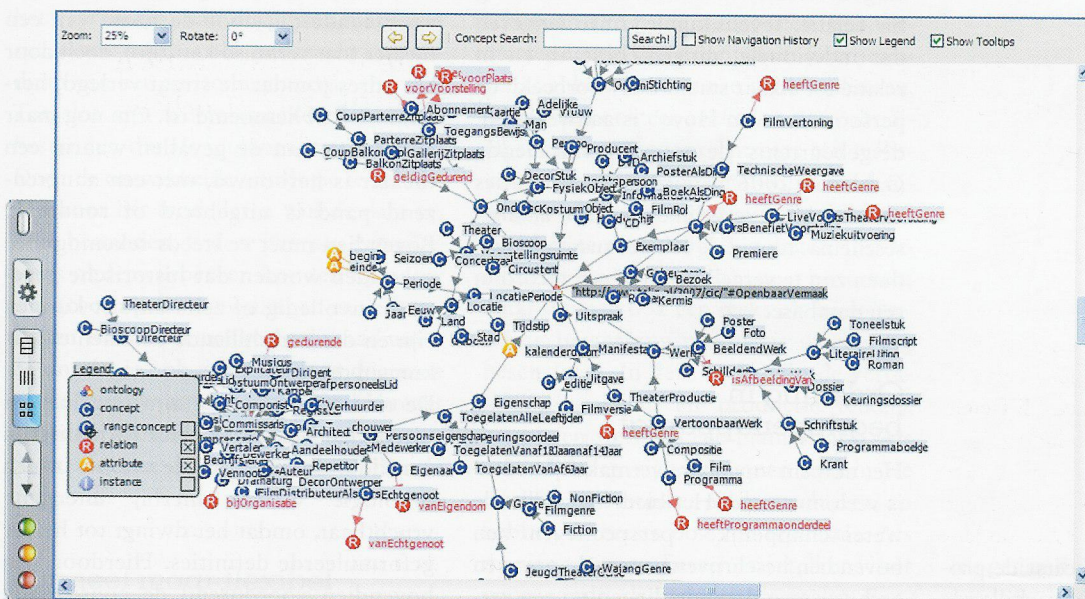
De bredere culturele en sociaal-economische context is bij uitstek van belang voor culturele gebeurtenissen zoals theater- en filmvoorstellingen. Het project Cultuur in Context (CiC) ambieert een groter en geïntegreerd publieksbereik van gedigitaliseerde collecties voor het domein 'openbare gemakkelijkheden', met een nadruk op de periode van circa 1800 tot 1950.* De term 'openbare gemakkelijkheden' wordt behalve als overkoepelende term voor theater- en filmvoorstellingen, ook gebruikt voor de kermis waar vroeger eveneens dergelijke voorstellingen plaatsvonden.

'Digitalisering is net zo verkokerd binnen de culturele instellingen als de fysieke collecties'

[illegible]

Screenshot van de door OWLDoc
gegenereerde html-documentatie.

Let op de koppeling naar een
wikipagina waar commentaren
kunnen worden toegevoegd



Screenshot van een visualisatie in OntoStudio van concepten en hun relaties in het domein 'openbare vermakelijkheden'

De bewaarde tekst of film is maar één aspect van een (film)voorstelling. De praktijk van filmdistributie, de plek van de voorstelling, de samenstelling van het publiek, de regieaanwijzingen, de acteurs en de reacties in de pers en dergelijke zijn minstens even interessante 'objecten' in het geheel van de voorstelling. Dit is nergens zo duidelijk als in het Theater Instituut Nederland (TIN). Een voorstelling is een gebeurtenis, een ontmoeting van spelers en toeschouwers in een ruimte, die na afloop voorgoed is verdwenen. Voor het TIN blijft dus een grote leegte waar andere musea een overdaad aan kunstvoorwerpen kunnen tonen. In plaats daarvan is het TIN een expert geworden in het verzamelen van context in paraferalia, zoals decortekeningen, foto's,

kostuums, tekstboekjes, geluidsbanden, portretten, affiches en recensies waarmee je een voorstelling kunt reconstrueren en documenteren. Ook hun database is volledig ingericht op contextinformatie en verbindt de meest uiteenlopende objecten met elkaar.

Ontologie als bindmiddel

Een belangrijke doelstelling van het CiC-project is het integreren van data van verschillende onderzoekers en verschillende instellingen, zodat deze op een geünificeerde manier kunnen worden doorzocht en bekeken. Data-integratie is overigens pas nuttig als er op betekenisniveau overeenkomsten en verbanden zijn. Binnen het CiC-project zijn dat bijvoorbeeld

acteurs die in films en theaters hebben gespeeld, of gebouwen waar zowel film- als theatervoorstellingen werden gegeven. Het integreren van data betekent dat die instellingen en onderzoekers hun data beschikbaar moeten stellen. Idealiter blijft iedereen daarbij zijn eigen data beheren en vindt op het moment van zoeken een vertaalslag plaats.

Deze aanpak heeft echter vele technische haken en ogen. Denk bijvoorbeeld aan toegangsrechten, communicatieprotocollen en veiligheid van computernetwerken die niets met het onderwerp te maken hebben. Er is daarom voor gekozen om data te converteren naar een gemeenschappelijk model en deze in een door de Universiteit van Amsterdam beheerde database op te slaan.

‘Een noodzakelijke volgende stap in het digitaliseringsdenken is het sturen op de verbondenheid van collecties’

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 OWL. OWL is een taal ontworpen voor de W3C, de organisatie die de technische standaarden voor het web beheert. Je kunt ermee uitdrukken dat in de wereld klassen van ‘dingen’ en relaties daartussen kunnen worden onderscheiden, bijvoorbeeld personen, documenten en gebeurtenissen. En ook de ‘is aanwezig bij’-relatie. Tevens kun je ermee aangeven dat individuele ‘dingen’ bestaan en in relatie tot elkaar staan. Een voorbeeld: de persoon ‘Ivo van Hove’ ‘is aanwezig bij’ de gebeurtenis ‘de première van Hedda Gabler in 2006’. De klassen en relaties zijn daarbij te vergelijken met een databaseschema, terwijl de relaties tussen individuen zijn te vergelijken met een record in een database.

Ordering in begrippenapparaat

Het domein ‘openbare vermakelijkheden’ is veelomvattend. Het biedt verschillende wetenschappelijke perspectieven en 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 de historici worden 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 prosopografische bestanden met data over zitplaatsen, abonnementen, directies en programmering van theatervoorstellingen in Rotterdam van het Instituut voor Media en Cultuurwetenschappen van de Universiteit Utrecht. Prosopografische bestanden bevatten gegevens van

gemeenschappelijke kenmerken van een bepaalde maatschappelijke groep mensen, zoals beroep, inkomen, sociale afkomst, stand en woonplaats.

Ordering in het begrippenapparaat bleek de moeilijkste fase in het project. Het gedeelde uitgangspunt was vragen te kunnen beantwoorden omtrent: wie, wat, waar en wanneer. 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. Zelfs een plaatsaanduiding is noch eenduidig door de naam van een theater (dat verhuisd kan zijn), noch door het adres (omdat de straat verlegd, hernummerd of hernoemd is). Om nog maar te zwijgen van de gevallen waarin een theater is herbouwd, met een aangrenzend pand is uitgebreid of rondreist. Bovendien moet er steeds rekening mee gehouden worden dat historische gegevens onvolledig of zelfs onjuist kunnen zijn en dat verschillende bronnen elkaar kunnen tegenspreken.

De meeste problemen zijn opgelost na uitvoerige discussies. 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 voordurend de consistentie met eerder genomen beslissingen worden getoetst. Wel zagen de verschillende deelnemers zich regelmatig geconfronteerd met de beperkingen van een formele taal als OWL, waarin geen ruimte is voor nuances, kanttekeningen en voorbehouden.

Verdere ontleding

Na het vastleggen van een hoofdstructuur was de belangrijkste opgave het uitwerken van details voor specifieke deelgebieden. Daarvoor zouden verschillende deelnemers gelijktijdig aan de ontologie moeten 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

Op die manier kunnen allereerst de problemen die unificatie van verschillende data met zich meebrengt onderzocht worden. Om 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 er tussen, te kunnen uitdrukken. Hiertoe is een ontologie ontwikkeld.

Formalisering van de werkelijkheid

Een ontologie is een formalisering van de werkelijkheid. Ontologiën geven dus niet zozeer betekenis aan begrippen, maar leggen een begrippenkader vast en documenteren eigenschappen en betekenis van die begrippen. De eigenschappen van een begrip dat 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.

indrukwekkende kluwen van bollen, pijlen en kleuren.

Het beste geschikt voor onze doeleinden bleek een plugin voor Protégé, OWLDoc. Deze plugin genereert een website waarin de structuur via de 'is een'-relatie duidelijk zichtbaar wordt gemaakt en waarmee alle concepten afzonderlijk eenvoudig kunnen worden bekeken. Een website kan bovendien makkelijk online toegankelijk gemaakt worden voor alle betrokkenen.

Terugkoppeling door de inhoudelijk deskundigen via e-mail en mondeling overleg zorgde voor veel verbeteringen van de ontologie. Een wikipagina voor elk concept voor het plaatsen van suggesties en het vasthouden van discussies bleek niet succesvol, maar is ook pas in een laat stadium van het project ingevoerd.

Tot slot zijn we teruggekeerd naar de bestanden die als uitgangspunt dienden. Door de ontwikkelde ontologie te gebruiken voor het modelleren van de data uit deze bestanden, is deze aan een grondige toets onderworpen en konden opnieuw een groot aantal aanvullingen en verbeteringen worden gemaakt. In dit stadium is de ontologie ook gebruikt door docenten en studenten van de Hogeschool Utrecht als basis voor een user interface die uniform toegang geeft tot de verschillende collecties.

Formeel maar flexibel

De interesse van verschillende wetenschappers in het project gaat eerder uit naar het perspectief van de bezoeker van een voorstelling dan naar de kunstenaar als creatieve oorsprong. Toch is als centraal ordeningsprincipe 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 diverse instellingen en collecties vaak verschillend is. Een theaterproductie, bijvoorbeeld de encenering 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 en de theaterproductie zien als een werk met als maker de regisseur Van Hove.

In de ontologie laten wij beide perspectieven toe: 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 weliswaar formeel maar zeer flexibel beschrijvingssysteem. We hoeven ons bijvoorbeeld niet meer tot drie lagen (boek-encenering-uitvoering) te beperken, maar kunnen ook een reeks als boek-vertaling-bewerking-productie-uitvoering beschrijven. In deze reeks zijn de vertaling, bewerking en productie telkens zowel een werk als een manifestatie. Als werk hebben zij telkens een maker (de vertaler, de bewerker en de producent) en hebben de eerste ook een realisatie. Als manifestatie realiseren zij telkens het voorgaande element uit de reeks. Eenzelfde dubbele rol vinden we terug bij een aantal andere concepten, zoals de verfilming. Op meerdere plekken in de ontologie zijn dit soort overwegingen en beslissingen terug te vinden.

Ontologie als Zwitsers zakmes

In Cultuur in Context heeft het ontwerpen van een ontologie meerdere doelen gediend. Het belangrijkste is geweest dat het als katalysator diende voor de inhoudelijke discussies over begrippen in het domein van openbare vermaken. Dit is zeer nuttig gebleken juist door de dwang van de ontologie tot een formalisering te komen. In dit proces is ook ervaring opgedaan met ondersteunende tools. Deze tools laten nogal wat te wensen over op dit moment.

De ontologie heeft nog twee verdere doe-

'Ordering in het begrippenapparaat bleek de moeilijkste fase in het project'

len gediend. Zo heeft zij bij de technische implementatie van de data-integratie geholpen door de betekenis van de data te duiden. Verder zijn de belangrijkste concepten in de ontologie de basis geweest voor de ontwikkeling van een interface die uniform toegang geeft tot verschillende data van verschillende partijen. <

* Zie het eerste artikel van dit tweeluik over het project Cultuur in Context: Nancy van Asseldonk, 'Cultuur in Context. Erfgoeddata in nieuwe samenhang'. In: *InformatieProfessional*, november 2008.

Met dank aan de deelnemers van het Cultuur in Context-programma voor hun bijdragen.

Rogier Brussee en Christian Wartena zijn verbonden als wetenschappelijk onderzoeker aan het Telematica Instituut in Enschede. Harry van Vliet is senior wetenschappelijk onderzoeker bij het Telematica Instituut en Lector Cross-media Content aan de Hogeschool Utrecht, Faculteit Communicatie en Journalistiek.