

Knowledge Graph voor weginfrastructuur, Het kloppend hart van een gemeenschappelijk dataplatform

Berwich Sluer

Pavement expert, &NBSP Partners in Solutions B.V.

berwich@nbsp4solutions.nl

Natascha Poeran

Pavement senior consultant, &NBSP Partners in Solutions B.V.

natascha@nbsp4solutions.nl

Hans Schevers

Linked-Data expert, Building Bits B.V.

hans@buildingbits.nl

Lukas Hopman

Adviseur Data Gedreven Wegverharding

Rijkswaterstaat - Ministerie van Infrastructuur en Waterstaat

luuk.hopman@rws.nl

Samenvatting

Momenteel wordt informatie over de levenscyclus van wegverharding over een lange periode verzameld door verschillende ketenpartners met behulp van verschillende systemen en in verschillende bestandsformaten. De versnippering van informatie, de vele (bestands)formaten en het ontbreken van een gestructureerde data- en informatieomgeving maken het beheer van de weginfrastructuur vatbaar voor veel herbewerking, informatieverlies, beoordelingsfouten, verkeerde interpretatie van de informatie en belemmert innovatie en ontwikkeling in de weginfrasector. Waardevolle gegevens over het ontwerp- en bouwproces worden zelden overgedragen naar volgende fasen van de levenscyclus van de weg. Het gebrek aan interoperabiliteit, d.w.z. naadloze uitwisseling van gegevens tussen systemen, kan leiden tot schade en verlies van gegevens, evenals een toename van de kosten en tijd voor alle betrokken sectorpartijen.

Om de technische uitdagingen het hoofd te bieden, wordt een (gedistribueerde) knowledge graph op basis van Linked Data gezien als de ideale oplossing. Zo'n Knowledge Graph bevat alle data van de weginfrastructuur. De beschikbare datasets kunnen zich op verschillende locaties bevinden, maar zijn direct verbonden via een (netwerk van) Weginfra ontologie(ën). Dit netwerk van ontologieën is ook Linked Data en kan alle concepten, relaties en attributen moeten omvatten, idealiter voor de hele levenscyclus van weginfra-assets. Het linken van alle datasets aan deze ontologie(ën) moet zorgen voor interoperabele datasets, waardoor de fragmentatie van datasets drastisch wordt verminderd en levenscyclusmodellering van wegen haalbaar wordt.

Dit artikel legt de knowledge graph-benadering uit en richt zich op de ontwikkeling van een ontologie voor weginfrastructuur in Nederland. Het beschrijft ook een eerste generatie applicaties die met deze knowledge graph werken.

Trefwoorden

Ontologie, knowledge graph, Linked Data, verharding, weginfrastructuur

1. Introductie

Momenteel wordt informatie over de levenscyclus van wegen over een lange periode verzameld door verschillende ketenpartners met behulp van verschillende systemen en bestandsformaten. De versnippering van informatie, de vele (bestands)formaten en het ontbreken van een gestructureerde data- en informatieomgeving maken het beheer van de weginfrastructuur vatbaar voor veel herbewerking, informatieverlies, beoordelingsfouten, verkeerde interpretatie van de informatie en belemmert innovatie en ontwikkeling in de weginfrasector [1]. Waardevolle gegevens over het ontwerp- en bouwproces worden zelden overgedragen naar volgende fasen van de levenscyclus van de weg. Het gebrek aan interoperabiliteit, d.w.z. naadloze uitwisseling van gegevens tussen systemen, kan leiden tot schade en verlies van gegevens, evenals een toename van de kosten en tijd voor alle betrokken sectorpartijen. Bovendien gaat de huidige enorme hoeveelheid gegevens over verharding, wegenbouw en weginfrastructuur in de vele verschillende gegevensbronnen en gegevensformaten de menselijke controle en beheer(s)baarheid ver te boven. Relaties tussen verhardingsconstructies, eigenschappen en prestaties van verhardingsmaterialen, monitoring van prestaties en onderhoudsbehoefte van verhardingen zijn onmogelijk te leggen en vast te stellen, hetgeen innovatie, ontwikkeling en evolutie in de weginfrastructuursector naar een moderne geïndustrialiseerde en geprofessionaliseerde industrie belemmert.

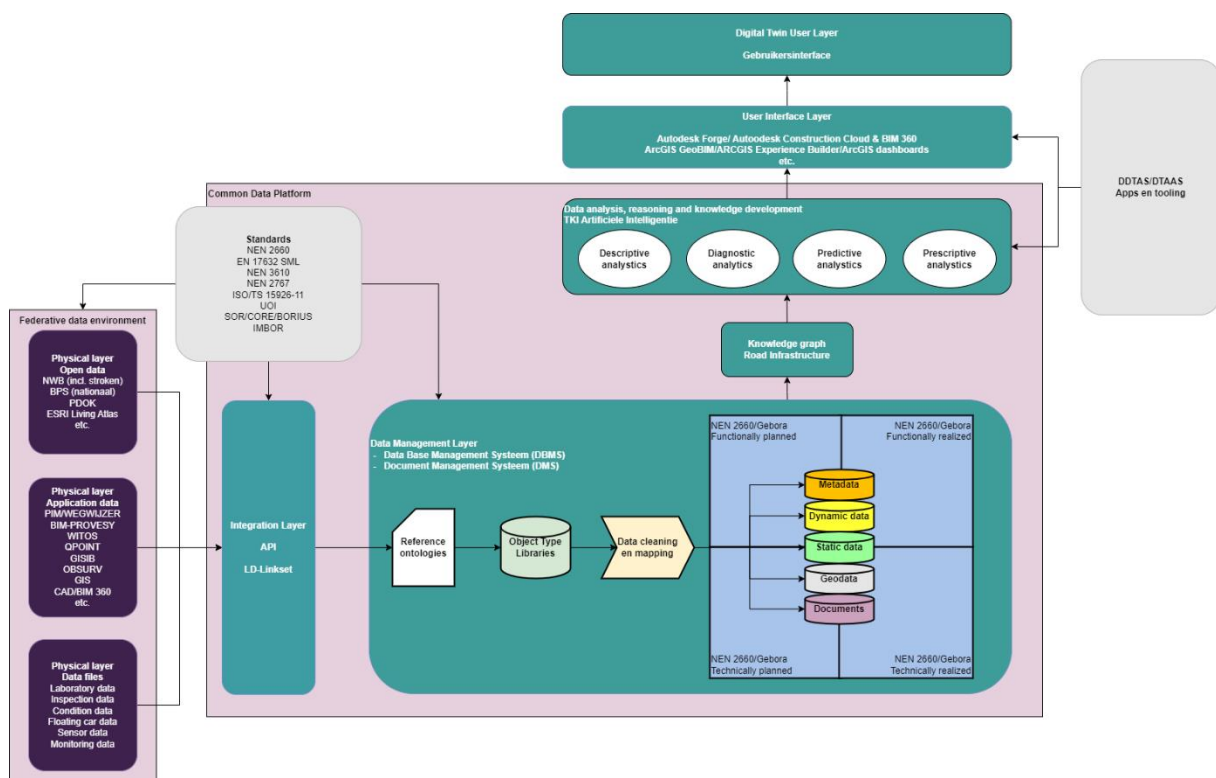
De huidige ambitie van koplopers onder de infrastructuur assetmanagers is het opzetten, implementeren en toepassen van digital twins voor het beheer van data van de gehele levenscyclus van de objecten in het domein, en het gebruik van deze data voor het optimaliseren van kosten, prestaties en risico's voor (datagedreven) assetmanagement. In deze context betekent een digitale tweeling een digitale weergave van de werkelijkheid, die het mogelijk maakt om alle gegevens van de realiteit van de assets in de levenscyclus te structureren, organiseren, toe te passen en zelfs de werking van assets virtueel te simuleren. Om deze ambitie te dienen is het standaardiseren en structureren van data één van de belangrijkste voorwaarden.

Mede door de eerder genoemde ambitie is er momenteel veel (internationale) aandacht voor datamodellering en datastandaardisatie om op een efficiënte, en vooral effectieve manier om te kunnen blijven gaan met de enorme stroom aan data en informatie. Het belangrijkste doel van deze datamodellering en datastandaardisatie is om te komen tot een situatie waarin data wordt gecentraliseerd en daarmee onafhankelijk of losgekoppeld wordt van alle systemen en applicaties waarin deze data worden gebruikt. Het beheer van data en informatie wordt hierdoor veel eenvoudiger. In plaats van wijzigingen door te voeren in alle systemen waarin overeenkomstige gegevensitems worden toegepast, vinden de wijzigingen slechts plaats op één centrale plek in de database van herkomst. Alle systemen en applicaties maken dan gebruik van dezelfde uniforme, gestructureerde data ('**single source of truth**').

In Nederland hebben onder andere Rijkswaterstaat (Ministerie van Infrastructuur en Waterstaat), de provincie Gelderland en het Havenbedrijf Moerdijk in samenwerking met marktpartijen al diverse proof of concepts (PoC's) uitgevoerd voor het delen en/of uitwisselen van assetdata van de opdrachtgever en aanlegdata van de aannemer binnen één centrale data-omgeving. In deze centrale data-omgevingen zijn gegevens uit verschillende bronnen geïntegreerd, gestandaardiseerd en gestructureerd met behulp van kleine specifieke

gegevensbrongerelateerde ontologieën, waarbij alle gegevens federatief worden beheerd bij de eigen bron. In deze PoC's is het principe en de gewenste werking van een digital twin voor assetmanagement op kleine schaal geverifieerd en gevalideerd. De weg lijkt vrij voor praktische implementatie van productieomgevingen van digital twins voor assetmanagement voor de hele weginfrastructuursector.

Op basis van de zeer positieve ervaringen in de pilot-experimenten is besloten om een gemeenschappelijk dataplatform te ontwikkelen op basis van een federatieve data-omgeving [2]. Het raamwerk voor dit gemeenschappelijke dataplatform wordt geïllustreerd in figuur 1. In deze federatieve data-omgeving worden data gedeeld en hergebruikt door alle partijen die toegang hebben tot het dataplatform, terwijl alle data worden beheerd bij de bron van herkomst.



Figuur 1: Visie op een gemeenschappelijk dataplatform voor weginfrastructuur [2,14,15]

De belangrijkste doelen voor de ontwikkeling van een gemeenschappelijk dataplatform voor weginfrastructuur zijn:

- De implementatie en het gebruik van Digital Twins voor weginfra assetmanagement
- Implementatie van datagedreven management binnen organisaties
- Levenscyclusbeheer van specificaties, activiteiten en prestaties van weginfrastructuur
- Kennis van wegverhardingprestaties te optimaliseren, te analyseren, te leren, te ontwikkelen en toe te passen
- Stimuleren van de ontwikkeling van intelligentie, innovatie en professionaliteit in een digitaal getransformeerde weginfrastructuursector.

Om de technische uitdagingen het hoofd te bieden, wordt een (gedistribueerde) knowledge graph op basis van Linked Data gezien als de ideale oplossing.

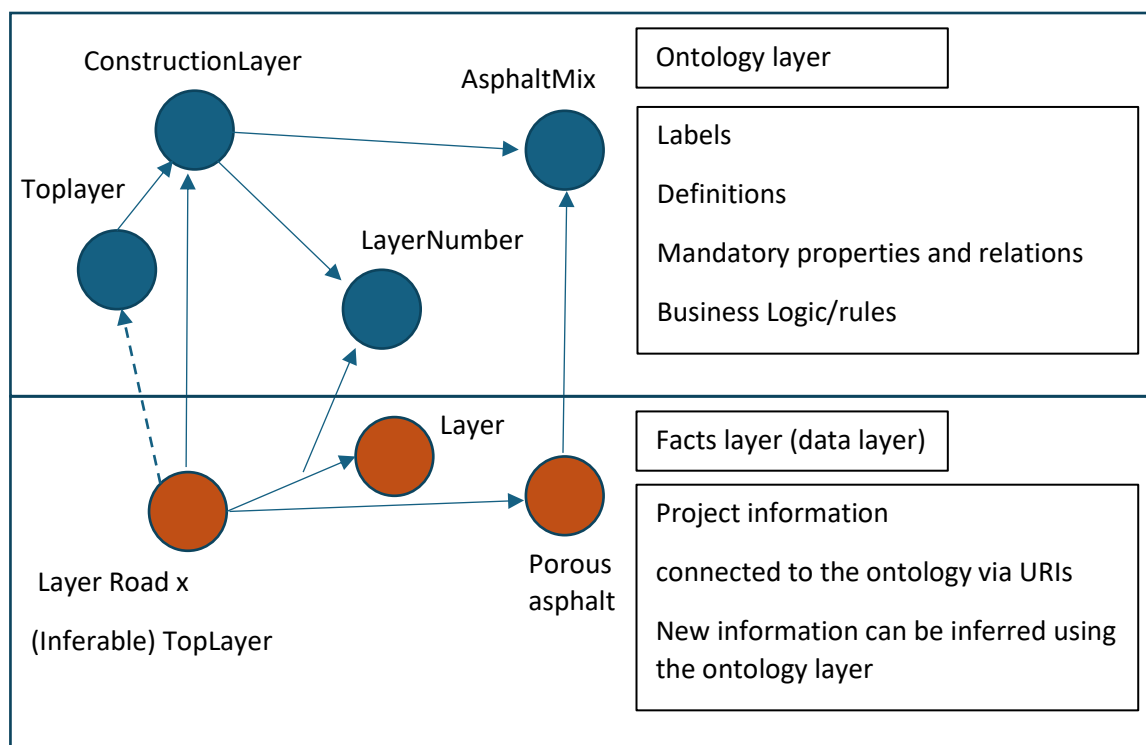
Dit paper beschrijft de voortgang in Nederland met de ontwikkeling van gedeelde data-omgevingen op basis van een knowledge graph, een domeinontologie en het gebruik van Linked Data voor de weginfrastructuursector.

2. Gekoppelde gegevens – Linked Data

Linked Data (LD) is een technologie van het W3C (<https://www.w3.org/wiki/LinkedData>), die gestructureerde gegevens kan verbinden en delen via het web met behulp van standaardprotocollen zoals HTTP, URI's en RDF (Resource Description Framework) [3]. Het maakt het mogelijk om gegevens uit verschillende bronnen te koppelen en te bevragen. Toepassing van geformaliseerde modelleringsregels in de vorm van een ontologie maakt het ook mogelijk om conclusies te trekken over de gegevens, waardoor nieuwe gegevens worden gecreëerd die ook kunnen worden gekoppeld en bevroegd. De business case voor linked data is om data online vindbaar te maken, koppelbaar te maken met andere data en vervolgens deze data machine-verwerkbaar te maken, datasilo's af te breken, datasets te verrijken en datamanagement te verbeteren. Dit met als doel kennis en intelligentie te ontwikkelen, innovatie te ondersteunen en efficiëntie te verbeteren. Linked Data sluit heel goed aan bij de principes van het "datacentrisch manifest" (<http://datacentricmanifesto.org/>). Dit manifest introduceert een reeks duidelijke principes zoals (1) Data zijn een sleutelasset, (2) Data zijn zelfbeschrijvend en niet afhankelijk van een applicatie, (3) data worden uitgedrukt in open, niet-eigendomsrechtgebonden formaten, (4) de toegang wordt georganiseerd op de data (niet beheerd door applicaties) en (5) applicaties raadplegen/(her)gebruiken de data en kunnen iets toevoegen aan de data.

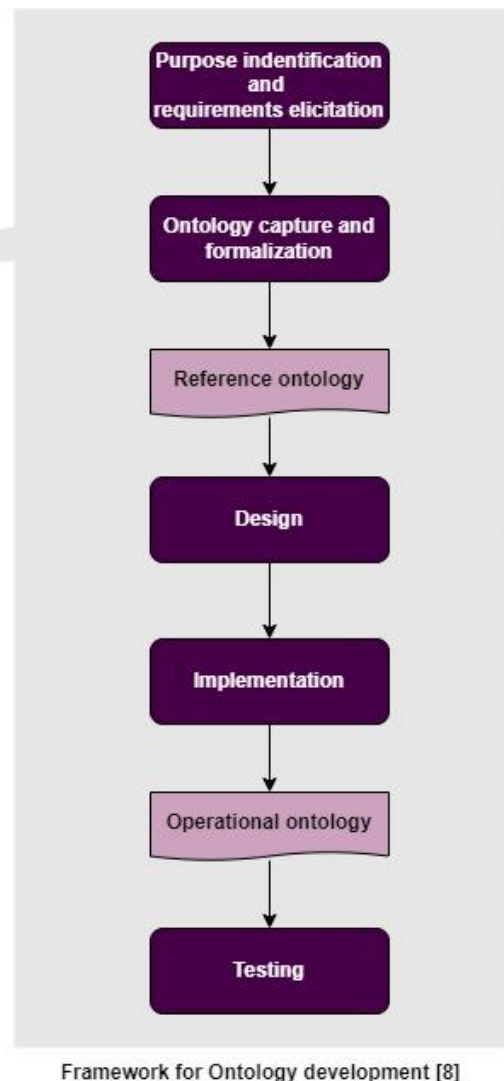
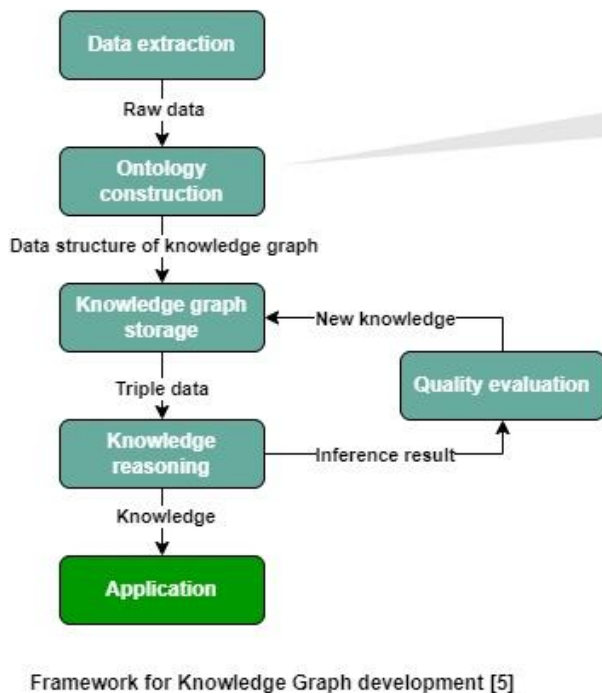
3. Linked Data Knowledge Graphs

Een knowledge graph beschrijft de verschillende concepten, entiteiten en relaties tussen entiteiten in de objectieve wereld op een gestructureerde manier, en biedt de mogelijkheid om enorme hoeveelheden informatie te organiseren, te beheren en te begrijpen [4]. In wezen is een knowledge graph een semantisch netwerk dat de relaties tussen entiteiten legt, gegevensbronnen effectief kan representeren, efficiënt complexe gerelateerde informatie kan vinden en semantische verwerkingsmogelijkheden heeft [5]. Figuur 2 toont een vereenvoudigde Knowledge Graph voor wegverharding. De ontologie beschrijft concepten als ConstructionLayer en Asphalt mix en definieert eigenschappen zoals LayerNumber en relaties met andere objecten. Hiërarchieën kunnen ook worden gebruikt, net als de taxonomie die concepten ordent van ruw naar meer gespecialiseerd. Toplaag kan een meer gespecialiseerd concept zijn met een bovenliggende constructielaag. Gegevens kunnen worden uitgedrukt met behulp van deze ontologie. In Linked Data worden alle data gemodelleerd met behulp van resources die een unieke URI hebben. Alle concepten, maar ook alle data, zijn hierdoor resources en hebben hun eigen unieke URI's.



Figuur 2: vereenvoudigde Knowledge Graph voor weg-infra

Een knowledge graph maakt gebruik van formele beschrijvingen (ontologieën) als fundament om entiteiten en hun eigenschappen en relaties binnen datasets uit verschillende bronnen te verbinden tot een samenhangende graph. Dit kan met behulp van Linked Data-technologie. Op deze manier worden datasets gestructureerd en interoperabel en kunnen ze worden bevraagd, uitgebreid met extra data en knowledge graphs en kunnen ze worden gebruikt om nieuwe data af te leiden met behulp van bijvoorbeeld op geformaliseerde modelleringsregels gebaseerde ontologieën [6,7].



Figuur 3: Knowledge Graph ontwikkelingsproces [5, 8]

Figuur 3 toont voorbeelden van ontwikkelkaders die kunnen worden gebruikt om ontologieën en knowledge graphs te ontwikkelen, te testen en te implementeren.

Graphs vormen de basis van veel moderne data- en analysemogelijkheden en vormen een robuuste ruggengraat voor elk AI- en analyseplatform [9]. Beter begrip van en samenwerking met zakelijke gebruikers, het organiseren en voorbereiden van gegevens voor downstream-processen, het blootleggen van verborgen inzichten, het verbeteren van de ontwikkeling van ML-modellen en het bieden van verklaarbare AI zijn slechts enkele van de toepassingen die worden aangedreven door verschillende graphtechnologieën en -technieken [9].

Knowledge Graphs combineren kenmerken van verschillende paradigma's voor gegevensbeheer:

- Database, omdat de gegevens kunnen worden opgevraagd via gestructureerde query's;
- Graph, omdat het kan worden geanalyseerd als elke andere netwerkgegevensstructuur;
- Kennisbron, omdat de gegevens erin formele semantiek dragen, die kan worden gebruikt om de gegevens te interpreteren en nieuwe feiten af te leiden [9]

4. Netwerk van Ontologieën

Een ontologie is een formele beschrijving van kennis als een reeks concepten binnen een domein, en de relaties die daartussen bestaan. Door de structuur van de kennis in een domein te beschrijven, is de ontologie de basis voor de Knowledge Graph om de gegevens erin vast te leggen. Omdat ontologieën worden gebruikt om gemeenschappelijke modelleringsrepresentaties van gegevens uit gedistribueerde systemen en databases te specificeren, maken ze database-interoperabiliteit, cross-database zoeken en soepel kennisbeheer mogelijk [10].

Een ontologie is een soort schema om datasets te structureren en te modelleren. Het kan ook worden gebruikt om labels, definities en attributen van concepten in de datasets te verstrekken. Andere toepassingen zijn validaties waarbij bijvoorbeeld (inkomende) datasets kunnen worden getest of ze voldoen (op SHACL gebaseerde ontologieën). Ook kunnen voor het afleiden van nieuwe gegevens (op OWL gebaseerde) ontologieën worden gebruikt. Er zijn verschillende andere toepassingen mogelijk, zoals het uitbreiden van de ontologieën met kaartinformatie om datasets te converteren of om interoperabiliteit te ondersteunen. W3C Linked data heeft verschillende gestandaardiseerde vocabulaires voor het opzetten van een ontologie, zoals SHACL dat is gericht op gegevensvalidatie en informatiespecificatie, OWL-vocabulaire gericht op inferentie en SKOS voor eenvoudigere kennismodellering (bijvoorbeeld thesauri).

Volgens SABIO [8] wordt er onderscheid gemaakt tussen referentie-ontologieën en operationele ontologieën. Een **domeinreferentie-ontologie** is een domeinontologie die is gebouwd met als doel een zo goed mogelijke beschrijving van het domein te maken. Het is een oplossingsonafhankelijke specificatie (conceptueel model) met als doel een duidelijke en nauwkeurige beschrijving van domeinentiteiten te maken met het oog op communicatie, leren en probleemoplossing. Zodra gebruikers het eens zijn over een gemeenschappelijke conceptualisering, kunnen operationele versies (machine-leesbare ontologieën) van een referentie-ontologie worden geïmplementeerd. In tegenstelling tot referentie-ontologieën, zijn **operationele ontologieën** ontworpen met de focus op het garanderen van gewenste computationele eigenschappen.

Zoals in verschillende bronnen is aangegeven, zouden domeinontologieën idealiter moeten worden ontwikkeld op basis van **fundamentele (toplevel) ontologieën**. Concepten en relaties in een domeinontologie moeten vooraf worden geanalyseerd in het licht van een fundamentele ontologie [8].

Volgens SABIO [8], figuur 3, worden in eerste instantie het doel van de ontologie en het beoogde gebruik ervan geïdentificeerd. Zodra het doel en het beoogde gebruik van de ontologie zijn gedefinieerd, worden de eisen die eraan moeten worden gesteld vastgelegd. Ontologie-eisen kunnen worden onderverdeeld in functionele en niet-functionele eisen. Functionele eisen verwijzen naar de kennis (inhoud) die door de ontologie moet worden vertegenwoordigd. Niet-functionele eisen verwijzen naar de kenmerken, kwaliteiten en algemene aspecten die geen verband houden met de inhoud van de ontologie. Niet-functionele eisen kunnen worden onderverdeeld in:

- (i) Kwaliteitskenmerken van de ontologie, die verwijzen naar kenmerken die de ontologie, als softwareproduct, moet vertonen, zoals redenering, prestaties en beschikbaarheid (in het geval van operationele ontologieën), bruikbaarheid (bijv. begrijpelijkheid), onderhoudbaarheid (bijv. uitbreidbaarheid)

- (ii) Projecteisen, dit zijn eisen die zijn gesteld aan het betreffende ontologieproject zelf, zoals proceseisen (bijv. naleving van procesmodellen en documentsjablonen), implementatie-eisen (bijv. of de operationele ontologie moet worden geïmplementeerd in een bepaalde machine-leesbare taal), leveringseisen (bijv. time-to-market), consensusgerelateerde eisen (bijv. wie het eens moet zijn met de ontologie)
- (iii) Eisen die verband houden met beoogd gebruik, zoals eisen met betrekking tot kennisbronnen (bijv. of de terminologie die in de ontologie moet worden gebruikt, moet worden overgenomen van standaarden), interoperabiliteitseisen (bijv. of de ontologie moet zijn gebaseerd op een specifieke fundamentele ontologie om de integratie ervan met andere reeds bestaande ontologieën te vergemakkelijken).

Als het domein van interesse complex is, zoals bijvoorbeeld het domein van de weginfrastructuur, wordt de ontologie modulair opgezet. **Ontologiemodularisering** bestaat uit het identificeren van modules (of sub-ontologieën), die afzonderlijk kunnen worden beschouwd terwijl ze onderling verbonden zijn met andere sub-ontologieën. De voordelen van het modulariseren van ontologieën zijn onder meer [8]:

- (i) om de ontwikkeling en het onderhoud van de ontologie te vergemakkelijken door deze op te delen in los gekoppelde, op zichzelf staande modules
- (ii) om het hergebruik van delen van de ontologie te vergemakkelijken
- (iii) om de prestaties te verbeteren door gedistribueerde verwerking mogelijk te maken.

De ontwerpfase van een ontologie is gericht op het overbruggen van de kloof tussen de conceptuele modellering van referentie-ontologieën en de codering ervan in termen van een operationele ontologietaal. De implementatiefase betreft het implementeren van de ontologie in de gekozen operationele taal en ten slotte betreft het testen van ontologieën de dynamische verificatie en validatie van de prestaties van de operationele ontologie op een eindige reeks testcases, tegen de gemodelleerde en verwachte prestaties [8].

In de huidige Nederlandse datamodelleringspraktijk wordt een referentie-ontologie vaak aangeduid als 'ontologie' en wordt de beschrijving van de gedetailleerde specificaties van de ontologie (definities, klassen, attributen) aangeduid met 'Object Type Library (OTL)'. Om verkeerde interpretaties te voorkomen zijn de definities van SABIO (fundamentele ontologie, referentie-ontologie, operationele ontologie) overgenomen. Implementatie of instantiatie van operationele ontologieën op databases resulteert uiteindelijk in knowledge graphs.

Door alle beschikbare relevante ontologieën samen te voegen ontstaat een netwerk van ontologieën. Dit netwerk van ontologieën vormt de basis voor de totstandkoming van de Knowledge Graph voor weginfra (WeginfraNL-KG).

5. Knowledge Graph voor de weginfrasector

Aan de ene kant wordt er veel valide data verspild. Wegbouwkundigen en onderzoekers kunnen de intrinsieke relatie tussen verhardingsconstructies, materiaalspecificaties en -prestaties en de resultaten van monitoring van verhardingsprestaties niet aantonen bij het nemen van beslissingen voor assetmanagement (instandhouding) en ontwikkeltrajecten (nieuwe ontwikkeling). Dit resulteert in beperking van de mogelijkheden om verhardingsconstructies en materiaalinnovatie en -ontwikkeling verder te optimaliseren. Aan de andere kant bemoeilijkt het niet delen van informatie de toepassing van kunstmatige intelligentie, het effectief en efficiënt zoeken en de ontwikkeling van

beslissingsondersteunende technologie voor weginfra. Deze aspecten vormen een ernstige belemmering voor de ontwikkeling van kennis en intelligentie in de weginfrasector [8].

Een Knowledge Graph voor de weginfrasector kan de gegevensintegratie van ongelijksoortige gegevensbronnen verbeteren en een uniform sectorbeeld van weginfragegevens creëren:

- Ontologieën binnen de knowledge graph kunnen helpen bij het verstrekken van duidelijke definities van entiteiten in datasets, waardoor de integratie en interoperabiliteit van gegevens worden verbeterd
- Ontologieën kunnen informatiespecificaties leveren ter ondersteuning van gegevensuitwisseling. Ook geautomatiseerde controles om datasets te testen aan de hand van de specificaties zijn mogelijk
- Datasets kunnen aan elkaar worden gekoppeld, waardoor de waarde van datasets toeneemt door experts toegang te geven tot meer gestructureerde data
- Datasets in Linked Data die verbonden zijn met dezelfde ontologieën worden interoperabel.

Een knowledge graph in Linked Data is de basistechnologie voor het beschrijven, koppelen en delen van informatie over wegenassets gedurende de volledige levenscyclus. Deze technologie is dan ook een ideale basis voor digital twinning van weginfra, ter ondersteuning van een datagedreven benadering van assetmanagement en ontwerptimalisaties, waaronder duurzamere wegennetwerken.

Een Knowledge Graph beschrijft entiteiten, attributen en hun relaties in de objectieve wereld in een relatief vaste en gestructureerde vorm, die kennis of informatie kan uitdrukken in een vorm die dicht bij de menselijke cognitieve wereld staat. Het biedt ook de mogelijkheid om grote hoeveelheden kennis en informatie over weginfrastructuur en verhandingen beter te organiseren, te beheren en te begrijpen. Met andere woorden, een knowledge graph is een soort kennisrepresentatie die de beschrijving van kennisentiteiten en de relatie tussen hen kan formaliseren. Daarom kunnen op basis van de kenmerken van een Knowledge Graph de functies van intelligent zoeken, redeneren en intelligent beantwoorden van vragen worden gerealiseerd.

Knowledge Graphs zijn een nieuwe technologie en basis voor systematische organisatie, diepteanalyse op basis van big data en intelligent beheer van kennissystemen voor weginfra-informatie. Ze hebben brede toepassingsmogelijkheden op het gebied van het ontwerp van verhandingsconstructies en voorspelling van materiaal en verhandingsprestaties in de weginfra [11].

6. Naar een Nederlandse weginfra Knowledge Graph

Zoals uitgelegd in de vorige paragrafen is het belangrijkste doel van een ontologie om het kader vast te stellen voor een Weginfra Knowledge Graph voor Nederland, WeginfraNL-KG. Deze knowledge graph is het kloppend hart van een gemeenschappelijk dataplatform voor weginfrastructuur in Nederland. De algemene doelstellingen van het gemeenschappelijk gegevensplatform zijn reeds beschreven in paragraaf 1.

De fundamentele (top-level) ontologie die wordt gebruikt is de Nederlandse Nationale Norm NEN 2660 "modellering van data in de gebouwde omgeving", die gelijk is aan de Europese norm EN 17632 "Semantic Modelling and Linking".

De scope van de eerste fase van de ontologie voor WeginfraNL-KG omvat concepten, relaties en attributen voor het registreren, structureren en delen van bouw- en assetmanagementgegevens (As Built-fase) tussen stakeholders (opdrachtgevers en aannemers) in de wegenbouw. Voor nadere analyse van wegenbouwprocessen voor het vaststellen van relevante datatypen en attributen voor deze fase wordt verwezen naar de paper 'Fundamental Framework for Functional Validation of Pavement Performance' [16]. De volgende fasen betreffen de toevoeging van een sub-ontologie voor sensordata, zoals meetgegevens over de toestand van de weg en laboratoriumonderzoeksgegevens, op basis van SOSA [12] en de implementatie van toestanden en gebeurtenissen voor het modelleren van de vier belangrijkste fasen van de levenscyclus van gegevens volgens de NEN 2660 en EN 17632 normen.

Eisen aan de ontologie voor WeginfraNL-KG

Niet-functionele eisen

- De ontologie moet relevante bestaande vocabulaires en ontologieën zoveel mogelijk naleven, integreren en hergebruiken
- De ontologie maakt een onderscheid tussen algemene concepten (ontologie op het hoogste niveau) en taak-/toepassing specifieke concepten
- De ontologie maakt specialisaties en uitbreidingen mogelijk
- De ontologie maakt diepe modellering mogelijk, waardoor metamodellen worden verbonden met instantiaties
- De ontologie moet open access mogelijk maken
- De ontologie moet gebaseerd zijn op FAIR-principes.

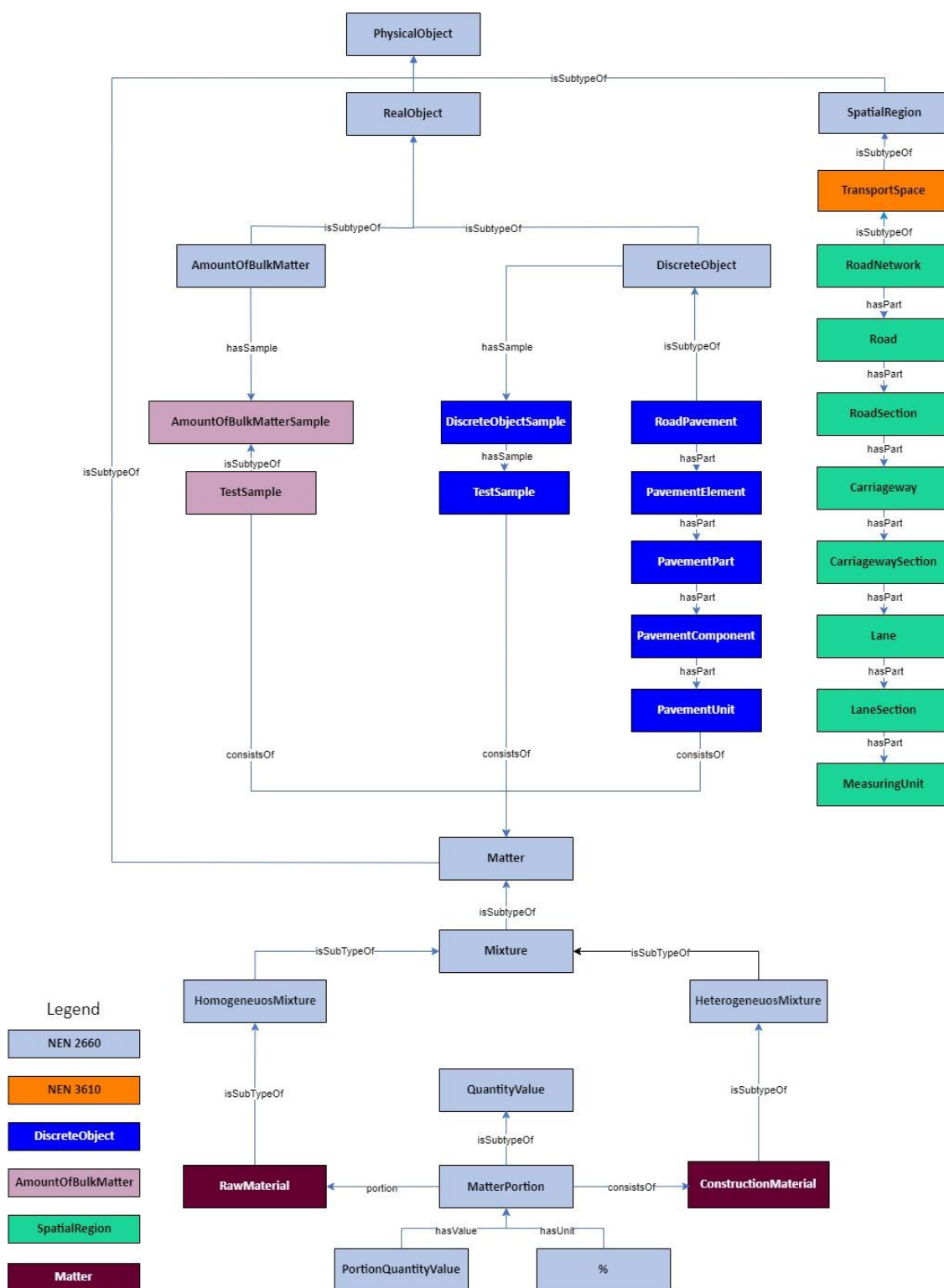
Technische eisen

- De ontologie wordt in een duidelijk gevisualiseerde vorm weergegeven en specificaties worden in een uniform gestructureerd (database)formaat opgesteld
- De ontologie is geïmplementeerd in SHACL
- De ontologie wordt automatisch gecontroleerd op consistentie en correctheid
- De ontologie kan worden bevraagd (bijvoorbeeld met behulp van SPARQL).

Referentie ontologie

De referentie-ontologie als basismodel voor een asfaltverharding is gevisualiseerd met DRAWIO (figuur 4) en de specificaties van de referentie-ontologie zijn verzameld in een eenvoudig gestructureerd format in EXCEL:

| Voorwerp/concept | Attribuut | Bron | Beschrijving/definitie |
|------------------|-------------------|------|---|
| Wegconstructie | Project-id | PIM | Unieke id van de wegconstructie |
| Wegconstructie | Contract-id | PIM | Identificatie van het contract voor de aanleg wegconstructie |
| Wegconstructie | Aanlegdatum | PIM | Datum van aanleg van de wegconstructie |
| Asfaltmengsel | Mengselcode | PIM | Mengcode van het asfaltmengsel voor communicatie in de praktijk |
| Asfaltmengsel | Streefdichtheid | PIM | Streefdichtheid van het asfaltmengsel |
| Asfaltmengsel | Holle ruimte | PIM | Gemeten holle ruimte van het asfaltmengsel |
| Asfaltmengsel | Watergevoeligheid | PIM | Gemeten watergevoeligheid van het asfaltmengsel |



Figuur 4: Referentieontologie als basismodel voor een asfaltverharding

Operationele ontologie

EXCEL is toegepast om domeinexperts te ondersteunen bij het creëren van de operationele ontologie voor Weginfra. Daartoe is een reeks tabellen gedefinieerd (EXCEL-sheets) en aan domeinexperts uitgelegd voor gebruik: tabellen voor het registreren van klassen, datatype-eigenschappen, relaties (objecttype-eigenschappen) en enumeratietypen. Het werkblad voor

de registratie van klassen, bijvoorbeeld, bevat onder andere kolommen zoals localname, label, description en parent.

| | A | B | C | D |
|---|-------------------|----------------------------------|-------------------|-------------|
| 1 | localname | label | parent | description |
| 2 | ConstructionLayer | Construction layer | PhysicalObject | |
| 3 | TopLayer | top layer of a road construction | ConstructionLayer | |
| 4 | AsphaltMixture | Asphalt Mixture | Material | |
| 5 | | | | |
| 6 | | | | |

Er is een converter gemaakt om deze tabellen om te zetten naar SHACL-gebaseerde Linked Data, wat resulteert in een taxonomie van klassen met labels en beschrijvingen. De bovenstaande tabel toont een deel van de beschreven taxonomie. Ook is een datatype property sheet geconstrueerd waar eenvoudige eigenschappen zijn gedefinieerd met een localname, label, beschrijving en datatype zoals string, decimaal, datum en ook een (QUDT) eenheid. Een andere sheet is opgezet om deze eigenschappen te verbinden met klassen. De schermafbeelding in figuur 5 laat zien dat een bepaalde klasse eigenschappen heeft die direct verbonden zijn en 'geërfde' eigenschappen heeft. Deze laatste zijn eigenschappen die verbonden zijn met klassen die parents zijn van de geselecteerde klasse. Dezelfde benadering is gevolgd voor relaties (objecteigenschappen in LD). Er is een werkblad gemaakt om de relaties te definiëren en een werkblad om ze aan een klasse te koppelen. Daarnaast is er een werkblad opgezet voor enumeratietypen waar vooraf gedefinieerde lijsteigenschappen zijn ingevoerd.

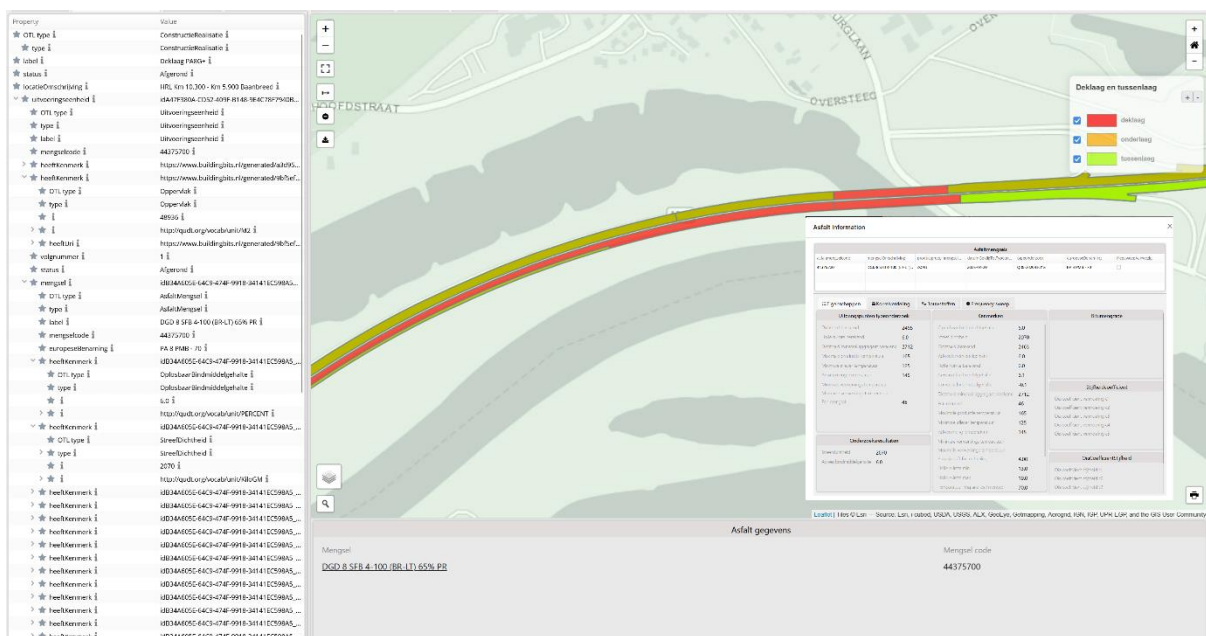
The screenshot displays the Wistor ontology viewer interface. On the left, a 'Taxonomy' panel shows a hierarchical tree of classes. The 'Wegconstructie' class is selected, and its sub-classes, including 'Wegconstructie wegtipe 1' through 'Wegconstructie wegtipe 7', are visible. The main panel on the right shows the 'General' tab for the 'Wegconstructie' class. It includes the URI 'https://weginfra.nl/otl/def#WegConstructie', a label 'Wegconstructie', and a definition: 'De gehele opbouw van een constructie voor wegverkeer vanaf de natuurlijke ondergrond tot aan het rijpervlak'. Below this, a 'Kernwoorden' (Keywords) table lists properties like 'Unieke identificatie van de wegconstructie' and 'Omschrijving van de wegconstructie'. The 'Uitgaande relaties' (Outgoing relations) table shows relationships to 'Constructie-element' and 'Uitsluitend object monster'. The 'Inkomende relaties' (Incoming relations) table shows relationships from 'Type wegconstructie' and 'Type weginfrastructuurobject'. The 'Enumeraties' (Enumerations) table lists 'Type wegconstructie' and 'Type discreet object'.

Figuur 5: Screenshot van de operationele ontologie viewer

7. Eerste toepassingen van knowledge graph apps voor weginfra

Een eerste reeks toepassingen die gebruik maakt van de knowledge graph-benadering is ontwikkeld. De eerste stap was het extraheren en converteren van gegevens uit verschillende systemen en databestanden. XML gebaseerde rapporten zijn via XSLT geconverteerd naar Linked Data, JSON API's zijn geconverteerd naar JSON-LD en SQL-systemen zijn geconverteerd via RML en hardcoded implementaties.

Omdat al deze data (ontologie en de 'projectinformatie') Linked Data zijn, was het eenvoudig om deze data in een triplestore (een linked data database) te plaatsen. Dit resulteerde in een database die kan worden bevraagd met behulp van de SPARQL-taal. Widgets van een applicatie kunnen SPARQL-query's gebruiken om de benodigde inhoud op te halen. Een SPARQL query kan gebruikt worden om alle "ConstructieRealisaties" (gerealiseerde verhardingsconstructies) op te halen met ruimtelijke informatie die ook als LD is opgeslagen met behulp van de Geo-SPARQL ontologie. Een andere widget kan alle eigenschappen en relaties van geselecteerde "Constructierealisaties" weergeven. In dit geval is de geselecteerde "ConstructieRealisatie" gekoppeld aan 'werk'realisaties die zijn aangesloten op Asfaltmengsels. De screenshot in figuur 6 laat zien dat de geselecteerde "ConstructieRealisatie" datatype eigenschappen heeft zoals wegoppervlakte, maar ook relaties met de weginfra ontologie en relaties met andere verbonden objecten zoals asfaltmengsel. Deze boom volgt de gegevensstructuur op de voet. Uiteraard is het heel eenvoudig om SPARQL te gebruiken om data op te halen en op een gebruiksvriendelijkere manier te presenteren. In deze applicatie wordt alle relevante informatie van een asfaltmengsel samengevoegd in één dialoogvenster.



Figuur 6: Screenshot WeginfraNL-KG van een gerealiseerde verhardingsconstructie

De bron van deze (gecombineerde) data is een SQL-gebaseerde applicatie (PIM) en een GIS-dataset. Via conversie wordt Linked Data uit dit systeem gehaald, waardoor de data veel begrijpelijker zijn dan de zeer complexe SQL-tabellen en de losgekoppelde GIS-dataset. Het

De Weginfra Ontologie kan worden gebruikt voor een (internationaal) platform waar weginfrastructuurgegevens kunnen worden verzameld, zodat de internationale sector of weginfrasector toegang heeft tot levenscyclusinformatie van verhardingen en weginfrastructuur.

De scope van de eerste fase van de ontologie voor WeginfraNL-KG omvat concepten, relaties en attributen voor het registreren, structureren en delen van bouw- en assetmanagementgegevens (As Built-fase) tussen stakeholders (opdrachtgevers en aannemers) in de wegenbouw. De volgende stap is het maken, opslaan en beheren van een Nederlandse nationale Knowledge Graph voor de weginfrasector, WeginfraNL-KG, en het ontwikkelen van de redeneer- en analysemodule om de Knowledge Graph in de praktijk toe te passen.

Een intrinsieke eigenschap van Linked Data is dat het kan worden uitgebreid. Dit betekent dat de ontologie van WeginfraNL-KG ook vrij eenvoudig kan worden uitgebreid. De volgende fasen in de ontwikkeling van WeginfraNL-KG zullen betrekking hebben op de toevoeging van een subontologie voor sensordata, zoals meetgegevens van de wegconditie en laboratoriumonderzoeksgegevens, op basis van SOSA en de implementatie van staten en gebeurtenissen voor het modelleren van de vier belangrijkste fasen van de levenscyclus van gegevens volgens de normen NEN 2660 en EN 17632.

De prestaties van de kleinschalige knowledge graphs in de gemeenschappelijke data-omgevingen van de pilotprojecten zijn uitstekend. Een van de belangrijkste onzekerheden voor grootschalige common data platforms is de praktische prestatie van grootschalige Knowledge Graphs in deze gemeenschappelijke dataplatforms. Dit aspect is één van de belangrijke onderzoeksthema's voor de implementatie van een gemeenschappelijke productieomgeving voor dataplatforms.

Aanbevelingen

Gebruik de Europese norm EN 17632 'Semantic Modelling and Linking (SML)' als de fundamentele (top-level) ontologie voor de ontwikkeling van ontologieën en rapporteer of publiceer praktijkervaringen met modelleerpatronen die niet (voldoende) in de normen zijn uitgewerkt.

Het wordt elke persoon, onderneming of organisatie die betrokken is bij de ontwikkeling van Knowledge Graphs aanbevolen om beschikbare referentie- of operationele ontologieën over te nemen of te hergebruiken of om op ontologieniveau te koppelen aan bestaande referentie- of operationele ontologieën. Op deze manier wordt het mogelijk om een zeer krachtige (inter)nationale Knowledge Graph voor verhardingen en weginfrastructuur te creëren door (inter)nationaal beschikbare data te koppelen en te implementeren.

Aangezien het bij het gebruik van knowledge graphs gaat om hergebruik van (sub)ontologieën en data, is het aan te bevelen om een internationaal platform op te zetten als open space om ontologieën en Knowledge Graphs voor de wegenindustrie te publiceren en te beheren en om succesvolle resultaten van pilotprojecten te tonen.

Het wordt aanbevolen om softwareontwikkelaars te stimuleren en te ondersteunen bij het creëren van effectieve en efficiënte tooling ter ondersteuning van de ontwikkeling, implementatie en het beheer van productieomgevingen voor knowledge graphs en common dataplatforms zoals de Nederlandse weginfrasector voor ogen heeft.

9. Literature

1. Hernando, D., Moins, B., Audenaert, A., Van den Bergh, W. (2024). Information and data sharing needs for pavements 4.0: From tendering to end of life. Bituminous Mixtures and Pavements VIII, 7
2. Sluer, B., Poeran, N. (2024). Roadmap Kenniskader Data & Informatie – Transitiepad Duurzame Wegverharding
3. Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5), 34-43
4. Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems (IJSWIS), 5(3), 1-22
5. Tan, J., Qiu, Q., Guo, W., Li, T. (2021). Research on the Construction of a Knowledge Graph and Knowledge Reasoning Model in the Field of Urban Traffic
6. Schevers, et al, (2007), Towards digital facility modelling for Sydney Opera House using IFC and semantic web technology, ICON (https://www.itcon.org/papers/2007_24.content.02481.pdf)
7. Lexi Rowland, Wouter Beek, Erwin Folder, Hans Schevers (2024-forthcoming), Building the Dutch National Land Registry in LD forthcoming, Semantic-Web-Journal
8. Almeida Falbo, de, R. (2014). SABiO: Systematic Approach for Building Ontologies
9. Ontotext. Knowledge Graphs in the Enterprise: The Story Behind the Hype – Manage Knowledge, Not data, White paper
10. Ontotext (2024). What Are Ontologies?, [What Are Ontologies? | Ontotext Fundamentals](#)
11. Yang, Z., Bi, Y., Wang, L., Cao, D., Li, R., Qianqian Li, Q. (2022). Development and application of a field knowledge graph and search engine for pavement engineering
12. W3C (2024). Sensor, Observation, Sample, and Actuator ontology, [Semantic Sensor Network Ontology \(w3.org\)](#)
13. Jean Paul Sebastian Piest, J., Benoiston Jales de Oliveira, V., Alencar Silva, de, P., Ricardo da Cunha Junior, M. (2024). A Domain Reference Ontology for Design Science Research Knowledge Bases
14. OMICRON Project (2024). Intelligent Road Asset Management Platform, [Omicron - Omicron project](#)
15. Brink, B. (2024). Artificiële Intelligentie voor levensduurverlenging, White paper TKI-Bouw.
16. Poeran, N., Sluer, B. (2024), Fundamental Framework for Functional Validation of Pavement Performance, paper Asphalt 4.0, Slovenië

10. Afkortingen

| | |
|------|---|
| API | Application Programming Interface, a software intermediary that allows two applications to talk to each other |
| FAIR | FAIR principles of findability, accessibility, interoperability, and reusability |

| | |
|--------|--|
| HTTP | Hyper Text Transfer Protocol, a method for encoding and transporting information between a client (such as a web browser) and a web server |
| JSON | JavaScript Object Notation, a lightweight format for storing and transporting data |
| LD | Linked Data |
| OWL | W3C Web Ontology Language |
| QUDT | Quantity, Unit, Dimension and Type. Base classes, properties, and restrictions used for modeling physical quantities, units of measure, and their dimensions |
| R2RDF | RDB to RDF Mapping Language |
| RDF | W3C Resource Description Framework, a standard for data interchange that is used for representing highly interconnected data |
| RML | RDF Mapping Language |
| SHACL | W3C Shapes Constraint Language |
| SKOS | W3C Simple Knowledge Organization System, a W3C recommendation designed for representation of thesauri , classification schemes , taxonomies , subject-heading systems , or any other type of structured controlled vocabulary |
| SOSA | Sensor, Observation, Sample, and Actuator ontology, a revised and expanded version of the initial SSN ontology |
| SPARQL | Standard query language and protocol for Linked Open Data and RDF databases |
| SQL | Structured query language, a standard language for database creation and manipulation |
| SSN | W3C Semantic Sensor Network |
| URI | Uniform Resource Identifier, a character sequence that helps identify a logical or physical resource connected to the internet |
| W3C | World Wide Web Consortium (W3C) for the development of standards and guidelines to help everyone build a web based on the principles of accessibility , internationalization , privacy and security |
| XML | eXtensible Markup Language, a software- and hardware-independent tool for storing and transporting data |
| XSLT | Extensible Stylesheet Language Transformations, a language originally designed for transforming XML documents into other XML documents or other formats such as HTML for web pages , plain text or XSL Formatting Objects , which may subsequently be converted to other formats, such as PDF , PostScript and PNG |