

Naar grotere zichtbaarheid en meer gebruik van de archieven

Linked open data bij het Nationaal Archief

Het Nationaal Archief¹ gelooft in de kracht van open data. Daarom wil het zoveel mogelijk data beschikbaar stellen. Niet alleen voor de overheid zelf, maar ook voor derden. Hierdoor kunnen erfgoedliefhebbers en andere geïnteresseerden op nieuwe manieren participeren en ontstaan er tal van mogelijkheden voor het gebruik van digitaal erfgoed.

TEKST ED DE HEER

Een van de nieuwe toepassingen is linked open data. Hierbij worden de data zodanig aangeboden dat er veel makkelijker, en zelfs automatisch, verbinding kan worden gelegd met andere informatie. Daardoor wordt de informatie rijker en nuttiger voor de gebruiker.

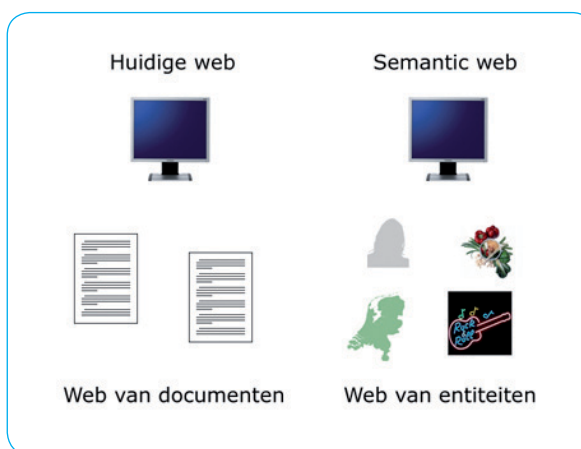
Dit artikel beschrijft de activiteiten die het Nationaal Archief op dit moment uitvoert en de ambities die de organisatie heeft op het gebied van linked open data.

Hieronder wordt eerst uitgelegd wat nu precies linked open data zijn en wat die voor relatie hebben met het zogenaamde semantisch web.²

Het semantisch web: een web van informatie waarin kleine brokjes kennis met elkaar verbonden zijn

Het semantisch web en linked open data

Sir Tim Berners-Lee, bedenker en grondlegger van het wereldwijde web (www), sprak als eerste over 'het semantisch web': een web van informatie waarin kleine brokjes kennis met elkaar verbonden zijn, de zogenaamde linked data. In zijn visie past het semantisch web in de ontwikkeling van het internet: van het web van documenten (Web 1.0) via Web 2.0, het internet als interactief en sociaal communicatiemedium, naar Web 3.0: een web van linked data met links waardoor aanvullende informatie kan worden



Figuur 1: Ontwikkeling van het web.

gepresenteerd. Het grote voordeel van het semantisch web t.o.v. eerdere versies is dat het bestaat uit machine leesbare informatie.

Het semantisch web bestaat uit kleine brokjes kennis die *triples* genoemd worden. Deze triples kunnen aan elkaar gekoppeld worden zodat er een web van kennis ontstaat: het semantisch web. Een triple bestaat uit een subject, een predicaat en een object.

Hieronder volgen drie voorbeelden.

Subject	Predicaat	Object
Eduard Douwes Dekker	had een aanstelling in	Nederlands-Indië
Eduard Douwes Dekker	is geboren op	2 maart 1820
Multatuli	is een pseudoniem van	Eduard Douwes Dekker

Tabel: Voorbeelden van triples.

Op basis van de term Eduard Douwes Dekker kan dan gelinkt worden met andere termen zoals Multatuli, en zo kan een netwerk van termen ontstaan. In het groot hebben we het dan over het semantisch web en specifiek voor de erfgoedsector over een *knowledge graph*. De termen die hierboven gebruikt worden, komen uit zogenaamde termenlijsten of vocabulaires. Op het semantisch web definiëren vocabulaires de concepten en relaties die gebruikt worden bij het beschrijven van een onderwerp of domein.

Netwerk Digitaal Erfgoed

Het Nationaal Archief is een van de vijf knooppunten, naast het Nederlands Instituut voor Beeld en Geluid, de Koninklijke Bibliotheek, de Rijksdienst voor het Cultureel Erfgoed en KNAW-DANS, binnen het Netwerk Digitaal Erfgoed (NDE)³. Op dit moment wordt er door het NDE hard gewerkt aan een gedistribueerd netwerk van erfgoed informatie. De essentie van dit gedistribueerd netwerk is de domein overstijgende knowledge graph van erfgoed informatie. Dit gedistribueerde netwerk moet ervoor zorgen dat data van alle erfgoedinstellingen, dus niet alleen van de vijf knooppunten, beschikbaar komen op het internet met behulp van linked open data.

De kenmerken van dit gedistribueerde netwerk van erfgoed informatie zijn:

- Gecontroleerde vocabulaires binnen de erfgoed domeinen worden met elkaar en met externe bronnen verbonden.
- Collectiebeheerders kunnen gebruikmaken van meerdere vocabulaires voor het beschrijven van hun collectieobjecten.
- Gebruikers zoeken en navigeren naar erfgoed informatie via concepten uit de knowledge graph (domeinoverstijgend, uit verschillende datasets/collecties).
- Gebruikers krijgen via concepten uit de knowledge graph suggesties naar erfgoed informatie.

Het Nationaal Archief en linked open data

In 2016 is het Nationaal Archief begonnen met een onderzoek naar de mogelijkheden van deze techniek. In eerste instantie zijn de zogenaamde 'nadere toegangen (NT) omgezet naar linked open data (LOD). Nadere toegangen zijn indexen⁴ op archieven. Deze Proof of Concept-fase (PoC) was positief en het Nationaal Archief en de Regionale Historische Centra (RHC's) besloten linked open data verder in te zetten om instelling- en domeinoverstijgend digitaal erfgoed te ontsluiten. Hiermee sluit het Nationaal Archief ook operationeel goed aan bij het Netwerk Digitaal Erfgoed. Na deze experimentele fase worden nu de volgende

acties opgezet om LOD binnen de archiefsector verder te professionaliseren: linked open data in productie, RHC's aansluiten op de LOD-store, verdere participatie NDE, het opzetten van goede communicatie en het maken van showcases. Hieronder volgt een nadere toelichting op deze onderwerpen.

Linked Open Data in productie

De PoC is zeker een goede stap geweest, maar het blijft nog maar een PoC. Het is nu zaak dat het Nationaal Archief LOD ook in de productieomgeving gaat zetten en dat het de mogelijkheid om LOD te maken goed gaat ondersteunen met de juiste tool(s).

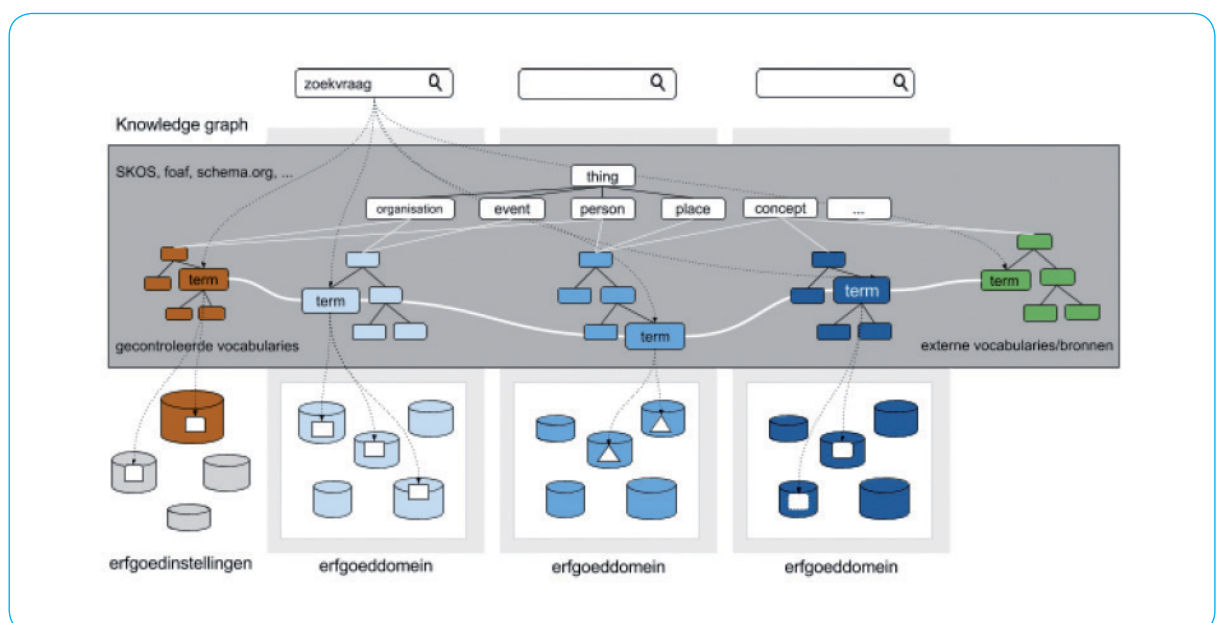
Een tweede belangrijke stap om LOD intern op orde te brengen is het integreren van LOD in de website. Dit is ook een nieuwe stap, omdat LOD zeer verschillende resultaten tot gevolg kunnen hebben. Dit vereist een zeer flexibele website.

Om LOD te kunnen maken is een softwarepakket aangeschaft, genaamd HUB3. Met deze software kunnen datasets op een redelijk eenvoudige manier omgezet worden naar linked open data.

Linked open data-voorziening voor de RHC's

Het programma Digitale Taken Rijksarchieven treft voorzieningen, die nodig zijn om de kerntaak van het Nationaal Archief en de RHC's die rijksarchieven beheren, ook in het digitale tijdperk te kunnen uitvoeren. Het Nationaal Archief en de RHC's zorgen samen voor de uitvoering van dit programma. Rijksinformatie die via de rijksinfrastructuur wordt aangeboden, moet volgens de principes van open data worden aangeboden. Het Nationaal Archief en de Regionale Historische Centra gaan op het gebied van belangrijke thema's zoals (linked) open data in de looptijd van DTR gezamenlijk beleid ontwikkelen.

Het Nationaal Archief gaat een voorziening opzetten waarmee de RHC's LOD kunnen maken en deze datasets kunnen publiceren in bijv. het Netwerk Digitaal Erfgoed of op hun eigen website.



Figuur 2: Het gedistribueerd netwerk van erfgoed informatie.

Voortzetten participatie NDE

Het Nationaal Archief kent binnen het NDE enkele rollen:

- *Participant*: Medewerkers van het Nationaal Archief participeren in diverse werkgroepen van het NDE. Het Nationaal Archief is o.a. betrokken bij de ontwikkeling van het Gedistribueerd Netwerk.
- *Leverancier van data*: De datasets die nu omgezet worden naar LOD kunnen aangeleverd worden aan het gedistribueerd netwerk. Zo kunnen andere instellingen deze data gebruiken om hun eigen informatie te verrijken.
- *Gebruiker van data*: Uiteindelijk wordt het Nationaal Archief ook gebruiker van het gedistribueerd netwerk. Het Nationaal Archief kan dan zijn eigen data verrijken met data van andere bronnen.

Communicatie

De PoC leidde tot een aantal aanbevelingen voor het gebruik van bepaalde vocabulaires. Zo zijn voor de wie-, wat-, waar- en wanneer-vragen vocabulaires voorgesteld die als het meest gangbaar en dekkend worden gezien voor de archiefsector. In navolging van deze PoC wordt dan ook een rondgang langs de RHC's gedaan om het gebruik van linked open data en het gebruik van vocabulaires toe te lichten en medewerkers op te leiden.

Showcases en Open Data Events

- Het Actorenregister⁵ bevat institutionele informatie van de overheid vanaf 1975. Hierin worden naamswijzigingen en taakveranderingen van de overheid bijgehouden. Het zou interessant zijn om informatie uit deze databron te koppelen aan andere bronnen zoals Parliament.com of DBpedia⁶ zodat het Actorenregister verrijkt wordt.
- Linked open data bieden grote mogelijkheden voor derden om de data te gebruiken voor apps en eigen toepassingen. Hierdoor krijgt het digitaal erfgoed meerwaarde voor de gebruikers en worden nieuwe gebruikers geïnteresseerd. Zo heeft het Nationaal Archief bijvoorbeeld meegedaan aan de Hackalod⁷ die op 9 en 10 februari door het NDE is georganiseerd.

Het Nationaal Archief wil de nieuwe techniek en werkwijze inbedden in de organisatie

Het Nationaal Archief heeft tijdens deze Hackalod de zogenaamde *Verkaufsbücher* (dit is een registratie van alle gedwongen verkopen van Joods onroerend goed tijdens de Tweede Wereldoorlog) gekoppeld aan data van het Joods Monument, een registratie van alle Nederlandse Joodse slachtoffers met woonplaats en plaats van overlijden. Het Nationaal Archief realiseerde een deel van zijn doelen en won de derde prijs.

Doorkijkje naar 2020

De PoC was een eerste stap met LOD. Hierna is het noodzakelijk om deze nieuwe techniek en werkwijze in te bed-



Foto: Denis Makheev | Shutterstock.com

den in de organisatie. Daar werkt het Nationaal Archief de komende twee jaar aan.

Waar staat het Nationaal Archief in 2020 met LOD? Gebaseerd op alle bovenstaande ontwikkelingen maakt het Nationaal Archief dan volop gebruik van linked open data. De LOD-expertise is geprofessionaliseerd en de LOD-store werkt optimaal. De RHC's zijn aangesloten en veel data van het Nationaal Archief worden gecombineerd met data van andere erfgoedinstellingen en/of het gedistribueerd netwerk van het Netwerk Digitaal Erfgoed. Dit alles zal leiden tot grotere zichtbaarheid en meer gebruik van het enorme potentieel van archieven. ●

Ed de Heer

Ed.deHeer@nationaalarchief.nl

Drs. Ed de Heer MIM is bij het Nationaal Archief informatie-analist en projectleider op het terrein van Linked Open Data.

Noten

- 1 <https://www.nationaalarchief.nl>
- 2 Meer informatie over het hoe en waarom van linked open data is ook te vinden op www.pilod.nl.
- 3 <http://www.netwerkdigitaal erfgoed.nl/>
- 4 <http://www.gahetna.nl/collectie/index>
- 5 <https://actorenregister.nationaalarchief.nl/welkom>
- 6 DBpedia is de Linked Open Data versie van Wikipedia <http://wiki.dbpedia.org/>
- 7 <http://hackalod.com/>

Zie ook www.od-online.nl