

Archiefontologie Nadere Toegangen NA (Linked Open Data)

Versie 1.0

Datum 28 november 2016
Status Definitief

Projectnaam	Archiefontologie Nadere Toegangen NA
Contactpersoon	Gijsbert Kruithof T +31 6 55 26 79 17 gijsbert.kruithof@nationaalarchief.nl Postbus 90520 2509 LM Den Haag
Auteurs	Gijsbert Kruithof
Versie	1.0 (definitief na goedkeuring stuurgroep Open Toegang)
Bijlage(n)	

Project

Archiefontologie Nadere Toegangen NA

Doel

Komen tot een concreet en voor de archiefsector bruikbaar implementatievoorstel om archiefmateriaal via Linked Open Data toegankelijk te maken.

Achtergrond

Een van de hoofddoelen van Archief 2020 is Open Toegang, dit doel blijft ook gehandhaafd in de nieuwe meerjarenvisie onder het speerpunt Toegang. Een van de belangrijkste middelen voor Toegang is het gebruik van open standaarden. Linked Open Data biedt bij uitstek de mogelijkheid om data beschikbaar te stellen op een open en gestandaardiseerde manier waardoor hergebruik en verbinden met andere (erfgoed)bronnen heel eenvoudig wordt.

Om Linked Open Data te gebruiken moet een keus gemaakt worden uit vocabulaires en ontologieën (terminologiebronnen) om de metadata mee te definiëren en technisch te implementeren. Er is al veel inventarisatiewerk verricht (o.a. in het kader van de NDE), maar er zijn nog nauwelijks concrete implementaties binnen de archiefsector. Er moeten bijvoorbeeld keuzes gemaakt worden welk schema je kiest om persoon of een gebeurtenis te coderen of een verwijzing naar een archiefbron, en hoe je dat precies doet. Deze keuzes wil je het liefst zoveel mogelijk afstemmen met andere archieven en erfgoedinstellingen, zodat "het wiel maar één keer hoeft worden uitgevonden".

Nationaal Archief en RHC's zijn van plan om Linked Open Data systematisch in te gaan zetten in de nieuwe infrastructuur. Daarvoor is het nodig om te weten welk model je hanteert bij implementatie en welke vocabulaires en terminologieën gebruikt kunnen worden. De Linked Open Data expertise die daarvoor nodig is, is nog niet aanwezig binnen het Nationaal Archief. Het voorstel is om een Linked Open Data architect (Gerard Kuys, Ordina) in te huren die een Proof of Concept kan uitvoeren. Op basis daarvan kan hij een concreet voorstel maken voor het ontwikkelen van een model voor het toepassen van Linked Open Data, dat in de hele archiefsector bruikbaar is. Dit is de start voor een Linked Open Data ontologie voor de archiefsector. Het project zal richtlijnen opleveren en een praktisch Kookboek voor archiefinstellingen om van archiefdata Linked Open Data te maken. De Linked Open Data ontologie en het LOD Kookboek zijn te beschouwen als een kennisproduct in de lijn van andere kennisproducten die vanuit Archief 2020 zijn opgeleverd aan het Nationaal Archief. Zo bouwt het voort op de Handreiking Linked Data Archieven van november 2016 (Archief 2020).

Met het projectresultaat weten Nationaal Archief en RHC's in 2017 hoe zij hun nadere toegangen kunnen gaan coderen en ontsluiten als Linked Open Data.

Resultaten

Het daadwerkelijk inzetten van Linked Open Data in de archiefsector is nieuw en te beschouwen als een ontdekkingstocht waarbij we van te voren niet precies weten waar we tegenaan gaan lopen. Onderstaande lijst van resultaten is gebaseerd op de huidige inzichten, maar het is nodig om binnen het project enige speelruimte te houden om de werkzaamheden en de resultaten aan te passen.

De nu geplande resultaten zijn:

1. Overkoepelend model om archiefdata om te zetten naar Linked Open Data.
2. Eerste Proof of Concept met het coderen van drie Nadere Toegangen als Linked Open Data, inclusief het linken van de archiefdata aan externe bronnen: november 2016.
3. Proof of Concept met een representatieve set aan archiefdata over het hele spectrum van personen, onderwerpen, gebeurtenissen, plaatsen enz. in februari 2016. In de POC zullen de archiefdata gecodeerd zijn als Linked Open Data, maar er zal ook een poging gedaan

- worden de data te linken aan andere bronnen. Met de POC kan de meerwaarde van Linked Open Data zichtbaar gemaakt worden.
4. Deelnemen aan de uitwerking van het Event model voor het modelleren van gebeurtenissen in Linked Open Data.
 5. Rapportage met verantwoording van gemaakte keuzes, uitdagingen, problemen en issues.
 6. Richtlijnen voor keuze en gebruik van terminologiebronnen en vocabulaires voor het nader toegankelijk maken van archiefmateriaal.
 7. Linked Open Data Kookboek voor de archiefsector: praktische handleiding hoe je archiefdata kan vertalen naar Linked Open Data: Welke entiteiten zijn er: wie, wat, waar, wanneer enz. en hoe codeer je ze als LOD.

Relatie met andere projecten

- Het project sluit aan op één van de NDE-projecten die is gedefinieerd in het NDE werkpakket Bruikbaar, namelijk het Project om het gebruik van terminologiebronnen af te stemmen met de rest van het erfgoedveld. Als projectleider is hiervoor aangetrokken Ed de Heer. Hij zal de voorstellen zoals die worden opgesteld in dit project afstemmen met de andere erfgoed domeinen. In het project van Ed de Heer zal een klankbordgroep worden ingericht waarin het werk van Gerard kan worden besproken en zondig bijgestuurd. Onder het project van Ed valt ook het communiceerbaar maken van de producten die Gerard oplevert en de daadwerkelijke communicatie daarvan. Doordat het project van Ed onderdeel is van het werkpakket Bruikbaar zal ook de aansluiting met het Erfgoedveld afdoende zijn gewaarborgd.
- De Vrije Universiteit onder leiding van professor Lora Aroyo start een project voor het opstellen van een Linked Open Data model voor gebeurtenissen. Er is tot nog toe weinig aandacht besteed aan gebeurtenissen in vergelijking met de entiteiten Persoon, Plaats en Tijd. Gebeurtenissen zijn echter voor historici en het archiefwezen bij uitstek van belang. Gerard zal vanuit zijn expertise kunnen bijdragen aan dit project en de projectresultaten van de VU kunnen onderdeel worden van een overkoepelend Linked Open Data model voor de archiefsector.
- Het project maakt aansluitend gebruik van de 'Handreiking Linked Open Data Archieven', die in november 2016 wordt gepubliceerd door Bob Coret, projectleider Archief 2020. In deze Handreiking is opgenomen het Stappenplan LOD, een product van de pilot Vierdaagse Nijmegen (Archief 2020) voor de implementatie van Linked Open Data in de organisatie.

Mijlpalen

- 11-12 nov. 2016:** Hack-a-Lod: De Nadere toegangen uit de Proof of Concept zijn beschikbaar als Linked Open Data en kunnen gebruikt worden in toepassingen tijdens de Hack-a-Lod.
- November 2016:** Presentatie van voorstel voor overkoepelend model voor Linked Open Data van archieven.
Bijdrage aan de open 'Kennissessie Nadere Toegangen Linked Data' op 23 november bij het Nationaal Archief.
- December 2017:** Bespreking van het voorstel van een overkoepelend Linked Open Data model voor archieven, met diverse gremia zoals de architectuurcommissie van NA & RHC's, DERA-architecten enz.
- Januari 2017:** Concept gebeurtenismodel
- Jan./febr. 2017:** Nadere Toegangen uit de Proof of Concept beschikbaar via het Nationaal Archief
- Febr./mrt 2017:** Linked Open Data Kookboek
- Maart 2017:** Rapportage van keuzes, uitdagingen, problemen en issues.

Doorlooptijd

In het project wordt een overkoepelend model "ontworpen" en er wordt een representatieve set van Nadere Toegangen in Linked Open Data gecodeerd. Dit is een complexe taak, die een forse investering vergt qua tijd en doorlooptijd. De complexiteit zit op drie niveaus:

- De diversiteit van de nadere toegangen is enorm en het aantal mogelijke manieren om het als Linked Open te modelleren zijn legio. Daardoor is het analytisch zeer complex.
- Voor een deel is het terra incognita. Sommige concepten in Nadere Toegangen zoals bijvoorbeeld filmkeuringen of visrechten zijn nog nooit gemodelleerd en gecodeerd als Linked Open Data en daar zijn ook geen modellen voor, die moeten dan deels nog ontworpen worden.
- Het modelleren moet afgestemd worden met diverse experts en partijen om te zorgen dat er een goed doortimmerd en gedragen voorstel ligt.

Gerard Kuys is een Linked Open Data expert van het eerste uur met een groot internationaal netwerk. Daardoor is hij bij uitstek geschikt voor deze complexe taak.

Gerard is begonnen in oktober en zal de laatste werkzaamheden afronden in maart 2017. Het project van Ed de Heer dat zorgdraagt voor de afstemming met het veld (zie Kennisborging) loopt nog verder door.

Kennisborging

De kennisfunctie van het NA is betrokken bij het verzamelen van specificaties, het borgen en het verspreiden van resultaten. Via het project van Ed de Heer bij het Nationaal Archief is er een relatie met het NDE. Resultaten worden gedeeld op het 'Kennisplatform toegang' in Pleio (informatie2020.pleio.nl). Daarnaast vindt kennisdeling plaats met het archief- en erfgoedveld, wetenschap, platforms en programma's in kennissessies en – bijeenkomsten. Op deze wijze vindt brede dialoog en afstemming plaats over de voorgestelde aanpak en opgeleverde producten.

Financiën

Inhuur van de Linked Open Data expert Gerard Kuys (Ordina) voor de duur van een half jaar voor 2 dagen in de week. Indicatie kosten: 26 weken, 16 uur per week à 83,92 euro per uur (excl. BTW) = 42.241,97 euro (incl. BTW).

Communicatie

Er komt een communicatieplan zodat de archiefsector weet wat er is opgeleverd, waarom, wat ze er aan hebben en wie er mee verder gaat. De communicatieboodschap wordt bepaald met de samenwerkingspartners.

Samenwerkingspartners

In het project wordt samengewerkt met diverse experts en (archief)instellingen die voorop lopen in het gebruik van Linked Open Data.

Dit zijn onder meer:

- De Vrije Universiteit (Jacco van Ossenbruggen, Lora Aroyo, Victor de Boer)
- Koninklijke Bibliotheek (Enno Meijers)
- Het NIOD (Edwin Klijn, Annelies van Nispen, Lizzy Jongma)
- Tresoar/Erfgoed Leiden (Hans Laagland)
- IISG (Richard Zijdeman)
- Huygens Instituut (Ronald Sluifster)
- Regionaal Archief Nijmegen (Ernest Verhees)
- Kadaster
- KOOP (Hayo Schreijer)
- De DBPedia gemeenschap
- APE / Europeana