

CONCEPT

Directie
RNE

Afdeling

Gegevensmanagement

Contactpersoon

*Friso Snijder
Gerald Groot Roessink
Marloes Dijkstra*

Datum

03-03-2020

Bijlage

Pitch architectuurscan v2

Verzoek ROSA scan

Aanleiding

RIO is een ketenvoorziening waarbij gegevens worden ingewonnen op een wijze die past bij verschillende contexten. Deze gegevens en metadata worden ontsloten binnen de keten, maar worden ook als open onderwijsdata beschikbaar gesteld aan verschillende doelgroepen. Eerder is een ROSA scan uitgevoerd op het register RIO. Vanuit Architectuurraad is geadviseerd om een vergelijkbare ROSA scan uit te voeren op de generieke ontsluiting voor het publiek.

Afspraak

Het gaat om de ketenvoorziening RIO. Instellingen vullen gegevens in verschillende contexten en hergebruik vindt plaats/zal gaan plaatsvinden door o.a. Inspectie van het Onderwijs, Kennisnet, SBB (Kies MBO), SAMBO ICT (Topicus), CINOP (actueel), maar ook Studielink, CITO, Studiekeuze123 zijn potentiële afnemers en zo zullen er nog velen zijn.

RIO wordt ontsloten als Linked Open Data (momenteel MBO en VAVO formeel/juridisch en onderwijskundig/verrijkt) (4* & 5* Tim Berners Lee¹). Het staat nog niet hard gepland, maar zodra RIO zover is zal er ook via onderwijsdata.duo.nl een publicatie gaan plaatsvinden (1* 2* & 3* Tim Berners Lee).

De Linked Open Data is een samenwerking met Kadaster. Instellingen vullen gegevens in tbv RIO. Het Gegevensmagazijn transformeert de gegevens naar Linked Open Data en voegt in deze transformatie metadata toe vanuit het Gegevenswoordenboek/Modelrepository. De data en metadata worden geleverd aan Kadaster en Kadaster stelt deze beschikbaar via een website, een moderne api en straks zal het ook mogelijk zijn voor gebruikers om een sparql query te gebruiken

De Linked Open Data Api is hoog beschikbaar en is niet bedoeld voor synchronisatieprocessen (d.w.z. dat de afnemer data kopieert naar eigen database en deze gebruikt voor andere processen).

De Open Onderwijsdata variant is een generieke voorziening voor OCW waar bijvoorbeeld Inspectie van het Onderwijs en RCE gebruik van kunnen maken voor het beschikbaar stellen van open onderwijs data. Hier wordt gebruik gemaakt van open source CKAN en hosting verloopt via ODCN. Open onderwijsdata wordt

¹ <https://www.open-overheid.nl/vraagbaak-open-data-5-sterren/>

beschikbaar gesteld via een moderne api en ook hier wordt het gegevenswoordenboek gecombineerd met de data. De gegevens over de publicatie worden ook automatisch doorgegeven aan data.overheid.nl.

Overige bijzonderheden qua RIO zijn:

- De RIO modellen worden bij Edustandaard 'afgetikt'
- de metadata van RIO zal ook als niet wettelijk register worden opgenomen in de stelselcatalogus (staat op de backlog bij stelselvoorzieningen KOOP).
- conform REST API Design Rules (Forum Standaardisatie)
- Verzamelde standaarden op gebied van Linked Data (Forum Standaardisatie)
- De RIO-api is een voorziening voor programmeurs. Eindgebruikers maken kennis met RIO via hun producten.

Verzoek en doel

Dit verzoek wordt aangevraagd door Friso Snijder. Ik ben werkzaam voor de Directie RNE en het organisatieonderdeel Service Eenheid Gegevens. Daarbinnen ben ik Productowner van het Gegevensmagazijn. Het Gegevensmagazijn houdt zich hoofdzakelijk bezig met het optimaal ontsluiten van gegevens ten behoeve van interne DUO afnemers voor hergebruik in bedrijfsprocessen (Beslissen, Mededelen, Effectueren).

Dit team verzorgt ook de publicatie van open onderwijsdata en onderhoudt de techniek hierachter en dit team doet ook de transformatie naar Linked Open Data.

Het doel van de ROSA scan is het aanvragen van een kwalitatieve scan op uitgangspunten, robuustheid, huidige implementatie en voorgenomen activiteiten zodat dit meegewogen kan worden in de verdere ontwikkeling. Zeg maar een soort van onafhankelijke second opinion met ROSA als kader (ketenperspectief).

Linked Data is al geruime tijd een principe dat in de ROSA wordt aangehouden. Met RIO hebben we dit principe full throttle geoperationaliseerd. Ik zou graag willen weten, indien het bevalt, hoe verder binnen de onderwijsketen? Zijn er andere toepassingen voor Linked Data?

Doel van de afspraak

Wat ik graag wil bereiken met de RIO ontsluiting is een optimale ontsluiting voor externe afnemers. Vanuit DUO perspectief lijkt het nu zeer wenselijk om dit op 1 manier aan te bieden voor iedereen met een ontsluiting zo dicht mogelijk bij de bron. DUO heeft hier gekozen voor een Linked Open Data variant (gegevensbeleid) die in mijn ogen zeer krachtig is.

Echter om dit in een keten te laten functioneren is een het van belang om de stakeholders te kennen, te weten welke producten de stakeholders hebben en welke klanten hierbij horen. Ook is het belangrijk om 'de staat van de ict' te kennen. Het is uiteraard ook belangrijk om te weten of deze stakeholders uit de voeten kunnen met wat we aanbieden nu en in de toekomst beredeneerd vanuit de totale gegevensbehoefte. Deze inzichten geven vorm aan een verdere invulling van de roadmap. Zover ben ik nu nog niet.

Op te lossen problemen

-DUO wil graag rolzuiver beheer. D.w.z dat het beheer daar plaats vindt waar het hoort (juist, tijdig, volledig en voorzien van metadata daar waar het beheer plaats vindt).

-Gegevensmanagement wil lean ontsluiten. Dus geen ellenlange ketenprocessen met verrijkingen en redundant beheer verderop in de keten.

-DUO wil graag betekenisvol ontsluiten. Het moet duidelijk zijn welke betekenis een gegeven heeft en in welke context deze gebruikt/hergebruikt kan worden. Afgeleid wil DUO graag dat betekenis daar gedefinieerd wordt waar beheer plaats vindt en dat deze betekenis ook wordt hergebruikt. Dus DUO typt geen BAG betekenissen over en dat doen we nu nog wel.

-DUO wil graag op een efficiënte manier ontsluiten (en uiteraard ook inwinnen)

-DUO wil optimaal voor de keten en BV Nederland zodat belanghebbenden goed geïnformeerd beslissingen kunnen nemen. Dienstaanbieders (softwareleveranciers, app bouwers en dergelijke) hebben hier een belangrijke rol in het creëren van toegevoegde waarde tussen gegevens en stakeholders.

-Linked data lijkt een methode waarmee op betrekkelijk eenvoudige wijze meerdere dataset aan elkaar gekoppeld kunnen worden, en waarmee een kennissysteem gevoed kan worden. Zo is het bijvoorbeeld vrij eenvoudig om over overheidsorganisaties heen te zien wat een organisatie weet van een bepaald onderwerp. Dit kan waardevolle informatie geven voor onderzoekers, beleidsmakers en wetenschappers (in een iets bredere context dan RIO)

Naast problemen oplossen of nieuwe toegevoegde waarde creëren introduceren nieuwe dingen uiteraard ook risico's. Als gegevens makkelijker gekoppeld kunnen worden en er allerlei hoog beschikbare api's ontstaan dan neemt de afhankelijkheid toe en heeft eventueel misbruik grotere gevolgen (fake nieuws/uitval). Ik ben met name geïnteresseerd in de voor nu juiste balans/ mitigerende maatregelen die redelijk/normaal zijn op dit gebied.

Er zijn nog overigens nog weinig partijen die hier actief om vragen. Het is voor velen een kijk op de werkelijkheid die net even anders is en waar nog niet altijd de toegevoegde waarde van wordt herkend/ waar nog een vertaalslag gemaakt moet worden naar wat dit betekent voor de processen bij de afnemende partij.

Aanleiding van de afspraak

RIO waarom:

<https://www.rio-onderwijs.nl/waarom-rio>

Linked data waarom:

https://www.pldn.nl/wiki/Conferentie:_Onderwijs_en_Linked_Data_-_wat,_waarom_en_hoe

Open (onderwijs) data waarom:

Hier zijn meerdere perspectieven op los te laten die allemaal waardevol zijn. Vanuit DUO willen we zo ruw mogelijk, meer en moderner aanbieden zodat dienstaanbieders hiervan producten kunnen maken die weer leiden tot beter geïnformeerde stakeholders (burger/beleidsmakers/wetenschap).

RIO waarom nu: Heel plat beantwoord heeft ICT een lifecycle. BRIN zoals het was is 'economisch afgeschreven'. BRIN is een ketenvoorziening. Als een ketenvoorziening vervangen wordt dan ga je in gesprek met stakeholders, neem je de meest recente inzichten mee en als het groots is dan maken we er meestal een programma van.

Linked data waarom nu:

De tijd is ongeveer rijp hiervoor en de vernieuwing van RIO past heel goed bij linked data dus het is een 'logisch' moment.

Open (onderwijs) data waarom nu:

De afdeling informatie producten Den Haag doet al geruime tijd aan open onderwijs data. Vanuit o.a. OSF maar ook andere partijen wordt al jaren gevraagd om meer en moderner aanbieden. OCW heeft dit geluid ook gehoord en in 2017 of 2018 opdracht verstrekt tot een stimulans op dit gebied. Vervolgopdrachten zijn overigens uitgebleven tot nu toe.

Waar staan we en wat is het vervolg

RIO wordt momenteel gevuld met een onderwijskundige werkelijkheid. RIO wordt per sector opgepakt. MBO/VO/PO/HO. Het gaat hier met name om het vullen van de onderwijskundige/verrijkte werkelijkheid.

Vanuit de optiek van ontsluiten wordt BRIN nu gepubliceerd op onderwijsdata.duo.nl voor alle sectoren.

Er is een linked open data ontsluiting gemaakt waarbij Kadaster nastreeft dat deze op 3-3-2020 live gaat. Deze ontsluiting betreft formeel/juridisch en onderwijskundig/verrijkt RIO (niet ROD).

Er is een tijdelijke workaroud gemaakt voor SAMBOICT/Topicus via onderwijsdata.duo.nl

Vervolgstappen zijn:

- Linked Open Data ontsluiting echt afronden voor MBO/VAVO en de resultaten breed delen.
- ROSA scan op dit stuk
- Stakeholders en hun producten en verwachtingen leren kennen
- De tijdelijke workaroud dagelijks geautomatiseerd actualiseren voor Topicus
- Gegevenswoordenboek versiebeheer loskoppelen van uri's

Wel in de planning maar nog niet geactiveerd

- Activeren VO (en mogelijk PO) deel via Linked Open Data
- BRIN open onderwijsdata vervangen voor RIO open onderwijsdata

Wensenlijst

- Realtime linked open data (SAMBO ICT)
- LOD API generieker maken

Risico's

Grotere onderlinge afhankelijkheid

'Fake nieuws'

Datakwaliteit past niet bij de verwachtingen van de gebruikers want niet duidelijk belegd/onvoldoende know how/niet te implementeren etc.