

Het huis dat Japie heeft gebouwd

Cultuurhistorie op weg naar het semantisch web

Ronald Wiemer (RACM)
Hans Nederbragt (Trezorix)
juni 2007

The house that Jack built
This is the farmer sowing his corn,
That kept the cock that crowed in the morn,
That waked the priest all shaven and shorn,
That married the man all tattered and torn,
That kissed the maiden all forlorn,
That milked the cow with the crumpled horn,
That tossed the dog,
That worried the cat,
That killed the rat,
That ate the malt
That lay in the house that Jack built.

Inleiding

In 2003 bracht de raad van cultuur advies uit over de wijze waarop de erfgoedsector de digitalisering en informatisering in samenhang zou moeten gaan aanpakken. De raad schrijft o.a. (eCultuur: van i naar e, *Advies over de digitalisering van cultuur en de implicaties voor cultuurbeleid*, Raad voor Cultuur, juni 2003)

Op internet kunnen kennis en collecties van verschillende instanties via zoekmachines en hyperlinks met elkaar worden verbonden. Bovendien kan content van de ene partij op de website van een ander in een nieuwe context worden geplaatst – en zo een andere betekenis of een ander en groter publiek krijgen.

De toenmalige staatssecretaris van cultuur, Medy van der Laan heeft dit advies overgenomen in de beleidsbrief “meer dan de som” van november 2003. In feite sluit zij hierbij aan bij een ontwikkeling die ook in breder kader plaatsvindt en dat te maken heeft met het werken in ketens in een genetwerkte samenleving.

KiCH (Kennis Infrastructuur Cultuurhistorie) is een ontwikkeling die weliswaar niet ontstaan is vanuit deze ontwikkeling maar er wel op meedeint. In KiCH gaat het er om Cultuurhistorische kennis toepasbaar te maken in het domein van de planvorming en ruimtelijke ordening. Daarnaast probeert KiCH het ontstaan van samenhang te stimuleren door te voorzien in een koppelingsmechanisme.

Dit koppelingsmechanisme is in eerste instantie een administratieve koppeling, die door bronhouders zelf moet worden gelegd. In de erfgoedsector is dit niet altijd voldoende. Er zijn tal van bronnen die de juiste structuur ontberen waardoor de juiste koppelvlakken lijken te ontbreken. Omdat voor de beleving van erfgoed contextualisering een wezenlijk aspect is, is binnen KiCH gezocht naar een methode waarbij informatie in een zodanige structuur wordt gegoten dat deze koppelvlakken vanzelf en ondubbelzinnig naar boven komen drijven.

Dit lukt echter alleen maar wanneer verschillende kennisdomeinen gebruik maken van dezelfde semantische standaarden, of wanneer er een voorziening is die thesauri, woordenlijsten en andere begrippenkaders aan elkaar koppelen. Dit laatste is het onderzoeksthema in het project Referentie Netwerken Architectuur.

Het kinderversje waarmee dit artikel begint, laat goed zien op welke wijze koppelvlakken in een keten kunnen worden vormgegeven. Op een vergelijkbare wijze is in een experiment van KiCH in samenwerking met het project Referentie Netwerken Architectuur een deel van de KiCH collectie gekoppeld met een collectie glazen voorwerpen.

Kennis Infrastructuur Cultuurhistorie (KiCH)

KiCH is ontstaan uit de beleidsnota Belvedere uit 1999, waarin geconstateerd werd dat het aanbod aan cultuurhistorische informatie sterk versnipperd was. Deze versnippering heeft tot gevolg dat cultuurhistorie in ruimtelijke ontwikkelingen vaak niet optimaal benut werd. Het ontwikkelingsprogramma KiCH moest hier een eind aan maken. In eerste instantie heeft KiCH zich beziggehouden met het samenbrengen van informatiebronnen op nationaal niveau. Deze bronnen hebben betrekking op archeologie, bouwhistorie en historische geografie. Hieruit is het webportaal KiCH (www.kich.nl) ontstaan, dat de gebruiker de mogelijkheid biedt zich een integraal beeld te vormen van de cultuurhistorie in ieder willekeurig gebied, zowel in een hedendaagse context, namelijk de huidige topografische kaart als in een historische. Voor dit laatste kan de gebruiker de cultuurhistorische elementen afbeelden op een kaartondergrond uit circa 1900.

Naast de fraaie kaarten die hiermee te maken zijn, maakt KiCH echter ook de beperking van de informatie duidelijk. De bronnen die nu toegankelijk zijn, zijn alle met een specifiek doel ontstaan en worden in het algemeen nog steeds vanuit dat perspectief beheerd. Dat betekent dat de informatiewaarde veelal hoog is binnen het kennisdomein waar de bronnen beheerd worden, terwijl voor andere gebruikers de begrijpelijkheid te wensen overlaat door het gebruik van jargon en het ontbreken van context.

Om dit probleem op te lossen is een aantal strategieën mogelijk. Ten eerste kunnen bronhouders gestimuleerd worden hun registraties zodanig te ordenen dat de toegankelijkheid (vindbaarheid en leesbaarheid) ook voor niet ingewijden gegarandeerd is. Omdat het vaak om grote collecties gaat die een lange traditie kennen vergt dit grote investeringen.

Een tweede mogelijkheid is de huidige bronnen zoveel mogelijk in te bedden in een context. Deze context kan worden geboden door objecten zoveel mogelijk met elkaar te verbinden. Hiervoor is het noodzakelijk dat er semantische relaties zijn tussen de objecten onderling. De structuur waarin deze semantische relaties worden gevat wordt een ontologie genoemd. Het CIDOC Conceptual Reference Model is zo'n ontologie. Het Referentie Netwerken Architectuur (RNA) Project past dit principe toe om

Het RNA project

Het RNA-project (Referentie Netwerken Architectuur) houdt zich op heel praktische wijze bezig met methoden, hulpmiddelen en technieken die gebruikt kunnen worden bij het realiseren van *dynamische kennissystemen*. In deze systemen is een belangrijke rol weggelegd voor *referentienetwerken*.

Een referentienetwerk is een verzameling *referentiestructuren* (thesauri, taxonomieën, trefwoordenlijsten, enzovoort, bedoeld is voor het vindbaar maken van webcontent) en gerelateerde content-metadata. Ook allerlei soorten *objectbeschrijvingen* (bijvoorbeeld uit collectiedatabases) zullen er vaak deel van uitmaken.

In deze context kan een *kennissysteem* beschreven worden als een combinatie van via het web beschikbaar gemaakte kennis enerzijds en een referentienetwerk anderzijds, en van alle relaties die tussen die twee bestaan. Zowel webcontent als referentienetwerk zijn dynamische verzamelingen. Het actualiseren van de relaties tussen beiden - noodzakelijk na elke mutatie waar dan ook in het kennissysteem - moet zo geautomatiseerd mogelijk kunnen plaatsvinden. Als daar aan voldaan wordt is er sprake van een *dynamisch kennissysteem*.

Het RNA-project richt zich vooral op twee vraagstukken met betrekking tot dynamische kennissystemen: (1) hoe kun je content op een betaalbare, flexibele en betrouwbare manier aan referentiestructuren koppelen, en (2) hoe kun je referentiestructuren op een betaalbare, flexibele en betrouwbare manier manieren samenvoegen tot een referentienetwerk dat een nauwkeurige en samenhangende vindbaarheid mogelijk maakt.

In het RNA-project wordt deelgenomen door een tiental culturele en kennisinstellingen, twee universiteiten en vier mkb-bedrijven. Het project is een SenterNovem PRIMA project, het staat op inhoudelijk gebied onder auspiciën van het Ministerie van OCW, en het loopt van begin 2005 tot en met eind 2007.

Meer informatie over het RNA-project is te vinden op www.rnaproject.org.

Het CIDOC-Conceptual Reference Model (CRM)

ICOM-CIDOC (*International Committee for Documentation of the International Council of Museums*) is het internationale platform voor de documentatiebelangen van musea en gerelateerde organisaties. Er zijn meer dan 750 leden in 60 landen. Onder auspiciën van CIDOC is sinds 1995 gewerkt aan de ontwikkeling van een structuur voor de beschrijving van het erfgoed. Inmiddels is de structuur gereed en geaccepteerd als ISO standaard (ISO 21127:2006). Het "Conceptual Reference Model" zoals deze structuur heet is een model waarin gebeurtenissen (*events*) centraal staan. Alle mogelijke vormen van gebeurtenissen zijn beschreven in het model. Voor het ontstaan van een fysiek object is er bijvoorbeeld een gebeurtenistype "E65 Creation", maar er zijn ook klassen die verplaatsing of verandering van eigendomsrecht beschrijven. Het *event* is zo belangrijk omdat in de erfgoedsector het *wanneer* zo'n belangrijke rol heeft. Het *wanneer* wordt in heel veel verschillende betekenissen

gehanteerd; de gebeurtenis maakt het mogelijk dit nauwkeurig te specificeren en daarmee de historie vast te leggen.

Elke entiteit of klasse heeft een nummer, herkenbaar aan de beginletter E. Naast gebeurtenis-entiteiten zijn er klassen die fysieke of virtuele zaken beschrijven, zoals personen, plaatsen en voorwerpen. Alle entiteiten kunnen aan andere entiteiten worden gerelateerd middels eigenschappen (*properties*). Ook de eigenschappen zijn geformaliseerd. Alle eigenschappen zijn gedefinieerd in twee verschijningsvormen omdat een eigenschap altijd twee richtingen kent; een auteur is schrijver van een boek, maar het boek is ook geschreven door een auteur. In totaal kent CIDOC-CRM 80 klassen die met 132 eigenschappen aan elkaar kunnen worden gerelateerd.

Het voert te ver om CCRM hier uitgebreid aan de orde te stellen, maar een voorbeeld verduidelijkt wellicht een en ander:

Het document waarin wordt uitgelegd hoe Dublin Core te vertalen is naar CCRM is geschreven door Martin Doerr. Martin Doerr is voorzitter van de CRM working Group en verbonden aan het *Institute of Computer Science, Foundation for Research and Technology, Vassilika Vouton* in Heraklion (ICS-FORTH, Griekenland) en tevens één van de grondleggers van CCRM. Het document kan als volgt in Dublin Core termen worden beschreven:

Type: text

Title: Mapping of the Dublin Core Metadata Element Set to the CIDOC CRM

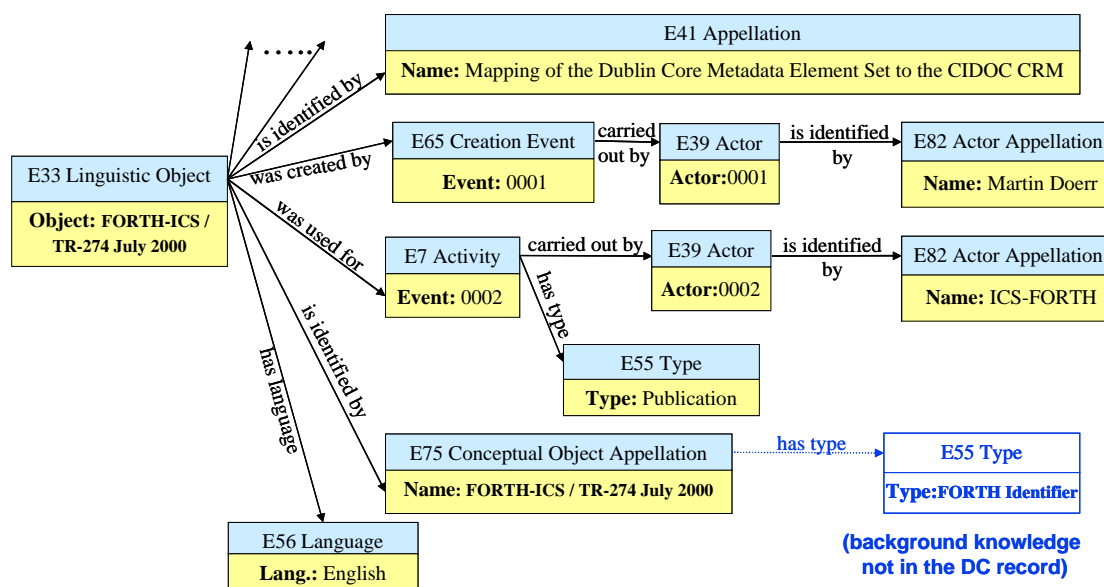
Creator: Martin Doerr

Publisher: ICS-FORTH

Identifier: FORTH-ICS / TR 274 July 2000

Language: English

In CCRM ziet dit er als volgt uit¹:



Duidelijk zijn de ketens te herkennen die ontstaan zijn door het koppelen van klassen met behulp van properties. Als zodanig voegt CCRM nog niet veel toe aan de DC beschrijving. De bedoeling is echter dat de bronhouder de betekenis van alle attributen probeert te expliciteren met behulp van de formulieren waarmee een registratie wordt bijgehouden of de bijbehorende documentatie. Wat hiermee wordt bereikt is dat informatie een eenduidige betekenis krijgt, ook buiten de context waarin wordt vastgelegd.

¹ (bron: Martin Doerr, Mapping of the Dublin Core Metadata Element Set to the CIDOC CRM, Technical Report FORTH-ICS/TR-274, Juli 2000)

In bovenstaand schema is dat bijvoorbeeld gedaan door een *E55 Type*-klasse te koppelen aan een *E75 Conceptual Object Appellation*. Deze kennis is niet aanwezig in de DC beschrijving, maar wel degelijk relevant buiten de context van ICS-FORTH.

Waar CCRM tevens op aanstuurt is dat eindpunten van ketens zo mogelijk administratief verwijzen naar (bij voorkeur universele) referentiebronnen. In de cultuursector zijn dat bijvoorbeeld de Universal List of Artis Names (ULAN, Getty), De Arts and Architecture Thesaurus (AAT, Getty) en de Theasurus of Geographic names (ook Getty).

Wanneer verschillende collecties conform de regels van CCRM worden gemodelleerd dan ontstaan koppelvlakken daar waar ketens overlappen. De mate van zekerheid over de koppelvlakken wordt bepaald door de hoeveelheid overlap. Wanneer het einde van een keten verwijst naar een referentiebron, dan is de mate van overlap met deze bron 100%.

CCRM wordt –ook in het buitenland- nog niet op brede schaal toegepast. Er zijn instellingsspecifieke implementaties in Oslo (The Museum Project), Helsinki (bronnen noemen) en Geneve (bronnen noemen).

Vanwege de behoefte naar contextualisering binnen KiCH én omdat Cidoc-CRM een ISO-standaard is, afkomstig uit de erfgoed sector en die bovendien gepositioneerd wordt als concept voor het semantisch web, is ook in Nederland een experiment uitgevoerd naar de bruikbaarheid ervan binnen de cultuurhistorie.

Wat is RDF?

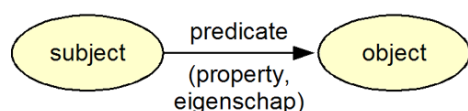
We zijn er van uit gegaan dat de in CCRM opgezette objectbeschrijvingen primair in een internet omgeving moeten kunnen functioneren. Twee standaarden zijn daar belangrijk: XML en RDF. Deze standaarden komen als het ware nog onder het CCRM-model te liggen.

XML (eXtensible Markup Language) is een standaard die is afgeleid van SGML (ISO 8879) en was oorspronkelijk bedoeld voor *electronic publishing* op grote schaal. XML is een simpel, bijzonder flexibel tekst formaat dat is uitgegroeid tot de belangrijkste standaard om data te kunnen beschrijven en deze via het internet uit te kunnen wisselen. (W3C Recommendation, 10 februari 1998).

RDF (Resource Description Framework) is een taal waarmee informatie over bronnen op het web uitgedrukt kan worden, zoals bijvoorbeeld *titel*, *auteur*, *copyright*, etcetera. Als je het concept “bronnen op het web” algemeen opvat, kan RDF veel ruimer gebruikt worden, namelijk om informatie te geven over dingen die op het web *geïdentificeerd* kunnen worden, ook al zijn ze niet letterlijk beschikbaar op het web. Producten van webwinkels zijn hier een goed voorbeeld van.

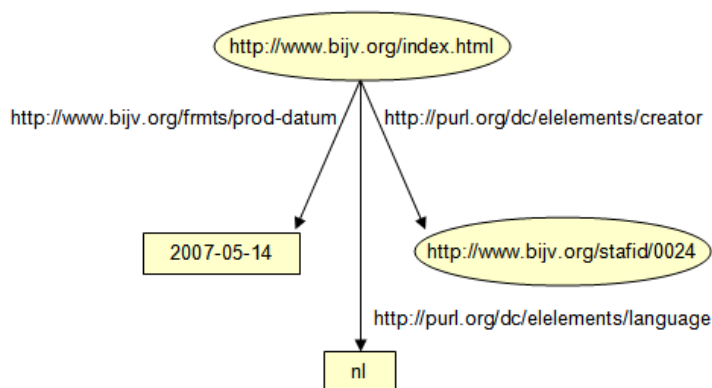
RDF is bedoeld voor situaties waarin deze informatie primair door software-applicaties verwerkt moet kunnen worden, dus niet in de eerste plaats door mensen. RDF zorgt voor een gemeenschappelijk kader voor het uitwisselen van deze informatie zonder dat oorspronkelijke betekenissen tijdens dit uitwisselingsproces verloren gaan. De mogelijkheid om informatie tussen verschillende applicaties te kunnen uitwisselen betekent dat ze beschikbaar gemaakt kan worden voor andere applicaties als die waarvoor ze oorspronkelijk gemaakt was.

RDF is gebaseerd op het idee dat “dingen op het web” door middel van een algemeen geldende standaard, de zogenaamde *Uniform Resource Identifiers* of *URI's*, geïdentificeerd kunnen worden en dat dergelijke bronnen beschreven kunnen worden in termen van eigenschappen (*properties*) en waarden van die eigenschappen. Het basiselement van het RDF model is het triplet: een bron (het *subject*) is gelinkt aan een andere bron (het *object*) door een verbinding die gelabeld is met een derde bron (*predicate*).



Een schema zoals hierboven wordt een RDF-graaf genoemd. We kunnen zeggen dat <subject> de *eigenschap* <predicate> heeft met *waarde* <object>. Een subject kan meerdere eigenschappen hebben en een eigenschap kan meerdere waarden hebben. In onderstaande afbeelding wordt de URI gegeven van een webpagina, met waarden voor de eigenschappen *creation-date*, *language*, en *creator*. Die eigenschappen zijn dus hier op zich ook vastgelegd in URI's.

Onderstaande RDF-graaf kan vertaald worden als: de webpagina <index.html> is aangemaakt op <14 mei 2007> in de taal <Nederlands> door een stafid met id <0024>.



RDF kan uitgedrukt worden in XML. RDF is door het flexibele gravenmodel uitermate geschikt als basis om complexe structuren in te beschrijven.

De RNA conceptuele architectuur

In het RNA-project wordt er gewerkt vanuit een architectuuropvatting die breder is dan CCRM. In deze architectuur wordt een onderscheid gemaakt tussen *domeinmodellen* en *applicatiemodellen*.

Een domeinmodel is een conceptueel datamodel dat als voornaamste doel representatie van kennis van een specifiek domein heeft. Voorbeelden uit het RNA-project van domeinmodellen zijn referentiestructuren (zoals de Art and Architecture Thesaurus), typologische beschrijvingen (zoals de RACM-typologie "glas") en objectbeschrijvingen (zoals de beschrijvingen van glazen objecten uit de pilot Nationale Referentiecollectie van RACM). CCRM kan opgevat worden als een bepaald domeinmodel.

Een applicatiemodel is een declaratieve structuur (in RDF) die een mapping aanbrengt tussen een domeinmodel en de entiteiten die een bepaalde applicatie feitelijk verwacht. In tegenstelling tot een domeinmodel bevat een applicatiemodel dus *taak-specifieke kennis*.

Om verschillende datasets aan elkaar te kunnen koppelen zijn er globaal gezien drie methoden van mapping:

- (1) directe mapping van bron-data naar een globaal conceptueel schema (bijvoorbeeld een combinatie van meerdere domeinmodellen);
- (2) elke dataset heeft een eigen lokaal schema, aanbrengen van verbanden gebeurt door direct mapping tussen deze schema's of door mapping via een overkoepelend globaal schema (bijvoorbeeld CCRM, e.d.);
- (3) elke dataset heeft een eigen lokaal schema, aanbrengen van verbanden gebeurt door mapping via een applicatiemodel.

Methode 2 en 3 lijken veel op elkaar, maar het grote verschil is dat bij methode 2 de mapping taak-onafhankelijk is, terwijl bij methode 3 de mapping juist gerelateerd is aan een bepaald applicatiemodel. Het is zeer goed mogelijk om een mix van alle drie de methodes toe te passen, en zoals we hieronder zullen zien is dat in de "KiCH-demonstrator" ook gebeurd.

Het experiment

Het experiment is er op gericht geweest een deel van de KiCH collectie te mappen naar CCRM en vervolgens samen te brengen met een gedigitaliseerde collectie glazen voorwerpen.

Deze glazen voorwerpen zijn afkomstig uit beerputten in de oude binnenstad van Dordrecht. Deze werden in vroeger tijden gebruikt als afvaldumpplaats. Ze vormen nu voor archeologen een schatkamer van informatie. Omdat de huizen waar de achtertuinen bij horen veelal nog bestaan en bovendien een monumentstatus hebben, vormen ze onderdeel van de verzameling KiCH objecten. Als er overeenkomstige informatie in beide collecties is, dan zouden de koppelvlakken vanzelf zichtbaar moeten worden wanneer beide collecties naar CCRM worden vertaald.

Mapping naar CCRM

De vertaling van een collectie naar CCRM wordt uitgevoerd in een aantal stappen. In de eerste stap worden tabelnamen en attributen vertaald naar object-triplets (zie figuur 1).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
2					Statement 1					Statement 3					
3															
4	Unit of Information	Class	D	Property	Class	D	Property	Class	D	Property	Class	D	Property	Class	
5	Rijksmonumentnr	E25 Man-Made feature (this monument)	R	P70 documents (is documented in)	E31 Document (this entry)	F	P1 is identified by (identifies)	E41 Appellation (Rijksmonumentnr)							
6					E31 Document (this entry)	R	P106 is composed of (forms part of)	E31 Document (the inventory as a whole)	F	P1 is identified by (identifies)	E41 Appellation (voorlopige rijksmonumentenkaa)				
7		E25 Man-Made feature (this monument)	f	P1 identified by (identifies)	E41 Appellation (Rijksmonumentnr)										
8	x_coord	E25 Man-Made feature (this monument)	F	P53 has former or current location (is former or current location of)	E53 Place (the place where this monument is located)	F	P87 is identified by (identifies)	E47 Spatial Coordinates (Rd X_coord + Rd y_coord)							
9	y_coord	E25 Man-Made feature (this monument)	F	P53 has former or current location (is former or current location of)	E53 Place (the place where this monument is located)	F	P87 is identified by (identifies)	E47 Spatial Coordinates (Rd X_coord + Rd y_coord)							
10	plaats	E25 Man-Made feature (this monument)	F	P53 has former or current location (is former or current location of)	E53 Place (the place where this monument is located)	R	P88 consists of (forms part of)	E53 Place (the place where the monument place is located in)	f	P87 identified by (identifies)	E48 Place Name (plaats)				
11										R	P88 consists of (forms part of)	E53 Place (the municipality the monument place is located in)			
12	gemeente						R	P88 consists of (forms part of)	E53 Place (the municipality the monument place is located in)	f	P87 identified by (identifies)	E48 Place Name	f	P3 Has note	E62 String (gemeente)
13										R	P88 consists of (forms part of)	E53 Place (the province the monument place is located in)			
14	provincie						R	P88 consists of (forms part of)	E53 Place (the province the monument place is located in)	f	P87 identified by (identifies)	E48 Place Name	f	P3 Has note	E62 String (provincie)
15													f	p3.1 has type	e55 Type (AMIC provincielijst)

Figuur 1

Deze objecttriplets worden afgeleid uit de tabellen en attributen die de oorspronkelijke bron beschrijven. Een triplet bestaat uit een entity-property-entity reeks. Triplets kunnen entiteiten gemeenschappelijk hebben en daarmee ketens vormen. Het geheel aan triplets kan worden weergegeven in een netwerk zoals is getoond het dublin core voorbeeld.

Deze stap lijkt eenvoudiger dan hij is. Cidoc-CRM kent twee ontwerpprincipes die in de meeste huidige registraties niet als leidraad hebben gediend:

- (1) de gebeurtenis staat centraal,
- (2) attributen hebben een eenduidige betekenis.

Indien gebeurtenissen als entiteit niet in de huidige collectie voorkomt, dan moeten deze worden afgeleid uit de tabellen en attributen die voorhanden zijn. In het experiment waar we het hier over hebben is tegen de regels gezondigd: er is geen gebruik gemaakt van het event-principe. Niet uit onwil, maar puur vanwege gebrek aan kennis en ervaring met Cidoc-CRM. Voor het verloop van het experiment is heeft dit echter geen gevolgen gehad.

Om aan het tweede ontwerpprincipe te voldoen is het noodzakelijk de inhoud van de registratie aan een nader onderzoek te onderwerpen. Zo blijken er tal van attributen te zijn waarvan de definitie nog veel ruimte overlaat voor verschillende interpretatie. Voor de bronhouder zelf hoeft dat geen probleem te zijn: vanuit het perspectief van bijvoorbeeld monumentenzorg is eenduidigheid niet altijd een noodzaak. Dat geldt bijvoorbeeld voor een attribuut als "naam object". Voor de monumentenzorgers is het adres, de kadastrale aanduiding of de coördinaten van het object van belang. Indien een gebouw een ook nog naam heeft dan is die naam een informatief gegeven en niet meer dan dat. De naam wordt niet beschouwd als identificerend attribuut. Eenduidigheid is geen noodzaak. Dat verandert als het gegeven in een andere context gebruikt wordt, bijvoorbeeld als element in een publiekswebsite, waar een willekeurige gebruiker een zoekactie uitvoert op basis van de naamgeving aan objecten. Dat is geen toevallig voorbeeld: in tal van (vooral historische) bronnen wordt naar plekken verwezen door middel van de naam van een locatie en niet naar het adres en ook een toerist die in Amsterdam op zoek is naar het Rijksmuseum zal zich laten leiden door de naam en niet door het adres.

Om een juiste vertaling van de eigen attribuutstructuur naar CCRM te maken kan het derhalve wenselijk zijn in de data zelf wijzigingen door te voeren, met het doel eenduidigheid te bereiken. Omdat het hier een experiment van beperkte omvang betrof is, is deze stap achterwege gelaten.

Implementatie in Sesame / facetnavigatie

De omzetting van de oorspronkelijke data naar het Sesame RDF-framework, inclusief facetnavigatie gaat als volgt:

De data zijn aangeleverd in spreadsheets. Inhoudelijke controle en eventuele aanvulling vond binnen dit formaat plaats. Ook omzetting naar het CCRM model werd binnen de spreadsheets gedaan. In de eerste plaats werden aan reeds bestaande database-velden CCRM-classes toegekend. En vervolgens werd de CCRM-structuur gecompleteerd door verdere CCRM-classes toe te voegen die voor het model nodig zijn. Voorbeeld van dit laatste zijn klassen die te maken hebben met "events" (hoewel we deze, wegens het veelal ontbreken van relevante data, zo spaarzaam mogelijk toegepast).

De ontstane CCRM-structuur werd vertaald naar een zogenaamd "map-bestand". Dit bestand bevat het datamodel in XML-formaat: de Classes, Properties en Attributes in een geneste structuur. Het wordt gebruikt als template voor de structuur van het op te bouwen RDF-bestand.

Via een gescipte conversie zijn vervolgens de spreadsheet en map-bestand gecombineerd tot een RDF-structuur met inhoud. Tegelijkertijd is er een "config-bestand" aangemaakt. Hierin worden gegevens voor de facetnavigatie opgeslagen: welke eigenschappen zijn er, wat zijn de daarbij behorende labels? Bovendien kan op basis van dit config-bestand een redactioneel bewerkbare pagina in het CMS worden gemaakt, waarin de set eigenschappen die bij de facetnavigatie ingezet wordt gewijzigd kan worden.

Tenslotte is het RDF-bestand in een Sesame-repository geplaatst. De Spectacle facetnavigatie-functionaliteit zorgt hierbij voor bevraging van de repository en toelevering van resulterende data, afgaande op de in het CMS geselecteerde set eigenschappen.

De demonstrator

Het doel van de demonstrator is om te laten zien dat objectbeschrijvingen van verschillende objecten, allen beschreven in het RDF/CCRM formaat, zonder verdere speciale programmering aan elkaar gekoppeld kunnen worden. Voorwaarde daarbij is uiteraard dat de verschillende soorten objecten gemeenschappelijke kenmerken (*facetten*) hebben.

De demonstrator bestaat uit de volgende onderdelen:

- Ruim driehonderd objectbeschrijvingen van *monumenten*, oorspronkelijk vastgelegd in IMKICH formaat, omgezet naar RDF/CCRM formaat.
- Ruim driehonderd objectbeschrijvingen van *glazen archeologische voorwerpen* in RDF/CCRM formaat. Deze zijn ook te vinden in de site van de [Nationale Referentie Collectie NRC](#).
- Ongeveer 90 objectbeschrijvingen van *complexen*, archeologische vindplaatsen, in RDF/CCRM formaat. Deze complexen zijn de vindplaatsen waar bovengenoemde glazen voorwerpen gevonden zijn.

Gemeenschappelijke facetten

De enige kenmerken die de glazen voorwerpen en de monumenten uit de demonstrator *in principe* gemeenschappelijk hebben zijn locatiedata. Maar in de objectbeschrijvingen van de glazen voorwerpen zijn deze locatiedata niet rechtstreeks aanwezig.

Er is hier een tussenstap nodig: in de objectbeschrijvingen van de glazen voorwerpen is een nummer opgenomen dat verwijst naar een complex, een vindplaats. De objectbeschrijving van dat complex bevat wèl locatiedata: plaats en straatnaam.

Ook de objectbeschrijvingen van de monumenten bevatten plaats en straatnaam. Monumenten en glas voorwerpen kunnen dus via het complex tot op straatniveau aan elkaar gekoppeld worden.

Vanuit de RNA-architectuur bezien zijn er twee datasets, die beschreven zijn volgens domeinmodellen in CCRM-formaat, gekoppeld middels een model (schema) dat heel specifiek op de betreffende applicatie gericht is.

Conclusie

Modelmatig lijkt CCRM binnen bepaalde kennisdomeinen een antwoord te kunnen geven op de vraag hoe het semantisch web vorm te geven. Er zitten echter ook wel haken en ogen aan.

Ten eerste zijn er zaken die wat lastig te implementeren lijken. Zo kunnen klassen in CCRM de eigenschappen van meerdere andere klassen overerven, kent iedere property twee richtingen en is het zelfs mogelijk sluiproutes te definiëren om daarmee de uitgebreidere ketens te omzeilen als die niet nodig zijn. Bestaande databasesoftware kan slecht met deze eigenschappen omgaan.

Bovendien ontstond tijdens het experiment de indruk dat het Cidoc model zich niet heel makkelijk leent voor representatie in RDF, de "oerbouwsteen" van het semantisch web. De manier waarop properties en klassen van Cidoc in RDF gemodelleerd moeten worden zorgt voor een moeilijk leesbaar resultaat - je zou kunnen denken dat de twee uiteindelijk toch niet op "natuurlijke wijze" bij elkaar passen. Verdere experimenten op dit gebied zouden zeer nuttig kunnen zijn.

Er waren ook problemen met Cidoc op domein-specifiek gebied, bij het gebruik ervan door inhoudsdeskundigen. Deze problemen lagen op twee vlakken:

(1) het blijkt inhoudelijk een heel moeilijk model te zijn,

(2) er zijn vaak geen data voorhanden waarmee de specifieke punten (zoals de focus op *events*) makkelijk zichtbaar gemaakt kunnen worden.

Deze beide punten zorgen er voor dat omgang met Cidoc moeilijk te leren is en dat het lang duurt voor de goede eigenschappen naar boven gehaald kunnen worden.

Waar CCRM wel bijzonder behulpzaam kan zijn is als referentiemodel. Het is niet nodig een technische implementatie van CCRM uit te voeren, maar de bevindingen van een logische mapping van een eigen model kan gebruikt worden om verbeteringen aan te brengen in de eigen datastructuren. Bij bestaande gegevenscollecties leiden deze structuurwijzigingen tot veelal ingrijpende aanpassingen in de content en daarvoor ontbreken veelal de middelen. Maar bij het ontwerp nieuwe structuren is deze beperking er niet en dan kan CCRM zeker een positieve bijdrage leveren aan de eenduidigheid van de informatievoorziening.