

# NETWERK DIGITAAL ERFGOED

## Schets van een infrastructuur voor het gedistribueerd netwerk van erfgoedinformatie

---

Discussienota

Juni 2016

Opgesteld door:  
Enno Meijers

Gereviewed door:  
Wilbert Helmus (NDE)  
Joop Vanderheiden (RCE)  
Johan Oomen (B&G)  
Erik Groeneveld (Seecr)  
Kees Hendriks (Zuiderzeemuseum)  
Martijn Arts (Total Active Media)

# Inhoudsopgave

<b>Inleiding</b>	<b>3</b>
<b>Centrale voorzieningen</b>	<b>4</b>
1. Het netwerk van terminologiebronnen (“Knowledge Graph voor Digitaal Erfgoed”)	4
2. Het collectieregister	4
3. Het objectregister	5
<b>Een groeimodel met vier instapmomenten voor erfgoedinstellingen</b>	<b>6</b>
Configuratie A - ‘blind’ systeem	6
Configuratie B - ‘knowledge graph enabled systeem’	7
Configuratie C - ‘Linked Data enabled’ systeem	7
Configuratie D - ‘Full Linked Data capabilities’	8
<b>Opbouw en synchronisatie objectenregister</b>	<b>9</b>

## Inleiding

Dit document schetst een infrastructuur die uitgaat van een geleidelijke overgang naar een op linked data gebaseerde doelsituatie voor het beheren en zichtbaar maken van erfgoed informatie. Deze uitwerking is een concrete vertaling van de visie die beschreven is in het position paper 'Naar een gedistribueerd netwerk van erfgoed informatie'. In essentie richt de visie in het position paper zich op domeinoverstijgende voorzieningen die de gebruiker zo direct mogelijk naar de erfgoed informatie brengt met zo minimaal mogelijke bewerkingen en transformaties onderweg.

Een belangrijk uitgangspunt is om de informatie zo veel mogelijk in lokale registratiesystemen te laten en alleen de informatie die nodig is voor het vindbaar maken naar een overkoepelend niveau te brengen. Op dit overkoepelende niveau zijn alleen de relaties tussen objecten en centrale referenties (voor wie, wat, waar, wanneer, etc) bekend. Door het open en bruikbaar publiceren van erfgoed informatie bij de bron vervalt op termijn de noodzaak voor het kopiëren en bewerken in de route tussen bron en gebruiker.

## Centrale voorzieningen

In het kader van het NDE - Werkpakket Bruikbaar worden een aantal domeinoverstijgende voorzieningen gerealiseerd die leiden tot een gedistribueerd netwerk van erfgoedinformatie. Aan het bestaande landschap van aggregatoren worden extra functionaliteiten toegevoegd die als centrale voorzieningen gaan functioneren.

Op termijn betekent dit dat de diensten die de aggregatoren nu leveren (o.a. doorlevering naar Europeana, museale presentatie van de Collectie Nederland, thematische presentaties) meer en meer op basis van deze centrale voorzieningen ingericht kunnen worden. Afhankelijk van de mate waarin lokale registratiesystemen kunnen samenwerken met de landelijke voorziening zal een tussenstap via een regionaal of landelijk aggregatieplatform voorlopig nodig blijven.

De centrale voorzieningen worden opgebouwd uit een aantal (gedistribueerde) dataverzamelingen die beschikbaar zijn via de verschillende knooppunten en netwerken waaruit het Netwerk Digitaal Erfgoed bestaat. Concreet kunnen de volgende drie type dataverzamelingen onderscheiden worden:

### 1. Het netwerk van terminologiebronnen (“Knowledge Graph voor Digitaal Erfgoed”)

Dit is de verzameling van gelinkte terminologiebronnen die als referentie dient voor het maken van verwijzingen naar plaatsen, personen, onderwerpen, tijdperiode, gebeurtenissen in de beschrijvingen van de collecties en objecten. Het doel hiervan is tweeledig: zorgen voor een hoogwaardig netwerk van referenties én het zo breed mogelijk implementeren hiervan.

Uiteindelijk is dit een open netwerk waar iedereen onder bepaalde voorwaarden aan kan bijdragen. De basis van deze gemeenschappelijke voorziening wordt gevormd door de knooppunten en andere organisaties die gezaghebbende terminologiebronnen en andere referenties onderhouden en beschikbaar stellen (bijvoorbeeld ook WikiData, Getty, OCLC, etc) .

### 2. Het collectieregister

Dit is een gestandaardiseerd register waarin alle collecties zijn opgenomen inclusief relevante metadata over toegang tot de collectie, rechtenmodel, actualiteit etc. Hiervoor kan zoveel mogelijk de opzet gevolgd worden die ook voor andere Open Data registers wordt gebruikt (zie Datahub, Open Overheid, etc). Picturae heeft onlangs op basis van hetzelfde model (CKAN) een eerste stap in deze richting gedaan met <https://opendata.picturae.com/>. In het NDE-werkpakket Bruikbaar zal dit gemeenschappelijke collectieregister opgebouwd worden, inclusief de noodzakelijke financiële en technische afspraken over duurzaam beheer op lange termijn.

Voor het uniek identificeren van instellingen en de collecties die ze beheren zijn er binnen de verschillende domeinen al diverse coderingsafspraken gemaakt (o.a. ISIL voor bibliotheken, museum codering voor collectieregistratie en bruikleenverkeer). Het collectieregister zal hier zoveel mogelijk op aansluiten. Verder wordt er aansluiting gezocht bij het dataregister dat momenteel opgebouwd wordt in het kader van het CLARIAH project.

### 3. Het objectregister

Het objectregister is de motor voor zichtbaarheid. Hierin staan de relaties tussen de termen uit de knowledge graph en de objecten bij de instellingen. Voor het inrichten van een specifieke dienst (bijvoorbeeld “Oorlogsbronnen” of “de Zuiderzee Collectie”) wordt het register bevraagd op relevante matches met de door de dienst geformuleerde criteria (specifieke plaatsen, periodes, verzameling termen).

Het objectregister levert een globale afbeelding van de erfgoed informatie die beschreven is in de collecties in het netwerk. Afhankelijk van de mogelijkheden die de registratiesystemen ondersteunen voor het leveren van de data (OAI of LD) wordt er een specifiek aggregatie- en indexeringsproces voor de dienst ingericht. Het objectregister kan ook gezien worden als een uitbreiding op het collectieregister.

Op termijn kan zal dit register zich ontwikkelen naar meer linked data gebaseerde platformen die informatieverzamelingen via ‘federated queries’ op basis van bijvoorbeeld Linked Data Fragments. Hoewel van beide registers maar één instantie getekend is, zal een federatieve opbouw (bijv. per knooppunt) zeker tot de mogelijkheden behoren.

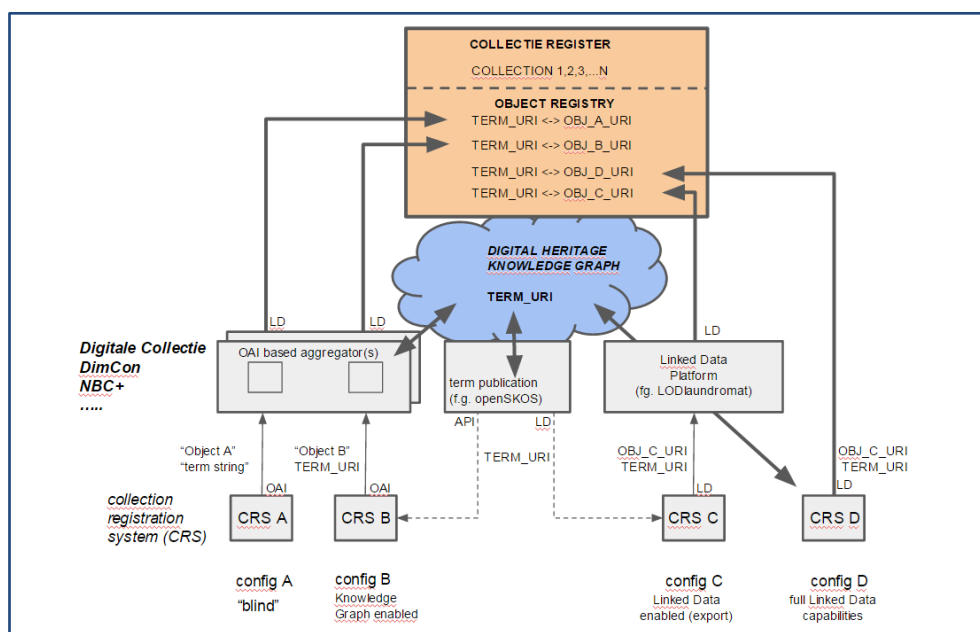
Ten opzicht van de huidige situatie ziet het NDE met deze aanpak de volgende verbeteringen:

- In de eerste plaats worden centrale referenties naar de bron gebracht waardoor matching bij de bron (en door de bronbeheerder!) gebeurt. Hiermee is de brondata direct van de correcte centrale URI voor een term voorzien. Dit levert een kwaliteitsverbetering op ten opzichte van matching of verrijking in tussen gelegen platformen en buiten het blikveld van de bronbeheerder zoals nu het geval is.
- In de tweede plaats gaat het nu primair om het consistent houden van het register in plaats van een volledige kopie van alle data. Dit is een belangrijke vereenvoudiging ten opzichte van de huidige situatie.
- De voorgestelde oplossingsrichting opent een groeiscenario naar een eindsituatie waar volledig op basis van Linked Open Data principes gewerkt wordt en waar het kopiëren van (delen van) data niet meer nodig is.

Het bieden van een groeimodel om geleidelijk van aggregeren over te gaan naar linken is een nieuwe opening in langlopende discussie.

## Een groeimodel met vier instapmomenten voor erfgoedinstellingen

Het schema hieronder toont het termenregister en het register voor collecties en objecten in samenhang met verschillende configuraties van de lokale collectie registratiesystemen. Het resultaat biedt uiteindelijk een uniforme ingang voor het selecteren en presenteren van de 'Collectie Nederland'.



Figuur 1: Termennetwerk en register voor collecties en objecten

Er bestaat een grote kloof tussen de huidige praktijk en de stip aan de horizon, waarin collectiebeheerders hun data beschikbaar stellen als linked open data, nauw verbonden met het termennetwerk en de registers van de centrale voorziening. Maar het is wel mogelijk om nu direct stappen te zetten die uiteindelijk naar dit eindplaatje leiden. Hieronder volgen de verschillende scenario's, uitgaande van de bestaande praktijk bij erfgoedinstellingen.

### Configuratie A - 'blind' systeem

In de eerste situatie (config A) is er sprake van een traditioneel collectieregistratiesysteem waarin geen gebruik gemaakt wordt van unieke identifiers voor objecten of gestandaardiseerde terminologie bronnen. Objecten worden gecodeerd met eigen identifiers in het registratiesysteem en de verwijzingen ('wie', 'wat', 'waar', 'wanneer'...) worden als tekst in velden geregistreerd, eventueel gebaseerd op een eigen lokale thesaurus of trefwoordenlijst.

De interoperabiliteit van het registratiesysteem bestaat ten hoogste uit een OAI-PMH interface waarmee de collectie metadata geharvest kan worden. Het harvesting platform dat de harvesting verzorgt (bijv. Digitale Collectie), leest de collectiegegevens periodiek in en probeert op basis van het matchen van de tekstvelden een verbinding te maken met de termen uit de terminologie bronnen.

Verder genereert het harvestingplatform een unieke (persistente) identifier (uri) voor het object op basis van de identificatie van de collectie en het object. Zo mogelijk wordt deze persistente URI's terug geleverd aan het collectieregistratiesysteem om het eventueel opnieuw genereren van een persistente identifier door een ander platform te voorkomen,

Als resultaat genereert het aggregatieplatform een linked data regel (triple) voor elke geslaagde match en schrijft dit weg in het objectregister (bijvoorbeeld: "TERM\_URI <-> OBJ\_A\_URI"). Voor de eigenschap die de relatie beschrijft, zou bijvoorbeeld edm:isRelatedTo gekozen worden. Voor het register is een dergelijke globale typering in eerste instantie voldoende omdat nadere specificiteit volgt na het opvragen bij het bronsysteem of het tussengelegen platform.

### Configuratie B - 'knowledge graph enabled systeem'

In de tweede configuratie (config B) is het collectieregistratiesysteem voorzien van een koppeling met het landelijke termennetwerk. Hoe de koppeling tussen een registratiesysteem en het termennetwerk exact vorm zal gaan krijgen moet nog verder uitgewerkt worden. Het doel is om de gestandaardiseerde termen laagdrempelig beschikbaar te maken voor de collectiebeheerder. Dit kunnen geografische begrippen zijn afkomstig van de historische geocoder, agents of concepten via OpenSKOS of verwijzing naar linked data resources die elders bestaan. In deze schets is het OpenSKOS platform als voorbeeld voor een dergelijke principe gebruikt.

De collectiebeheerder is nu in staat om tijdens het beschrijven termen te kiezen die een directe relatie met het termennetwerk hebben. In de metadata worden in dat geval niet alleen de tekstveld overgenomen, maar ook de URI voor de betreffende term. In situaties waar bestaande termen te kort schieten zou het zelf toevoegen van een term aan dit netwerk mogelijk moeten zijn. Afspraken hierover moeten verder uitgewerkt worden.

Net als in de eerste configuratie ondersteunt het registratiesysteem hooguit een OAI export. Via een aggregatieplatform wordt de metadata opgehaald. In dit geval is alleen het toekennen van een unieke identifier voor het object nodig. Er wordt vervolgens weer een regel linked data gegenereerd voor elke term die een gedefinieerde URI heeft en in het objectregister weggeschreven (bijvoorbeeld TERM\_URI isRelatedTo OBJ\_B\_URI).

Een variant voor dit scenario is het direct terugmelden aan het objectregister door het OpenSKOS platform. Zodra, via het OpenSKOS platform, een term wordt gekoppeld aan een object wordt het resultaat terug geschreven in het objectregister. Deze variant is mogelijk wanneer het object voorzien is van een geldige URI of wanneer het OpenSKOS platform het toekennen van object uri's kan faciliteren. De functionaliteit van het OpenSKOS platform zal in dit geval hiervoor verder aangepast moeten worden.

### Configuratie C - 'Linked Data enabled' systeem

In deze configuratie is het collectieregistratiesysteem niet alleen in staat om via een interface direct te koppelen met het termennetwerk maar voorziet het registratiesysteem ook in een exportmodule voor het genereren van linked data. In dit geval wordt door het collectieregistratiesysteem zelf de (persistente) URI voor het object aangemaakt. Het uitwerken van deze functionaliteit is de doelstelling van een gezamenlijk project van de NDE-werkpakketten Bruikbaar en Houdbaar.

Het collectieregistratiesysteem voorziet in deze situatie echter nog niet in een volwaardige linked data publicatie omgeving, de data is echter wel als download beschikbaar. Deze datadumps worden geladen in een linked data aggregatie omgeving (bijvoorbeeld een platform als de LODlaundromat). Dit platform destilleert de relevante regels uit de aangeboden linked data en schrijft deze weg in het register (bijvoorbeeld TERM\_URI isRelatedTo OBJ\_C\_URI). Het linked data platform ondersteunt verder ook het federatief doorzoeken van de verschillende linked datasets bijvoorbeeld op basis van Linked Data Fragments. Hiermee kunnen op eenvoudige manier eenvoudige linked data front-end applicaties ingericht worden.

#### **Configuratie D - 'Full Linked Data capabilities'**

Het lokale collectieregistratiesysteem is doorontwikkeld tot een volwaardige linked data beheer en publicatie omgeving dat gebruikt maakt van het termennetwerk en persistente URI's voor de objecten genereert. Het systeem is zelf in staat om rechtstreeks de links tussen de objecten en de termen in het register te schrijven (bijv. TERM\_URI isRelatedTo OBJ\_D\_URI). Daarbij is het ook mogelijk dat het systeem direct zoekvragen naar deze links of andere verbanden kan afhandelen en daarmee in feite een virtueel register met de andere registratiesystemen kan vormen.



## Opbouw en synchronisatie objectenregister

Het objectenregister kan gezien worden als een virtueel register dat mogelijk gedistribueerd over de verschillende domein wordt opgebouwd. Verder onderzoek is nodig om na te gaan in welke mate het gedistribueerd opbouwen zinvol is en of inzet van nieuwe technologieën voor het gedistribueerde web (o.a. blockchain, DHT, ipfs) interessant is.

Naast de relaties tussen objecten en termen kunnen ook relaties tussen objecten onderling relevant zijn. Ook dit kan in het register van gelegd worden (bijv. OBJ\_A\_URI isRelatedTo OBJ\_B\_URI). Op deze wijze kunnen de DERA-principes van “Verwijzen” en “Verwijsbaar zijn” ondersteund worden.

Het up-to-date houden van het objectenregister vanuit de verschillende configuraties is een belangrijk aandachtspunt. Het gaat hier om het synchroniseren van linked data stromen. Het verdient aanbeveling om na te gaan of het Resource Synchronization van Open Archives Initiatives<sup>1</sup> hiervoor een interessante optie is. In het kader van het CLARIAH project (Werkpakket 2 - Data en Infrastructuur) wordt momenteel voor een enigszins vergelijkbare situatie aan gewerkt.

Verder zal er ook nadrukkelijk aandacht besteed moeten worden aan het vasthouden de herkomst informatie (*provenance*) en het transparant houden van het update proces. Daarnaast is aandacht nodig voor het eventueel kunnen inrichten van rechtenmodellen en het beheer en onderhouden van het register. Hoewel deze onderwerpen hier als laatste aan bod komen verdienen ze een centrale plaats in het ontwerp van de centrale voorzieningen.

---

<sup>1</sup> <http://www.openarchives.org/rs/toc>