

# RONDE TAFEL

## *Linked Data*

# Colofon

**Ronde Tafel**  
*Linked Data*

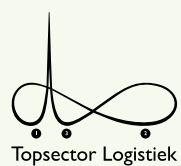
**Auteurs:**

Drs. L. J. Lapré, CGI Nederland B.V.

Ir. A. R. Toering, CGI Nederland B.V.

Geschreven in opdracht van Topsector Logistiek

Juni 2021



# Inhoudsopgave

<b>Inleiding</b>	<b>4</b>
De noodzaak tot datadeling	4
Problemen met datadeling	4
Linked Data als mogelijke oplossing	4
<b>Het onderzoek</b>	<b>6</b>
1. Beïnvloeding via Linked Data op Knooppunten	6
2. Veranderingen aangezet door Linked Data	7
3. Geleerde lessen bij het inzetten van Linked Data	7
4. Centrale aansturing bij Linked Data	8
5. FAIR en Linked Data	8
6. AI en Linked Data	9
<b>De sessie</b>	<b>10</b>
Problemen bij Datadeling	10
Presentaties	10
Discussie	11
<b>Conclusie</b>	<b>12</b>
<b>Expert interviews</b>	<b>13</b>
1. Wouter Beek	13
2. Jan Voskuil	17
3. Erwin Folmer	21

# Inleiding

Binnen de Topsector Logistiek is de vraag ontstaan of de Linked Data technologie meerwaarde kan bieden voor de transportsector. Hiervoor is er binnen Connekt een onderzoek uitgevoerd naar de ervaringen met Linked Data binnen verschillende organisaties. Deze resultaten zijn gedeeld met een aantal partijen in een sessie. Het onderzoek en de resultaten uit de sessie zijn in dit boekje weergegeven.

## De noodzaak tot datadeling

Veel organisaties komen nu tot de ontdekking dat het goed op orde hebben van data, voor de eigen organisatie, veel werk is en continue aandacht vraagt. Daarnaast wordt het ook steeds duidelijker dat een focus op de eigen organisatie niet voldoende is, aangezien er steeds meer data tussen organisaties uitgewisseld gaat worden.

Naar mate de problemen complexer worden, moet data van steeds meer organisaties met elkaar gedeeld gaan worden om tot goede oplossingen te komen. We zien deze trend terug in de vele dataknooppunten die nu in gebruik zijn, zowel binnen het bedrijfsleven als ook binnen de overheden. Je ziet het met name bij dataknooppunten waar publieke en private partijen samenkomen.

De grote opgaven waar wij nu voor staan, zoals de klimaatproblematiek, de energietransitie, het vervoer of de woningbouwopgave, zijn niet meer vanuit één enkel perspectief te benaderen. En juist de hoeveelheid benodigde invalshoeken leidt tot de noodzaak van het delen van data tussen organisaties. Hierbij komt het vaak voor dat de organisaties nauwelijks met elkaar bekend zijn.

## Problemen met datadeling

De data van een organisatie is vaak niet gereed voor uitwisseling. Dat komt omdat de data vaak gekoppeld is aan de noodzaak voor informatie bij een specifiek proces. En vanuit de behoeften van dat proces is deze data dan ook verzameld. We kunnen de data wel met elkaar delen maar dit leidt onmiddellijk tot een aantal problemen:

- Hoe moet de data geïnterpreteerd worden?
- Hoe accuraat is de data en kan ik de uitkomsten vertrouwen?
- Hoe is het beheer en de actualiteit geregeld?

Het is al moeilijk om deze vragen voor de data van de eigen organisatie te beantwoorden, laat staan voor de data van andere organisaties.

## Linked Data als mogelijke oplossing

Linked Data is een technologie en een manier van werken die juist ontstaan is om het delen van data eenvoudiger te maken. Het gebruikt semantiek zodat het duidelijk wordt waar de data precies over gaat en hierdoor wordt het eenvoudiger om data uit verschillende bronnen aan elkaar te 'Linken'. Dit staat ons toe om, gedeeltelijk geautomatiseerd, veel eenvoudiger data te koppelen en te verwerken.

Linked Data is een techniek waarbij data aan elkaar gekoppeld en geïnterpreteerd kan worden zonder dat de data verplaatst hoeft te worden. Dit ondersteunt het 'data bij de bron' principe waarbij data alleen nog bij de bron in de laatste versie aanwezig is en er niet meer op verschillende plekken kopieën ontstaan die allemaal net een andere waarheid beschrijven.

*Linked Data is structured data which is interlinked with other data so it becomes more useful through semantic queries. It builds upon standard Web technologies such as HTTP, RDF and URIs, but rather than using them to serve web pages only for human readers, it extends them to share information in a way that can be read automatically by computers. Part of the vision of linked data is for the Internet to become a global database*  
- Wikipedia

Er zijn meerdere sectoren waar Linked Data nu het uitgangspunt is om data met elkaar te gaan delen. Voorbeelden vinden we in Digitaal erfgoed<sup>1</sup> en het Beheer van de Openbare Ruimte<sup>2</sup>.

Veel goede voorbeelden van succesvolle inzet van Linked data zijn ook te vinden in de publicaties van het Platform Linked Data Nederland<sup>3</sup>.

<sup>1</sup> <https://netwerkdigitaalerfgoed.nl/activiteiten/linked-data/>

<sup>2</sup> [www.crow.nl/thema-s/management-openbare-ruimte/imbor/imbor-linkeddata](http://www.crow.nl/thema-s/management-openbare-ruimte/imbor/imbor-linkeddata)

<sup>3</sup> [www.pldn.nl/wiki/Publications](http://www.pldn.nl/wiki/Publications)


## Het onderzoek

In het onderzoek zijn een drietal experts bevestigd over hun ervaringen met Linked Data. Alle drie hebben ze ervaringen opgedaan in meerdere organisaties, artikelen gepubliceerd en lezingen gegeven. Ook zijn ze actief binnen de Linked Data gemeenschap in Nederland. De verzamelde input was de basis voor de sessie, aangevuld met een praktijk casus door één van de experts. De volgende drie experts zijn betrokken geweest:

- Wouter Beek, Triply<sup>4</sup>
- Jan Voskuil, Taxonic<sup>5</sup>
- Erwin Folmer, Kadaster<sup>6</sup>

Voor dit onderzoek zijn vier basisvragen gehanteerd voor de interviews. Tijdens de interviews kwamen nog twee aspecten regelmatig aan de orde; FAIR en AI. Deze zijn alsnog toegevoegd.

### 1. Beïnvloeding via Linked Data op Knooppunten

 *In welke mate beïnvloedt nieuwe technologie, zoals semantisch modelleren en Linked Data, de manier waarop straks **informatie-uitwisseling** georganiseerd zal zijn? Hoe vindt die beïnvloeding plaats?*

Dit waren de antwoorden op die vraag:


- LD stimuleert organisaties om zich meer bewust te worden van hoe hun data wordt gebruikt en richten zich daardoor meer op hun afnemers.
- Door LD verschuift de focus van het uitwisselen van data naar het delen van kennisdefinities en gezamenlijk komen tot standaarden.
- Door LD kan de data bij de bron blijven en laat daardoor meer decentrale en federatieve organisatievormen toe.
- LD maakt het afleiden van de bron van de data eenvoudiger waardoor herleidbaarheid beter geborgd is bij wijzigingen en herstel van fouten.
- LD kan beter data verbinden binnen ketens en tussen silo's dan andere technologieën.
- LD stelt ons in staat om niet alleen de data zelf, maar ook kennis van de data, opgedaan door medewerkers van de organisatie, te modelleren.
- LD stelt ons in staat om AI technieken toe te passen in netwerken van gekoppelde kennis en dan de bijbehorende data te raadplegen.

<sup>4</sup> [www.linkedin.com/in/wouterbeek](http://www.linkedin.com/in/wouterbeek)

<sup>5</sup> [www.linkedin.com/in/jan-voskuil](http://www.linkedin.com/in/jan-voskuil)

<sup>6</sup> [www.linkedin.com/in/erwinfolmer](http://www.linkedin.com/in/erwinfolmer)


## 2. Veranderingen aangezet door Linked Data

 *Op welke **veranderingen** moeten de bestaande organisaties, die nu een dataknooppuntfunctie hebben, zich gaan voorbereiden als deze technologie gemeengoed wordt? Wat gaat dit voor hen betekenen?*

Dit leidde tot de volgende antwoorden:

- Aangezien data opslag, transport en beheer straks 'geen issues' meer zijn, krijgt de organisatie kansen om zich meer te gaan richten op hun afnemers en kennisuitwisseling te gaan ondersteunen.
- We gaan van het beheren van de eigen data, naar het modelleren van kennis ingezet over meerdere organisaties en processen.
- Taken waar data handmatig bij elkaar wordt gebracht worden meer geautomatiseerd.
- Er komt meer vraag naar cognitieve kennis rond kennismodellen.
- Ontologieën en modellen op basis van LD zijn gemeengoed aan het worden.
- Er zijn sectoren en organisaties waar Linked Data bottom up tot stand kwam (Kadaster, Rijkswaterstaat, Schiphol).
- Maar nu ontstaan ook veel top down initiatieven waar LD meteen van af het begin wordt voorgeschreven (Digitaal Erfgoed, Beheer Openbare Ruimte/CROW, Overheidspublicaties).

## 3. Geleerde lessen bij het inzetten van Linked Data

 *Welke **lessen** zijn nu al af te leiden uit de bestaande implementaties van deze nieuwe technologie voor de organisatieontwikkeling van deze dataknooppunten? Welke patronen zien we nu al ontstaan?*

We kunnen hier een viertal domeinen onderscheiden waar de geleerde lessen betrekking op hebben:

### Proces:

- Maak gebruik van een open proces met korte cycli.
- Laat het proces bewaken door iemand die zich met innovatie bezighoudt.
- Zorg voor balans in het begin, probeer niet alles te modelleren.
- Je moet een leertraject in willen gaan.
- Zorg dat er een groep enthousiastelingen is.

### Stakeholders:

- Betrek de bronhouder vanaf het begin.
- Betrek de afnemer vanaf het begin.
- Identificeer de knowledge worker binnen de organisatie
- Eigenaarschap, toegankelijkheid en beveiliging vanaf dag één.
- Zorg voor link met strategisch niveau.
- Train je mensen.
- Zoek als organisatie elkaar op.


### Business:

- Zorg voor een snelle koppeling naar toepassingen.
- Zorg tijdig voor een goede interface.
- Zorg voor een goede eerste use case.
- Zet de business in de lead.
- Neem de organisatie steeds mee.

### Techniek:

- Koppel de techniek los van Linked Data.
- Geen enkel formaat is heilig.
- Betrek niet iedereen in de techniek.

## 4. Centrale aansturing bij Linked Data

 *Wat verandert er straks voor de **centrale aansturing** van informatie-uitwisselingsketens? Is die straks nog wel nodig?*

Hierop kwamen de volgende antwoorden:

- Aansturing vindt straks voornamelijk plaats op het niveau van gedeelde definities en sectorale kennismodellen.
- Er komt meer aansturing op een gezamenlijk integratie van databronnen voor hergebruik zodat niet elke partij afzonderlijk deze integratie moet gaan doen.
- Er komt meer aansturing op gezamenlijke datalabs en herbruikbare analyse elementen.
- Garanderen van openheid en transparantie hoeft niet meer door het centraliseren van data en centrale aansturing van processen. Meer decentralisatie en federatie wordt mogelijk.
- LD creëert nu verbonden eilanden in een aantal sectoren. Aansturing moet er nu komen om deze gekoppelde eilanden cross sectoraal te gaan verbinden. Aansturing blijft nodig om tot brede (boven) sectorale modellen te komen.
- Aansturing blijft ook nodig om risico's rond ongewenste toegankelijkheid af te blijven dekken. Als data eenvoudiger wordt om te delen en ook toegankelijker wordt, dan wordt aansturing op toelaatbaar gebruik steeds belangrijker.

## 5. FAIR en Linked Data

FAIR van het GO FAIR initiatief is een open source verzameling aan principes die het makkelijker maakt om digitale assets uit te wisselen tussen verschillende partijen en ze voor te bereiden op interpretatie en verwerking door computers met minimale menselijke tussenkomst. FAIR geeft tools voor het opstellen van goede metadata, protocollen voor authenticatie en autorisatie, best practices voor opslag en retentie, en instellingen voor het uitwisselen van asset- en planningsdata tussen verschillende partijen, bijvoorbeeld voor infrastructuur-beheerders<sup>7</sup>.

Alle experts benadrukten de rol van FAIR in combinatie met Linked Data:

- De inzet van de FAIR regels, op basis van LD, wordt in verschillende sectoren nu als vanzelfsprekend gezien.
- FAIR voegt governance toe aan de inzet van LD.
- FAIR heeft in de wetenschappelijke wereld zeer snel goede resultaten bereikt.
- FAIR helpt om de voordelen van LD beter te benutten in een ecosysteem van partijen die data uitwisselen.
- FAIR voegt een helder interoperabiliteitsdoel toe aan LD.
- LD is ideaal voor het toepassen van de FAIR principes.
- LD en FAIR worden steeds vaker in combinatie met elkaar ingezet.

<sup>7</sup> [www.linkedin.com/pulse/why-asset-data-must-fair-sido-case-jan-voskuil/](http://www.linkedin.com/pulse/why-asset-data-must-fair-sido-case-jan-voskuil/)



## 6. AI en Linked Data

De basisvorm waarin Linked Data wordt opgeslagen, de netwerk structuur van de objecten, kan in feite gezien worden als een manier om niet alleen betekenis maar ook kennis van een object op te slaan (knowledge graph). In die zin heeft Linked Data ook een relatie met Artificiële Intelligentie toepassingen.

*In knowledge representation and reasoning, knowledge graph is a knowledge base that uses a graph-structured data model or topology to integrate data. Knowledge graphs are often used to store interlinked descriptions of entities - objects, events, situations or abstract concepts - with free-form semantics. Since the development of the Semantic Web, knowledge graphs are often associated with linked open data projects, focusing on the connections between concepts and entities. They are also prominently associated with and used by search engines such as Google, Bing, and Yahoo; knowledge-engines and question-answering services such as WolframAlpha, Apple's Siri, and Amazon Alexa; and social networks such as LinkedIn and Facebook - Wikipedia*

De experts zagen ook deze verbanden nu al ontstaan:

- LD en functionele modellen bieden straks een belangrijke bron voor AI toepassingen.
- LD en AI kunnen straks ingezet worden voor automatische datakwaliteitsverbetering.
- Door LD te gebruiken voor AI blijft het antwoord herleidbaar naar de brondata.
- LD stelt ons in staat om deductie en inductie toe te passen in netwerken van gekoppelde kennis en dan de bijbehorende data te raadplegen.
- AI op basis van LD kan in de toekomst ook kennisintensieve taken automatiseren.
- AI is straks een efficiënt hergebruik van de investeringen die we nu doen in LD.

## De sessie

De sessie begon met een inventarisatie van de problemen die nu spelen bij datadeling zoals die door de deelnemers werden ervaren. Het doel van de sessie was dan ook om gezamenlijk te verkennen in welke mate Linked Data een oplossing biedt voor die problemen.

### Problemen bij datadeling

Hieronder is een overzicht weergegeven, afkomstig uit de sessie. Het laat de typische problemen zien die partijen ervaren met het delen van data.



Zoals we vaker zien is vertrouwen, of het ontbreken daarvan, een belangrijk struikelblok voor delen van data. Maar ook Semantiek (betekenis) en Governance scoren hoog.

### Presentaties

Hierna verzorgde Laurens Lapré een korte functionele introductie van Linked Data. Hierbij werden de volgende vragen behandeld: Wat is het? Wat maakt het anders dan bestaande oplossingen? Waarvoor wordt het nu ingezet? Wat zijn de voordelen?

Daarna gaf Jan Voskuil inzicht in het werk wat gedaan is voor Schiphol en het belang van Linked Data voor de vernieuwing van de Asset Informatie systemen. Duidelijk werd waar in de informatieketen Linked Data een rol had en tot welke baten dit leidden. Verder kwam aan bod welke lessen daar geleerd waren die ook voor andere organisaties relevant zijn.

Vervolgens werden de resultaten van het onderzoek gedeeld door Laurens Lapré, zoals die in het vorige hoofdstuk zijn beschreven, op basis van de interviews met de experts. Hierbij kwamen de volgende zaken aan bod: Terugkoppeling van interviews, impact op organisaties, geleerde lessen, impact op centrale aansturing en het toekomstperspectief. Er werd afgesloten met de strategische positionering die Linked Data nu bereikt heeft binnen het Kadaster in relatie tot de informatiedoelen van die organisatie.

## Discussie

Door de verschillende vragen uit het publiek ontstond er een interactieve sessie. Zo vroeg men zich af:

- Hoe verhoudt data analytics zich tot een bredere visie van een organisatie?
- In hoeverre zijn ketenpartijen bereid om data uit te wisselen?
- Hoe is een uitwisselingsinitiatief het beste georganiseerd?
- Wat betekent LD in de praktijk voor de interoperabiliteit?
- Hoe wordt er gekeken naar domein-overstijgende modellen?
- Hoe verhoudt linked data zich tot initiatieven in andere sectoren?
- Hoe borg je vervolgens de kwaliteit?

Dit was een van de vragen waar langer bij werd stilgestaan. De wrijving tussen business en semantiek werd gezien als een belangrijk aandachtspunt. Volgens de experts is hier nog veel werk te doen. Er is nog veel ruimte voor verbetering bij de validatietechnieken, om zo beter met de kwaliteitsverschillen om te gaan.

- Wat zijn de huidige regels op dit gebied? Wat voor coördinatie gaat dat opleveren?

Ook de coördinatie riep bij veel partijen vragen op. Zo zijn er bottom-up ontwikkelingen, maar ook horizontale ontwikkelingen.

- Hoe kan ik (het draagvlak van) bestaande semantische modellen achterhalen?

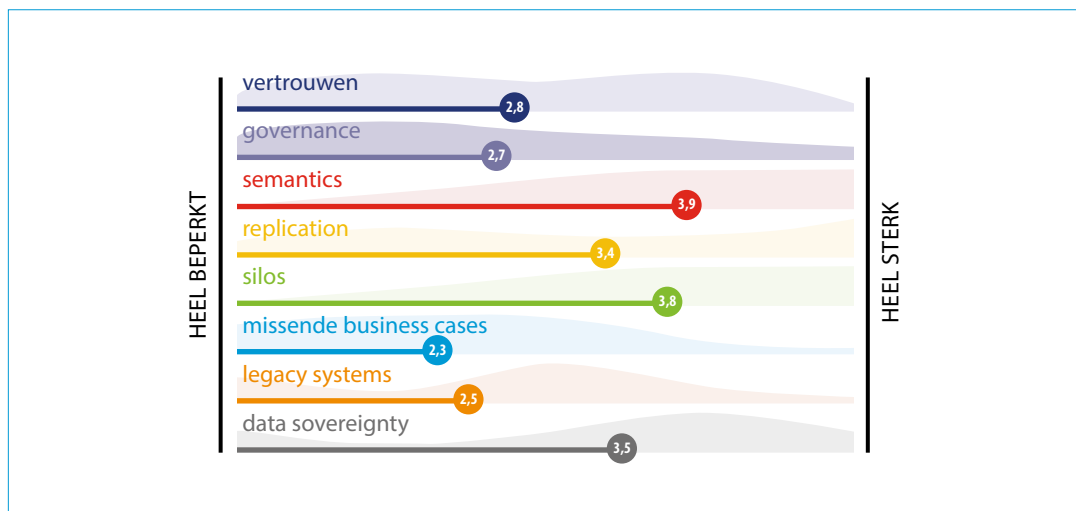
Volgens de experts is er inderdaad meer overzicht nodig. Wel kan je voor veel vragen terecht bij het loket of bij PLDN. Verder werd er gesproken over verschillende toepassingen, en wat dit betekent voor de opzet en organisatie tussen partijen. Zo werd er stilgestaan bij de kansen en uitdagingen van initiatieven tussen Ministeries. De centrale vraag die hierbij ontstond was hoe je het beste Linked Data eilanden weer kunt verbinden.

Verdere vragen die in de discussie aan bod kwamen waren:

- Wat betekent Linked Data voor data eigenaarschap?
- Tot welk niveau pas je FAIR toe?
- In hoeverre wil je data en algoritmes centraal dan wel decentraal hebben?
- Hoe volwassen is Linked Data precies? In hoeverre is er sprake van maatwerk?

## Conclusie

In de sessie is gezamenlijk verkend in welke mate Linked Data nu een oplossing lijkt te bieden voor de typische problemen bij datadeling die eerder waren aangegeven. De concluderende vraag was hier dan ook in welke mate Linked Data de top problemen adresseert die eerder waren geïnterviewd:



We zien dat met name op het gebied van duidelijke Semantiek en het doorbreken van Silo's dat de toegevoegde waarde van Linked Data onderkend wordt. Meerdere partijen gaven ook aan dat de governance inrichting meer aandacht verdient. Het opbouwen van vertrouwen moet hierbij een belangrijk aandachtspunt zijn. Dat vereist samenwerking.

Al met al kon worden geconcludeerd dat Linked Data een groot potentieel heeft voor veel partijen, maar dat het ook nog veel vragen oproept. De uitdagingen waar veel partijen voor staan vraagt om openheid en samenwerking. Het voortzetten van kennissessies kan bijdragen aan het stapsgewijs inzetten van een leerproces om met deze technologie om te gaan en de potentie ervan ten volle te gaan benutten.

## Expert interview: Wouter Beek



Dr. Wouter Beek is entrepreneur en wetenschapper. In 2018 richtte hij samen met Laurens Rietveld Triply op, een organisatie die zich richt op Linked Data software en services. Als onderzoeker heeft hij meerdere wetenschappelijke artikelen geschreven over Linked Data. In 2020 publiceerde hij, samen met o.a. Erwin Folmer, een paper met de titel: *Business Rationale for Linked Data at Governments: A Case Study at the Netherlands' Kadaster Data Platform*. In dit paper beschrijft hij een kenniskloof. Hij constateert dat er relatief weinig onderzoek is gedaan naar de businesskant van Linked Data. Hierdoor ontbreekt er kennis over business rationales om Linked Data te implementeren.

### 1. In welke mate beïnvloedt nieuwe technologie, zoals semantisch modelleren en linked data, de manier waarop straks informatie-uitwisseling georganiseerd zal zijn? Hoe vindt die beïnvloeding plaats?

Wat je over het algemeen ziet is dat data dichterbij de bron komt. Dat is de richting die het opgaat. Dus waar data vroeger een heel lang traject doorging, sluit Linked Data dichterbij de bron aan. Een ander aspect is verandering in de stappen tussen de bron en het gebruik van de bron. De stappen wordt steeds meer geautomatiseerd, en kennis wordt zo beter vastgelegd en behouden.

Linked Data geeft je technisch een ander manier om je IT-infrastructuur vorm te geven. Tegelijkertijd speelt zich nog een ander, diepgaander proces af. Doordat technisch meer mogelijk wordt, zijn er ook organisatorische veranderingen die je eventueel zou kunnen doorvoeren. Een grotere set organisatie modellen wordt realiseerbaar. Organisatie modellen die in het verleden vanzelfsprekend waren, kun je nu ter discussie stellen. Er zijn een aantal organisatorische transformaties die mogelijk worden gemaakt door het gebruik van Linked Data (kennisgraven).

### ” Linked Data, als disruptieve technologie, beïnvloedt organisaties en creëert uitdagingen (en kansen) voor business development<sup>8</sup>.

In productieketens wordt veel gebruik gemaakt van Linked Data. De uitwisseling van Linked Data over een hele productie heen, waar bijvoorbeeld drie leveranciers op elkaars tussenproducten wachten en waar de reserves heel klein zijn. Door Linked Data kunnen de reserves dan nog kleiner worden gemaakt. Daarnaast is er de biomedische wereld. Hier wordt veel geïnnoveerd met behulp van kennisgraven. Bijvoorbeeld voor het genereren van hypothesen. Als derde denk ik aan de overheid. Daar heeft Linked Data een groot voordeel te bieden, door verschillende takken beter met elkaar te laten samenwerken. Bijvoorbeeld op het gebied van assetmanagement. Linked Data heeft zich dus echt bewezen in de praktijk, op meerdere plekken.

Linked Data kan ook iets betekenen voor A.I. toepassingen. Hierbij zie ik twee verschillende stromen. Ten eerste aan de invoer kant. Met schonere data kun je betere resultaten behalen. Kennisgraven en Linked Data zorgen voor verbeterde invoer voor algoritmes, en dus betere resultaten. Daarnaast zie ik ook een tweede toepassing, namelijk bij het vastleggen van de uitkomsten. De beperkte uitlegbaarheid van de uitkomsten van A.I. algoritmes is nu nog vaak een probleem.

<sup>8</sup> Folmer, Erwin & Ronzhin, Stanislav & Hillegersberg, Jos & Beek, Wouter & Lemmens, Rob. (2020). *Business Rationale for Linked Data at Governments: A Case Study at the Netherlands' Kadaster Data Platform*. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.2984691.

De waarom vraag kun je echter ook terugvoeren naar de kennisgraaf. Dus ik zie kennisgraven als een belangrijke component in toekomstige AI-technieken, zowel aan de verbeterde invoer als ook in verbeterde opslag en uitlegbaarheid van uitvoer. Het combineren van statistische A.I. methodes met vastlegging die begrijpelijk is voor mensen in kennisgraven is een innovatief veld. Ik verwacht dat het om een blijvende ontwikkeling gaat. Een onomkeerbare trend. De integratie van A.I. en kennisgraven zorgt voor een nieuwe manier om het IT-landschap vorm te geven.

**2. Op welke veranderingen moeten de bestaande organisaties, die nu een dataknooppuntfunctie hebben, zich gaan voorbereiden als deze technologie gemeengoed wordt? Wat gaat dit voor hun betekenen?**

Op langere termijn is het zo dat een Linked Data project kan zorgen voor efficiëntere rollen binnen de organisatie. Rollen met veel handmatig werk zullen dan minder nodig zijn. Vaak zie je dan een verlegging van de soort taak. Van data handmatig bij elkaar brengen en verrijken naar meer geautomatiseerd werk en controlestappen. Van manipulatie naar curatie. Die transitie zie ik heel duidelijk. Of dat betekent dat een afdeling van een organisatie inhoudelijk fundamenteel verandert of misschien niet langer nodig is, is moeilijk te zeggen. Dat hangt natuurlijk ook af van andere zaken binnen de organisatie. Wat je tenminste ziet is een verlegging van de soort van taak. Men is nu veel meer bezig met het cureren van een proces, wat kan zorgen voor minder benodigde mankracht en kosten.

**” De voordelen van Linked Data zijn soms meer organisatorisch van belang dan technisch.**

Ik zou willen beargumenteren dat veel van de voordelen van Linked Data juist op business niveau interessant zijn. Herleidbaarheid is vaak een heel belangrijk principe voor de organisatie. Betrouwbaarheid ook. De voordelen van Linked Data zijn soms meer organisatorisch van belang dan technisch.

Het is belangrijk om Linked Data op de juiste plek binnen de organisatie te laten landen. Linked Data in het technische team is niet voldoende. Vaak hebben organisaties modellers in dienst. Daar leg ik het liefst de verantwoordelijkheid voor het onderhouden van het Linked Data kennismodel neer. Op het moment dat een Linked Data traject overgaat van in de steigers naar productie vind ik eigenlijk dat het bij het modellen-team moet worden neergelegd. Een andere groep die je binnen de organisatie moet betrekken bij de Linked Data ontwikkelingen zijn de data scientists. Ik merk wel dat je als organisatie eigenlijk een ander soort data scientist nodig hebt. Iemand die niet alleen goed data aan elkaar kan plakken, maar ook een soort knowledge worker is. Die meer vanuit kennis het onderhoud kan uitvoeren. Zo iemand hoeft niet altijd super technisch te zijn.

Bij Kadaster is Linked Data sinds een jaar of drie echt een productiedienst geworden. Het is daarmee een belangrijke pijler geworden voor de infrastructuur van de organisatie. Het is niet makkelijk meer weg te denken. Ondertussen zien we dat het Data Science Team van Kadaster vaak organisatie-brede vragen moet beantwoorden. We zien dat daar eigenlijk altijd een Linked Data voorziening voor nodig is. Dus voor de geïntegreerde organisatiebrede vragen van Kadaster wordt nu heel breed gebruik gemaakt van Linked Data.

**” ...het is de grootste publieke Linked Data implementatie in Nederland, en een van de grootste in de publieke sector wereldwijd.**

### 3. Welke lessen zijn nu al af te leiden via de bestaande implementaties van deze nieuwe technologie voor de organisatieontwikkeling van deze dataknooppunten? Welke patronen zien we nu al ontstaan?

Er zijn een aantal lessen waar ik aan moet denken.

**Les 1:** Zet de business in de lead

**Les 2:** Betrek de bronhouder vanaf het begin

**Les 3:** Betrek de afnemer vanaf het begin

**Les 4:** Identificeer de knowledge worker binnen de organisatie

Ten eerste, zet de business in de lead. Het liefst over de gehele keten. In het verleden zijn we te veel met een technische insteek aan de slag gegaan. Het belang van Linked Data wordt veel sneller herkend door business dan door IT. In de praktijk spreekt een Linked Data project dus veel meer aan als je het vanuit de business start. Verder zit het op de verkeerde plek als het door een puur technisch team wordt opgepakt. Het is weleens voorgekomen dat teams van ontwikkelaars niet overtuigd waren van de voordelen van Linked Data. Juist de business kant was dan doorslaggevend. Bijvoorbeeld m.b.t. het voldoen aan bepaalde wetgeving.

**» *Linked Data zit op de verkeerde plek als het door een puur technisch team wordt opgepakt.***

Ten tweede, betrek de bronhouder van de data vanaf het begin. Dit is een van de belangrijkste aspecten om mee te beginnen. Hier zit namelijk de kennis. De bronhouders zijn in sommige Linked Data projecten te weinig aangehaakt, met als gevolg dat er vanuit de bron niet geleverd kon of mocht worden. Dan zit je Linked Data aanpak meteen al op het verkeerde spoor, want vervolgens verrijk je een afgeleide van de bron. Zonde, want Linked Data is veel meer dan een syntactische operatie, het is kennis vastleggen. Je moet de bron ondersteunen bij het opslaan en beschikbaar maken. Zonder bronhouder beperkt het zich tot een losse technische exercitie.

Ten derde, voortbouwend op het vorige punt, betrek de consument/afnemer ook vanaf het begin. Hierdoor ontstaan de interessante use cases. Die zijn bij Linked Data eigenlijk voor het oprapen. De meeste interessante vragen die op dit moment heel moeilijk binnen organisaties te stellen zijn gaan over twee of meer bronsystemen heen. En met Linked Data is dat juist heel makkelijk. Met andere woorden, resultaten met hoge impact zijn eigenlijk juist heel dichtbij. Door de afnemer vanaf het begin te betrekken, blijven de use cases ook in het vizier.

Ten vierde, het is essentieel dat je knowledge workers binnen de organisatie identificeert en dat ze vanaf het begin betrokken zijn bij het traject. Een falend Linked Data project is een project waarbij technenuten inhoudelijke beslissingen gaan nemen over de data. Dan weet je eigenlijk al zeker dat het mis gaat. Ook daarom moet juist de business in de lead zijn, en moet er een plek zijn voor de knowledge worker.

**» *Een falend Linked Data project is een project waarbij technenuten inhoudelijke beslissingen gaan nemen over de data.***

#### **4. Wat verandert er straks voor de centrale aansturing van informatie-uitwisselingsketens? Is die straks nog wel nodig?**

Het belangrijkste centrale element voor de overheid is natuurlijk de afspraken die we maken als samenleving. De wetgeving is de centrale afspraak. Vervolgens is de vraag: hoe kunnen we een open en transparante democratie het beste inrichten, met verschillende overheidsorganisaties en IT-systemen? Het traditionele antwoord is centraliseren. Door de komst van Linked Data kan het speelveld echter veranderen. De vraag hoe je een open en transparante democratie kunt regelen kan anders worden beantwoord. Je krijgt een middel voor veel meer decentrale organisatie.

**” Door de komst van Linked Data kan het speelveld veranderen. De vraag hoe je een open en transparante democratie kunt regelen kan anders worden beantwoord.**

In de publieke sector zie ik nog een aantal organisatorische uitdagingen. In de commerciële wereld is het bijvoorbeeld zo dat er harde afspraken worden gemaakt tussen partijen over de beschikbaarheid van data. Een te lage beschikbaarheid kan dan leiden tot een boete. In de praktijk durft de overheid dat nog niet aan. Dat levert problemen op. Op die manier ontstaan decentrale systemen zonder mechanisme om permanente beschikbaarheid van data af te dwingen. Verder worden er nog vaak verkeerde keuzes gemaakt op decentraal niveau. Wanneer het bij een grote overheidsorganisatie wordt uitgezet is Linked Data vaak succesvol. Dat is bewezen. Echter, bij het decentraliseren ontstaan er problemen. Men koopt verkeerde producten in of gaat zelf bouwen. Zelf met de techniek aan de slag. Zo ontstaan er weer nieuwe standaarden. Juist wat je niet wilt.

Ik denk dat de ‘oude wereld’ en de ‘nieuwe wereld’ nog heel lang zullen blijven voortbestaan. Soms is de ‘oude wereld’ ook niet per se een oude wereld. Ik zie Linked Data meer als gradiënt. We introduceren een aantal aspecten van Linked Data binnen de organisatie. We denken na over wat bepaalde concepten precies betekenen, waarom data vaak onnodig wordt gekopieerd. Vervolgens zetten we een traject in om steeds meer gebruik te maken van Linked Data technologieën. Je gaat niet van 0 naar 100. Je gaat heel geleidelijk misschien richting de 100.

Ik zie als belangrijke volgende stap in de ‘nieuwe wereld’ het verbinden van de eilanden. Linked Data begint bij een aantal organisaties in productie te gaan. Hier en daar is dat al zo. De grote volgende stap is de verbinding tussen die eilanden. De echte meerwaarde komt pas als je naar het grotere netwerk gaat. Bijvoorbeeld als verschillende overheidsinstanties onderling Linked Data uitwisselen, of wanneer dit in de private sector tussen leveranciers en afnemers gebeurt. De beloofde meerwaarde van Linked Data zit in die directe aansluiting en dat gaat zich nu beproeven denk ik, binnen nu en de komende vijf jaar.

**” De beloofde meerwaarde van Linked Data... gaat zich nu beproeven denk ik, binnen nu en de komende vijf jaar.**



## Expert interview: Jan Voskuil



Jan Voskuil is architect en entrepreneur. Hij is al een kleine 20 jaar bezig met de materie rondom Linked Data. Als architect werkt hij voor het Kennis- en Exploitatiecentrum voor Officiële Overheidspublicaties (KOOP). KOOP ondersteunt overheidsorganisaties in Nederland om elke dag informatie zoals wet- en regelgeving, parlementaire stukken, vergunningen en internationale verdragen te publiceren. Hierbij wordt gebruik gemaakt van Open- en Linked Data. Ook werkt hij voor Schiphol, met focus op assetmanagement. Hij is CEO van Taxonic, een consultancy bedrijf op het gebied van semantische technologie in business processen. Hij heeft meerdere publicaties op zijn naam over het semantische web en Linked Data.

### 1. In welke mate beïnvloedt nieuwe technologie, zoals semantisch modelleren en linked data, de manier waarop straks informatie-uitwisseling georganiseerd zal zijn? Hoe vindt die beïnvloeding plaats?

Linked Data is een krachtig concept en brengt veel potentie met zich mee. Het is daarnaast heel anders dan andere technologieën. Men denkt weleens goed uit de voeten te kunnen met Linked Data vanwege XML-kennis. Dat valt vervolgens enorm tegen. Ik werk nu al zo'n tien jaar met Linked Data in een projectmatige context, dus buiten het laboratorium. In begin is dat niet makkelijk geweest, in die zin dat het lastig was om echt meters te maken. De laatste jaren begint het echt te lopen. Daar krijg ik veel energie van. Het is bijzonder om te zien hoeveel er mogelijk is. Ik heb er veel vertrouwen in dat het nog heel lang blijft doorgroeien en veel groter gaat worden.

FAIR is een idee waar je het moeilijk oneens mee kan zijn. Vervolgens moet het geïmplementeerd worden en dan begint het gekrakeel. Bijvoorbeeld bij samenwerking in de ondergrond. Omdat de FAIR-principes heel abstract zijn kunnen die op oneindig veel manieren geïmplementeerd worden. In de praktijk zie je dat dat snel leidt tot discussies en verschillende partijen die verschillende dingen doen, zonder tot onderling compatibele data te komen. Dan schiet je het doel voorbij. Dat is het nadeel. Het grote voordeel is dat het als gedachtegoed de essentie samenvat. Data moet: I) vindbaar, II) toegankelijk, III) interoperabel en IV) herbruikbaar zijn. Ook al zijn de principes niet per se meteen makkelijk uit te werken, het is en blijft een duidelijke boodschap. Dat helpt enorm mee om de voordelen van Linked Data te benutten. Je overtuigt mensen niet met een techniek, maar met een doel. FAIR zet een doel neer, en zorgt voor een heldere onderbouwing. De principes zijn nu helder en duidelijk vastgelegd in een publicatie in een artikel in Nature. Daarnaast beschrijft het artikel het grote maatschappelijke belang van het doel dat met deze principes wordt gediend. Het artikel heeft veel gezag en dat helpt enorm in de discussie.

De uitwerking is de uitdaging waar we met z'n allen voor staan. Het bewijs dat je met Linked Data de FAIR-principes kan implementeren is in de medische wetenschap allang ontdekt. In andere domeinen breekt dat inzicht nog maar mondjesmaat door, maar het ziet er goed uit.

**” Je overtuigt mensen niet met een techniek, maar met een doel.**

Linked Data stelt je in staat om te modelleren op basis van kennis, in plaats van op basis van bijvoorbeeld assets. Door de verschillende projecten die zijn opgepakt begint er ook een basis te ontstaan voor een eventuele volgende stap: AI. Dat is iets wat nu enorm actueel is. De link tussen semantisch web en AI is cutting edge technology. In de academische wereld wordt er veel aandacht aan besteedt.

De hoop is nu dat AI een belangrijk aandeel kan leveren in het genereren van kennis uit grote hoeveelheden data. Denk bijvoorbeeld aan predictive maintenance in de assetmanagementwereld. Het begint nu echt van de grond te komen. Wij werken bijvoorbeeld nu met een partner die planningen optimaliseert. Dat kan bij grote organisaties zo 10% van het onderhoudsbudget schelen. Het grote probleem bij het project is dat er veel, heel veel data nodig is, en die worden overal vandaan gehaald. Vervolgens moet de data eerst verschillende keren worden getransformeerd voordat het bruikbaar is. Heel arbeidsintensief. Stel nou dat je dat kunt rechttrekken door FAIR-data te gebruiken, met gedeelde modellen, dan wordt het veel krachtiger. De limiet is niet de techniek, maar de hoeveelheid bruikbare data die je met beperkte tijd en middelen beschikbaar kunt krijgen. Dus, om echt stappen te maken in de toekomst, heb je FAIR-data nodig.

**” ...om echt stappen te maken in de toekomst, heb je FAIR-data nodig.**

FAIR-data zorgt voor interoperabiliteit. Domeinen die vroeger niks met elkaar te maken hadden kunnen worden gekoppeld. Het is belangrijk om te zorgen dat iedereen dezelfde definitie hanteert bij een object, zoals bijvoorbeeld een brug. Lukt dat niet, zorg dan in ieder geval dat je het weet van elkaar dat er verschillende definities zijn. Met een stukje software kan dat inzichtelijk worden gemaakt. Op dat moment heb je FAIR-data. Linked Data is daar speciaal voor ontworpen. Bovendien ontstaat er een leerproces wanneer men de generieke elementen uit verschillende definities kan halen. De ultieme droom is dat AI dit zelf kan doen.

## **2. Op welke veranderingen moeten de bestaande organisaties, die nu een dataknooppuntfunctie hebben, zich gaan voorbereiden als deze technologie gemeengoed wordt? Wat gaat dit voor hun betekenen?**

Wat je in eerste instanties zag bij een aantal organisaties is een start vol enthousiasme. Vervolgens is er veel tijd nodig om de vaardigheden eigen te maken, om zo de kansen optimaal te benutten. Als dat punt eenmaal is bereikt, kan het hard gaan. Nu, tien jaar later, komt het echt van de grond. Dat is inspirerend om te zien. Het breidt zich uit als een olievlek. Dat neemt natuurlijk niet weg dat het tien jaar gekost heeft. Je ontkomt niet aan de leercurve. Dat is meteen de grootste uitdaging. Hoe zorg je ervoor dat je je mensen opleidt? Hoe zorg je ervoor dat je de kennis in huis krijgt? Dat zijn organisatorisch de grootste barrières bij deze technologie. Ik heb gemerkt dat de beloftes van Linked Data aanspreken bij management. Bijvoorbeeld het realiseren van i) waarde zonder een specifieke applicatie of ii) toenemende herbruikbaarheid van data. Die willen daar wel in investeren.

Bij de Nationale Politie wordt ook breed ingezet op Linked Data. Die zijn, net als Kadaster, door een behoorlijke leercurve heengegaan. Daar begint het nu echt te lopen. Het beheer van referentiegegevens gaat nu veel soepeler. Het wordt zo duidelijk en aantoonbaar dat je eigenaarschap van data goed moet beleggen, en niet aan toeval moet overlaten. Je krijgt vanzelf meer eigenaarschap. Veel betere werkprocessen.

Ook bij Zorginstituut Nederland (ZIN) wordt gewerkt met FAIR-data. Een van de initiatieven is de Personal Health Train (PHT). De organisatie is verantwoordelijk voor standaarden van informatiehuishouding binnen de zorg. Dat is een erg gedecentraliseerde wereld, waar heel veel verschillende soorten informatie worden uitgewisseld. In de praktijk is het merendeel XML. Er wordt al een paar jaar gewerkt aan een nieuw fundament, samen met o.a. de universiteit Twente. Er worden op basis van domeinkennis modellen gebouwd. Bij de PHT zie je ook dat het FAIR Data stuk echt een enorme versnelling krijgt. Het is een goed voorbeeld van hoe je aan de hand van een visie, geformuleerd op strategisch niveau, stap voor stap een techniek als Linked Data kan laten landen in de organisatie. Het is trouwens belangrijk om te beseffen dat het hier heel duidelijk niet om Open Data gaat. Het gaat om gevoelige informatie die heel goed beveiligd is.

Als je kijkt naar de ontwikkelingen bij TNO, ZIN en de Nationale Politie dan zie je dat het wel echt om een trend gaat. De lijst van organisaties die gebruik maakt van Linked Data wordt groter en groter. Bij KOOP is nu veel ervaring opgedaan en is echt een volgende fase bereikt. Hetzelfde geldt natuurlijk voor Kadaster.

### 3. Welke lessen zijn nu al af te leiden via de bestaande implementaties van deze nieuwe technologie voor de organisatieontwikkeling van deze dataknooppunten? Welke patronen zien we nu al ontstaan?

Er zijn een aantal lessen geweest die samen de belangrijkste ingrediënten vormen denk ik.

- |  |
|--|
| <b>Les 1:</b> Je moet een leertraject in willen gaan                       |
| <b>Les 2:</b> Eigenaarschap, toegankelijkheid en beveiliging vanaf dag één |
| <b>Les 3:</b> Train je mensen  |
| <b>Les 4:</b> Zorg dat er een groep enthousiastelingen is                  |
| <b>Les 5:</b> Zorg voor link met strategisch niveau                        |
| <b>Les 6:</b> Zoek als organisatie elkaar op                               |

Ten eerste, je moet een leertraject in willen gaan. Voor veel organisaties is het nieuw. Door een gebrek aan kennis worden er soms verkeerde ontwerpbeslissingen gemaakt. Er zit echt een leercurve aan. Een Linked Data traject kan je van tevoren beter inschatten als je er een beetje gevoel voor gaat krijgen. Je moet er op een gegeven moment induiken en mee aan de slag. Het is bijna niet te voorkomen dat je achteraf andere, betere mogelijkheden in het traject ziet. Het belangrijkste is dat je wel bewust met een leertraject kan starten, zodat je rekening kan houden met nieuwe inzichten.

Ten tweede, er moet vanaf dag één duidelijkheid bestaan over eigenaarschap, toegankelijkheid en beveiliging van de data. Het relevante kernpunt van FAIR in deze context is toegankelijkheid. Zorg dat je grip op de data hebt. Zorg dat je altijd weet wie de eigenaar is. Zorg als eigenaar dat je precies weet wie het gebruikt als dat belangrijk is. De toegankelijkheid en vooral dus ook de beveiliging moet vanaf dag één goed ingeregeld zijn, anders ga je het later niet meer goed krijgen.

Ten derde, train je mensen. Er zijn mensen nodig die kennis van zaken hebben. Zonder die kennis ga je het niet redden. Als mensen maar wat gaan doen dan werkt het niet.

Ten vierde, ik denk dat het belangrijk is om een groepje believers erbij te betrekken. Die zien de kansen en nemen de organisatie mee. Het is heel moeilijk, ook gezien het eerste punt, om de kansen van tevoren in te vullen.

Ten vijfde, het is belangrijk om het enthousiasme ook te vertalen in sponsorship op strategisch niveau. Er bestaan goede voorbeelden van hoe Linked Data een plek kreeg in de lange termijn-visie van grote organisaties. Er moeten mensen bijzitten die voor de langere termijn willen gaan. Die zich er bewust van zijn dat het een leertraject is. Er zullen nou eenmaal weleens dingen misgaan. Het Kadaster verhaal spreekt op dit punt boekdelen. Die hebben fantastische dingen gedaan.

Als laatste, zoek als organisatie elkaar op. Ik moet denken aan digitaal stelsel omgevingswet. Daar zijn voorbeelden van modellen die zijn ontwikkeld door medewerking van meerdere organisaties. Samenwerking is zeker een thema waar je naar moet kijken.

#### 4. Wat verandert er straks voor de centrale aansturing van informatie-uitwisselingsketens? Is die straks nog wel nodig?

Er is bij Linked Data sprake van bottom-up als ook top-down activiteit. Afgelopen tien jaar was er vooral veel bottom-up beweging. Slimme mensen met innovatieve ideeën. Management ging er niet altijd in mee of liet het maar een beetje gaan. De FAIR-beweging maakte veel energie los op strategisch niveau. Verschillende organisaties konden het gedachtegoed van FAIR koppelen aan hun toekomstvisie en bijbehorende doelen. Veel organisaties snappen dat ze moeten blijven leren om zo mee te kunnen gaan met technische innovaties. Die hebben niet de luxe om te blijven vasthouden aan de bestaande manier van werken.

Domeinspecifieke modellen zijn ontzettend belangrijk. Zonder die modellen kom je er niet. Bij KOOP rond ik nu een overkoepelend model af, voor alle overheidsinformatie. Dat kan over van alles gaan, van het aantal covid meldingen per gemeente tot een publicatie van de Staatscourant. Daar heb je een generiek model voor nodig. Dus we gingen opzoek naar domeinoverstijgende modellen. We hebben onder meer gekozen voor de Provenance-standaard. Dat gaat over: hoe kun je veranderingen modelleren? Dat noemen we een *upper ontology*. Dat kan je toepassen op publicaties, maar ook op organisaties. Deze ontologie wordt in allerlei sectoren gebruikt. Mochten twee domeinen ergens overlappen, dan ben je natuurlijk heel blij als je vaste ankerpunten vindt.

De technologie kan ook al snel vragen oproepen rondom sturing. Toen ik betrokken was bij een project waarin werd samengewerkt in de ondergrond kwam dat duidelijk naar voren. Een van de eerste dingen waarover men begon was het borgen van toegang tot de data. Voor een van de deelnemers was het erg actueel. Deze publieke organisatie had data beschikbaar gemaakt voor een datascience-project en vervolgens verscheen er online onverwachts een rapport vol met conclusies. De eigenaar van de data heeft daar veel last van gehad. Die werd benaderd met vragen. Er heerste ongeloof. Het ging om gevoelige data, namelijk assets in de ondergrond. Dus dat moet goed worden gemanaged. Dat is een belangrijk aandachtspunt.

Nog een ander punt wat hierop voortbouwt. Wanneer data eenvoudiger te delen is, betekent dat nog niet dat eigenaarschap niet belangrijk is. Juist wel. Je moet je bronnen hebben wanneer je een conclusie trekt. De analyse en het gebruik van de data moet verantwoord zijn. Je moet ook verantwoordelijkheid hebben om de juiste mensen en alleen de juiste mensen toegang te verlenen tot de data. Dus ik denk dat eigenaarschap alleen nog maar belangrijker gaat worden. Ik verwacht dat het sturingsaspect wat minder prominent wordt. De aandacht zal dus verschuiven. De randvoorwaarden voor gebruik van data worden belangrijker. Mensen gaan steeds meer zelf aan de slag, met een eigen stappenplan. Het vaststellen van die randvoorwaarden wordt echt een nieuw vakgebied. De rol centrale aansturing krijgt daarmee dus eigenlijk een andere invulling. Dat kan als gevolg hebben dat bestaande dataknooppunten meer op een standaardisatie-instituut gaan lijken.

Het fundament is en blijft een gedragen semantisch model. Zonder model houdt het op. En het opstellen van zo'n model valt niet mee. Ze worden al snel te complex of hoog over. Dat kan nog best lastig worden. De modellen moet je zien als de motor voor het geheel. Een computer moet de semantiek kunnen thuisbrengen. Het hele idee is dat je niet hoeft te programmeren om met specifieke data om te kunnen gaan. Organisaties zullen dus ook verantwoordelijkheid gaan krijgen over de semantische modellen van de data, naast de data zelf. Een Linked Data-expert die ik ken benadert het als volgt: haal de semantiek bij de bron. Dat betekent dat je niet alleen de data ophaalt, maar ook de betekenis. De betekenis is echter niet van de organisatie zelf. Het moet gedragen zijn door een bredere groep. Tegelijkertijd moet er wel iemand eigenaar zijn van het model, bijvoorbeeld voor het versiebeheer of beantwoorden van vragen. Ik moet ook denken aan data.nl. Daar wordt gewerkt aan het bouwen van community's rondom data in domeinen. Het probleem wat ontstaat is dat men eigenlijk in de data moet duiken voordat echt duidelijk is waar de grootste kansen liggen. Het betekent op een andere manier omgaan met je informatiehuishouding. Een omslag in het denken.

## Expert interview: Erwin Folmer



Erwin Folmer is senior adviseur bij Kadaster sinds 2015 en docent bij de Universiteit Twente. Gedurende vijf jaar, van 2012 tot 2017, was hij projectleider bij Platform Linked Data Nederland (PLDN). Hij richtte zich op het kenbaar maken van het potentieel van Linked Data, en het adresseren van technische en organisatorische barrières.

### 1. In welke mate beïnvloedt nieuwe technologie, zoals semantisch modelleren en linked data, de manier waarop straks informatie-uitwisseling georganiseerd zal zijn? Hoe vindt die beïnvloeding plaats?

We gebruiken nu meestal het woord kennisgraaf. We bedoelen (grotendeels) hetzelfde. De kennisgraaf en de mogelijkheid tot integraal bevragen gaat misschien wel de killerapp voor Linked Data worden. Met een kennisgraaf kan je over de verschillende data silo's heen. Zo kan je allerlei data verbinden op een manier die zorgt voor integrale beschikbaarheid.

Het gebruik van kennisgraven kan ook invloed uitoefenen op maatschappelijke kwesties. Denk bijvoorbeeld aan fake news, aan ongegronde beweringen in de politieke arena. Hoe ga je daar nou eigenlijk mee om als samenleving? Men stelt de vraag: hoe ben je tot die conclusie gekomen? Hoe heb je dat bepaald? Er is al veel voortgang geboekt in openheid, aangezien veel data en rapporten intussen openbaar zijn gemaakt. Aan de conclusie kant kunnen nog stappen worden gezet. Door hier meer openheid te creëren, wordt het voeren van inhoudelijke discussies vergemakkelijkt. Je kan dan als het ware meekijken met de analyse. De redenering is dan beter te volgen. Dat is eigenlijk een volgende stap: van open data naar ook open analyse en open query's. Zo ontstaat een complete flow van openheid en transparantie.

Verder geloof ik sterk in het concept FAIR. Er zijn periodes geweest waarin ik het vooral daarover had, in plaats van bijvoorbeeld Linked Data. Ik zie FAIR nu als Linked Data plus governance. Naar mijn idee wordt er te vaak Linked Data gepubliceerd zonder aandacht voor governance. Denk aan een eenmalige publicatie, zonder metadata of contactgegevens. Je weet niet hoe vaak die geüpdatet wordt of wat de status is. Ondertussen beseft men dat je niks hebt aan het publiceren van een dataset als er geen aandacht voor governance is. Daarom praat ik het liefst over FAIR data. In de wetenschappelijke wereld heeft FAIR erg snel zeer goede resultaten geboekt. Op de universiteit waar ik werk raden ze nu een cursus aan over dit onderwerp. Dan heb je echt iets bereikt vind ik. Helaas moet ik wel vaststellen dat het in de academische wereld behoorlijk is blijven hangen. De overgang naar de buitenwereld verloopt moeizaam.

Linked Data kan ook een belangrijke bijdrage leveren aan nuttige A.I. Het kan allemaal nog een stuk slimmer. Er zit een enorme kracht in A.I., maar het hangt ook af van hoe slim je de data inhangt. We moeten daarnaast ook echt van de black box af. Je moet niet diep hoeven graven voor de reden van een bepaalde conclusie.

## 2. Op welke veranderingen moeten de bestaande organisaties, die nu een dataknooppuntfunctie hebben, zich gaan voorbereiden als deze technologie gemeengoed wordt? Wat gaat dit voor hun betekenen?

Ik denk dat we nu op een kantelpunt staan. Een paar jaar terug verliepen de implementaties nog een beetje moeizaam. De techniek werkte nog niet optimaal. Nu ontstaat er op een aantal punten geen enkele discussie meer. Sterker nog, het zijn non-issues geworden. Bijvoorbeeld het gebruik van ontologieën. Ieder nieuwe ontologie is nu op Linked Data gebaseerd. Ook het integreren van databases via Linked Data is volledig geaccepteerd. Er ontstaat meer discussie als je het hebt over de database zelf als Linked Data. Maar ook dit wordt steeds makkelijker. Uiteindelijk is dat wel waar je naartoe wilt. We zijn bij het punt aangekomen dat het voor de hand ligt om direct met Linked Data aan de slag te gaan. Dit hangt natuurlijk af van de context. Een use case waarbij het draait om migreren vanuit een bestaande situatie, naar Linked Data in de database zelf, zie ik in de komende vijf jaar nog niet gebeuren.

” *We naderen misschien wel een punt waarbij het voor de hand ligt om direct met Linked Data aan de slag te gaan.*

Bij Kadaster hebben we meerdere dingen gedaan. Het semantische deel is bewezen en loopt goed. De data zit nu dicht bij de bron. Verder is er een belangrijke relatie met integraal bevragen. Een ander doel is data voor iedereen. Traditioneel is het vooral de geo-community die wordt bediend. Echter, de API's en Linked Data bedienen verschillende doelgroepen. We zien dat de API de grootste doelgroep bedient.

” *APIs kun je zien als 'meest gestelde vragen' voorziening. Welke vragen je kunt stellen aan de API is vooraf gedefinieerd. Welke gegevens je in je antwoord terugkrijgt, ook. Daardoor is zo'n API heel gemakkelijk in gebruik, maar niet flexibel. Als je iets anders wilt dan de vragen die de API ondersteunt, zul je meerdere APIs moeten bevragen en de antwoorden zelf combineren<sup>9</sup>. Verder is het zo dat een API nog steeds aanbod gedreven is. Er wordt wel veel beter geluisterd naar de klant, maar uiteindelijk bepaalt de data aanbieder hoe de API eruit ziet.*

Een ander traject waar ik aan moet denken is haal centraal, een grote gemeentelijke ontwikkeling. Gemeentes hadden zich als doel gesteld om alles bij de bron op te vragen. Hier zijn mooie grote stappen gemaakt rondom API's. Men kwam er namelijk achter dat ze geen enkele bron konden uitzetten, omdat diezelfde bron ook werd gebruikt voor het uitvoeren van analyses. Zo gingen gemeentes de Linked Data wereld in.

## 3. Welke lessen zijn nu al af te leiden via de bestaande implementaties van deze nieuwe technologie voor de organisatieontwikkeling van deze dataknooppunten? Welke patronen zien we nu al ontstaan?

Er zijn een aantal lessen geweest denk ik.

Les 1: Koppel de business case en de toepassing los

Les 2: Koppel de APIs los van Linked Data

Les 3: Geen enkel formaat is heilig

Les 4: Maak gebruik van een flexibel proces

Les 5: Zorg voor balans in het begin

Les 6: Betrek niet iedereen in de techniek

<sup>9</sup> Van den Brink. (2021, 26 januari). Linked data: onmisbaar voor doorontwikkeling in samenhang. Geraadpleegd: Geonovum website.

Allereerst, bij Linked Data is er niet sprake van één business case of één toepassing. Het zijn er zo veel. Je moet gaan loskoppelen.

Ten tweede, koppel de API los van het Linked Data verhaal. Hier zijn we eerder de fout ingegaan. We hadden destijds gelijk APIs uitgeleverd en daar werd veel gebruik van gemaakt. Echter, het switchen van technieken leverde voor Linked Data problemen op. Voor ons was het een belangrijke les om APIs los te zien van Linked Data.

Ten derde, de formaten zijn niet heilig. Niemand weet waar het precies heen gaat, welke techniek of toepassing de grootste doelgroep krijgt. Zorgt een opmars van de een bijvoorbeeld voor de overname van de ander, of blijven ze allen bestaan? Hoe evolueren de functionaliteiten zich over de tijd? Dat zijn lastige vragen. Wat zeker is, is dat de techniek van alles mogelijk maakt.

Ten vierde, je moet flexibel zijn in je proces. Ik geloof nu veel meer in de losse treintjes aanpak. We noemen het bij Kadaster een datahub. Het is een soort interne distributie laag. Je moet het zien als verschillende treinen die worden gevoed. Die mogen best naast elkaar bestaan. Ik geloof daar inmiddels een stuk meer in. De hoeveelheid treintjes en de prioriteit per trein kunnen veranderen over de tijd. Dat vraagt om een modulaire aanpak. Ik merk dat de ontwikkelingen op dit gebied nog gaande zijn.

Ten vijfde, zorg voor balans in het begin. Begin niet met een dataset met miljoenen objecten en een datamodel dat bestaat uit een handvol klassen/attributen. Dat werkt niet super goed. Een gebrek aan balans in de dataset is niet ideaal voor Linked Data. In de erfgoedwereld bestaan rijke modellen met niet als te veel instances. Dit is vaak een ideaal startpunt/scenario om met Linked Data te beginnen.

Als laatste, betrek niet iedereen in de techniek. Het is niet altijd noodzakelijk om alle details door te nemen met beslissers. Zo kan je het misschien beperken tot de systematiek van kennisgraven.

#### **4. Wat verandert er straks voor de centrale aansturing van informatie-uitwisselingsketens? Is die straks nog wel nodig?**

Het is interessant om te zien dat meerdere bewegingen zich tegelijkertijd manifesteren. Bij Kadaster was het duidelijk een bottom up beweging en werd vanuit de inhoud gestart. Om gegeven moment was het management overtuigd. In de erfgoedwereld is het volstrekt anders gegaan, namelijk top down. Vanaf het begin was het Ministerie betrokken bij het proces. In strategische documenten was de beoogde techniek altijd terug te vinden. Het idee was duidelijk: we gaan het gewoon doen.

Door Linked Data kan informatie toegankelijker worden, voor iedereen. Zelf stel ik me een ladder voor. Zo'n ladder gaat stapsgewijs van GIS developer naar software developer, data scientist, analist, projectmanager als ook de burger. Ik vind dat de burger bepaalde basisvragen moet de burger kunnen stellen.

Als het gaat om sturing, dan merk ik dat Linked Data nog veel vragen oproept. Er is absoluut aandacht nodig voor de negatieve risico's. Bij Kadaster speelt bijvoorbeeld de vraag: hoe kan men waarborgen dat er geen analyses plaatsvinden die juridisch niet mogen? Ik weet dat we minimaal de plicht hebben om te monitoren, loggen, en eventueel in te grijpen bij discutabele analyses. We hoeven denk ik niet op voorhand te blokkeren. Verder zijn er nog andere vragen zoals: wat gebeurt er als de gebruiker zijn eigen data ergens aan kan koppelen? En stimuleer je fout gedrag door iets makkelijker te maken? Hoe men kijkt naar deze vraagstukken hangt ook een beetje af van het politieke klimaat natuurlijk. Ik geloof erg in openheid en transparantie. Ik vind het belangrijk dat je moet kunnen uitleggen wat je doet. We mogen er allemaal wat van vinden.

Een van de centrale vraagstukken is data eigenaarschap. Ik denk dat het een kwestie van tijd is voordat we erachter komen dat de discussie rondom regie over gegevens een tussenfase is. Met meer inzicht heb je het probleem namelijk niet opgelost. De crux is dat de data van de burger moet zijn. De Vlaamse overheid heeft nu stappen gezet richting een soort data depot van de burger, waar je zelf kan instellen met wie je de data deelt. Het basisidee, dat de data van mij is, dat ik bepaal met wie ik het deel, en dat bedrijven dat ook doen, zie ik ook als killerap van Linked Data. Voor de langere termijn is dat een belangrijke use case. Misschien wil je als publieke organisatie een depot provider zijn, of wil je een Kadaster depot voor ieder burger. De burger houdt het zelf in beheer, en kan mandaten uitgeven om anderen in de data te laten kijken. Digitale kluisjes dus. We denken minder in berichtenuitwisseling maar in stukjes informatie en rechtstreekse toegang. De informatie in mijn kluis is uniek.

Bij de sturing van FAIR vallen mij ook een paar dingen op. Ik krijg namelijk steeds meer een bijmaak van de verschillende partijen die zich bezighouden met FAIR, zoals stichtingen, evaluators en bureaus. Het voelt niet helemaal open, wat juist haaks staat op de FAIR-principes. Dat is eigenlijk heel jammer. Anders had FAIR-data het veel verder kunnen schoppen denk ik.

Ook zijn we klaar met de open data movement denk ik. De volgende stap is kwaliteit, in plaats van kwantiteit. De volgende stap is koppeling maken. Voorheen was er bij gemeentelijke data voor iedere koppeling een apart linksetje nodig. Daar zijn nu betere manieren voor bedacht. Daardoor verloopt de governance veel soepeler, en dat betekent flinke winst.



**Connekt/Topsector Logistiek**

Ezelsveldlaan 59

2611 RV Delft

+31 15 251 65 65

[info@connekt.nl](mailto:info@connekt.nl)

[www.connekt.nl](http://www.connekt.nl)

