

Rapport

Technische oplossingen scan

UOI Unieke Object Identificatie & UOI-code-stelsel

Geonovum



Inhoudsopgave

1	Inleiding	3
1.1	Aanleiding & context	3
1.2	Opbouw van dit document	3
1.3	Werking van elk UOI-code mechanisme	3
2	Technische oplossingen UOI-code	4
2.1	Overzicht onderkende oplossingen	4
2.2	Identificatie met een UUID	4
2.3	Identificatie met een URI	4
2.4	Identificatie met blockchaincode	5
2.5	INSPIRE samengesteld code systeem	5
2.6	GTIN/GLN code systeem	7
2.7	RPUID-code systeem OGC	7
3	Technische oplossingen UOI-code-stelsel	8
3.1	Vinden van wegen naar de databronnen	8
3.2	Ontsluiten van de data uit databronnen	9
3.3	Semantisch duiden	10
Bijlage 1		14
	Samenstelling onderzoeksteam	14



Hoofdstuk 1

Inleiding

In dit hoofdstuk wordt de aanleiding en context voor de scan/inventarisatie naar technische oplossingen voor een UOI-code en UOI-code stelsel als onderdeel van de UOI-vervolgonderzoek beschreven. Ook wordt een korte leeswijzer via het schetsen van de opbouw van dit document gegeven.

1.1 Aanleiding & context

Gedurende de periode najaar 2020 en voorjaar 2021 hebben binnen het vervolgonderzoek UOI achtereenvolgens de volgende activiteiten plaatsgevonden:

- Een validatie van het eerste onderzoeksresultaat,
- Het opstellen van een Programma van Eisen voor een UOI-code & een UOI-code-stelsel
- Het inventariseren van technische oplossingen
- Het opstellen van een ontwerprapport voor de UOI-code en alternatieven voor een UOI-code stelsel
- Een consultatie van het PVE en het ontwerprapport

Dit rapport bevat de scan / inventarisatie naar technische oplossingen voor de UOI-code en een UOI-code stelsel.

1.2 Opbouw van dit document

In hoofdstuk 2 komen achtereenvolgens de resultaten van de inventarisatie naar technische oplossingen voor een UOI-code aan bod. Aan het einde van dit hoofdstuk geven we een beeld in welke mate deze oplossingen voldoen aan de in het Programma van Eisen geformuleerde eisen aan een UOI-code.

In hoofdstuk 3 komen achtereenvolgens de resultaten van de inventarisatie naar technische oplossingen voor een UOI-code-stelsel aan bod. Aan het einde van dit hoofdstuk geven we een beeld in welke mate deze oplossingen voldoen aan de in het Programma van Eisen geformuleerde eisen aan een UOI-code-stelsel.

1.3 Werking van elk UOI-code mechanisme

Zoals in de eerdere UOI-documenten reeds is betoogd, bestaat een UOI-code stelsel functioneel uit meerdere onderdelen die samen het beoogde effect (*altijd kunnen blijven vinden van gegevens over een object en de objectrelaties tussen registraties*) kunnen realiseren. Dat zijn achtereenvolgens de volgende onderdelen:

- A. Uniek identificeren van het object
- B. Vinden van de wegen naar de databronnen over het object
- C. Ontsluiten van de data uit de databronnen
- D. Semantisch kunnen duiden van de gevonden data

Technische oplossingen voor een UOI-code stelsel bestaan uit een combinatie van oplossingen voor deze vier onderdelen. We beschrijven de gevonden technische oplossingen voor deze onderdelen



Hoofdstuk 2

Technische oplossingen UOI-code

In dit hoofdstuk worden de gevonden technische oplossingen voor de UOI-code beschreven.

2.1 Overzicht onderkende oplossingen

In de onderstaande tabel zijn de onderkende technische oplossingen voor een UOI-code weergegeven.

Paragraaf	Oplossing	Korte toelichting
2.2	UUID	Internationale methode voor genereren van unieke codes
2.3	URI	Internationale methode voor aanduiden van bronnen
2.4	Blockchain	Internationale methode voor genereren van unieke HASH
2.5	INSPIRE	Europees samengesteld code-systeem (NEN 3610 volgt deze)
2.6	GTIN	Internationale methode voor de uitgifte van product-ID als barcode
2.7	RPUID	Amerikaanse methode voor aanduiden van Geo-objecten

Daarnaast is gekeken naar de analogie in de identificatie van personen [BSN-nummer](#), generieke objectidentificatie binnen de ICT-wereld via de [OID-code](#) en roerende zaken zoals boeken [ISBN-nummer](#) en [DOI-code](#). Deze systemen voor identificatie zijn niet gericht op objecten in de gebouwde omgeving met een locatie en om die reden worden deze hier niet nader beschreven.

2.2 Identificatie met een UUID

De eerste technische oplossing voor het genereren van een unieke object identificatie is de UUID-code generatie. De UUID-code staat voor [Universally Unique Identifier](#). De UUID ook wel GUID genoemd is een internationale methode om unieke persistente identificatoren te genereren. Deze methode wordt al lang gebruikt. Hoewel de methode geen 100% garantie biedt dat er altijd een unieke code wordt gegenereerd, is de kans op creatie van twee dezelfde codes zeer klein. Een UUID is een getal van 128 bit en wordt veelal geschreven in hexadecimale vorm. De GUID-code is een implementatie van Microsoft van de UUID-code. GUID staat voor [Globally Unique Identification](#)

2.3 Identificatie met een URI

De tweede technische oplossing voor het genereren van een unieke UOI-code is de URI-generatie. Een URI is een [Unique Resource Identifier](#). Geonovum heeft een aanzet gemaakt naar een [Nationale URI-strategie voor Linked data van de Nederlandse Overheid](#). Uniform Resource Identifiers (URI's) zijn de gestandaardiseerde manier om op het internet dingen (pagina's met informatie, objecten, datasets) uniek te identificeren. Dit kan door middel van een Uniform Resource Name (URN) of met een Uniform Resource Locator (URL). Bij een URN krijgt een object een unieke naam toebedeeld. Bij een URL krijgt een object een adres (een `https://URI`) waarmee het op internet kan worden gevonden (de ons welbekende website adressen). Door URI's te gebruiken, kun je op een unieke manier naar een object verwijzen en ze daardoor uniek onderscheiden. Het internet protocol zorgt er voor dat gepubliceerde URL's uniek zijn. Dit garandeert unieke UOI-codes. Het advies met de huidige stand van zaken is om stabiele `https://URI`'s te gebruiken voor geo-objecten en datasets.

Binnen INSPIRE en NEN3610 is afgesproken om op elk object op basis van de bestaande interne identifier een URI te geven, waarmee binnen de INSPIRE en NEN3610 context het object uniek te identificeren én op



te vragen is. Lees ook de aanzet tot een [nationale URI-Strategie en de beschrijvingen van de uitwerkingen daarvan op <https://www.geonovum.nl/themas/uri-strategie>](#)

en

https://www.geonovum.nl/uploads/documents/D1-2013-09-19_Towards_a_NL_URI_Strategy.pdf

Het is denkbaar met een URI-strategie een unieke UIO-code decentraal samen te stellen door een combinatie van elementen te kiezen die samen de URI vormen. Denk aan <bron> <domein> <localID>. De combinatie van <bron> en <domein> wordt in dit verband ook wel als de <namespace> aangeduid.

Dat zou in geval de BAG dan een UOI-code kunnen worden

.

Bij de BAG bestaat de objectnummering uit:

- gemeentecode (4 posities)
- objecttypecode (2 pos.)
- objectvolgnummer (10 pos.)

Op deze wijze ziet de identificatie van een Pand (objecttypecode 10) dat is ontstaan in de gemeente Amersfoort (gemeentecode 0307) en toen objectvolgnummer 367968 heeft gekregen, eruit als: 030710000367968.

2.4 Identificatie met blockchaincode

Een derde oplossingsrichting maakt gebruik van zogeheten [blockchain-technologie](#). Met blockchain-technologie wordt een unieke UOI-code middels een zogeheten HASH gegenereerd. Het voordeel van deze technologie is dat geen tweede gelijke code kan worden uitgegeven omdat het blockchain-ontwerp zorg draag voor een altijd unieke HASH. De blockchain HASH wordt namelijk samengesteld uit de aaneengeschaalde bron vermeldingen in gedecentraliseerde gecontroleerde en digitaal gecertificeerde blockchain-register (databank). Bij een dergelijke UOI-code generatie wordt via een API door een partij een nieuwe UOI-code (HASH) aangevraagd en verstrekt. Deze wordt vervolgens in de eigen registratie aan het desbetreffende object aanvullend vermeld. In de blockchain (gedecentraliseerd federatief werkende databank) wordt deze code ook verwerkt overeenkomstig de blockchain principes.

Voorbeelden van blockchain gebaseerde oplossingen in het domein van de gebouwde omgeving zijn:

- De weg die een bouwproduct volgt vastleggen (bouwtoevoerlogistiek) [Tradelens](#) / [Providence](#)
- Kredietwaardigheid bij aankoop en financiering van een bouwwerk [Muxe.online](#)
- Materialenpaspoort (zekerheid bij achterhalen van materialen in gebruikte bouwmaterialen) ([blockmaterials](#))
- Marktplaats bouwmaterialen ([Excesmaterialexchange](#))
- Renovatiepaspoort (zekerheid bij achterhalen bouwonderdelen in bouwwerken) ([Recheck](#))

2.5 INSPIRE samengesteld code systeem

Naast de drie generieke code generatiesystemen zijn er ook op nationaal en Europees niveau samengestelde code systemen in gebruik. Ook hier is het doel een unieke objectcode te verstrekken. We benoemen hier de [INSPIRE-aanpak](#). De Nederlandse geo-standaard NEN 3610 volgt dezelfde methode die ook wordt toegepast in een aantal Nederlandse geo-(basis)registraties.

Volgens de richtlijnen van INSPIRE moet elk object (feature) voorzien worden van een unieke identificatie: de [InspireId](#). Belangrijk hierbij is, dat de identificator **uniek** en **persistent** moet zijn. Persistent betekent, dat de identificator gedurende de levensduur van het object ongewijzigd moet blijven. Een *InspireId* bestaat uit de volgende onderdelen:



- Namespace (*verplicht*)
- LocalId (*verplicht*)
- VersionId (*optioneel*)

Namespaces

Een *Namespace* heeft als belangrijkste doel, dat het de ruimte definieert waar binnen de *localId* uniek is. Om te voorkomen, dat verschillende dataproviders dezelfde *Namespace* gebruiken, en om consistentie te verkrijgen, is er in Nederland (en Europa) een zogenaamd *Namespace Register* opgezet. De *Namespace* wordt geregistreerd in het [nationale namespace register](#). De verantwoordelijkheid voor uniciteit van de namespace ligt bij het namespace register.

LocalId

Hoe maak je unieke id's voor INSPIRE ruimtelijke objecten aan?

Elk ruimtelijk object van een dataset (of dataset series) dat via INSPIRE toegankelijk wordt gemaakt, moet voorzien worden van een unieke identificatie zodat derden ondubbelzinnig aan het object kunnen refereren. Dataproviders dienen dus – buiten eventuele eigen ID's die de organisatie gebruikt – de ruimtelijke objecten van een unieke INSPIRE identificatie te voorzien.

In hoofdstuk 14 ('Identifier management') van het [Generic Conceptual Model](#) wordt algemene uitleg gegeven over het gebruik van unieke identifiers. Per thema kunnen er nog specifieke uitbreidingen staan in **paragraaf 5.2.1.4** van de data specificatie. Binnen INSPIRE datasets en dataset series heet de unieke code het **InspireId**. De dataprovider is verantwoordelijk voor uniciteit van locale id's binnen zijn eigen context. In de internationale context van INSPIRE levert de combinatie van namespace (die de dataset identificeert) en localId die binnen de dataset uniek is) een identifier op die uniek is binnen INSPIRE.

VersionId

Een InspireId kan ook een (optioneel) versionId hebben. Dit versienummer kan gebruikt worden als er van hetzelfde object op verschillende tijdstippen verschillende versies in omloop zijn die van elkaar onderscheiden moeten worden

De NEN 3610 versie hiervan wordt toegepast in een aantal Nederlandse geo-registraties maar ook binnen het PDOK-, en KLIC-platform

Voorbeeld

Deze is van administrative boundaries, (bron PDOK) met als voorbeeld Ede:

Administrative boundary

Namespace + localid = feature_id

Naam	Waarde
gid	193
namespace	NL.BRK.AU
localid	0228
nationalcode	0228
nationallevel	3rdOrder
nationallevelname	Municipality
name	Ede
feature_id	NL.BRK.AU.0228



2.6 GTIN/GLN code systeem

Iedereen is wel bekend met de EAN-streepjescode op producten zoals we die in de supermarkt aantreffen. Deze methode is inmiddels ook uitgebreid naar vele andere sectoren en wordt door de [GS1-organisatie](#) breed gefaciliteerd. Het bijzondere van deze GTIN-code is dat deze is opgebouwd uit een aantal onderdelen. De zo opgebouwde productcode wordt formeel ingeschreven in het [GEPiR¹-register](#) en kan door een ieder worden geverifieerd. [EAN/GTIN-code](#) is opgebouwd uit drie onderdelen <organisatiecode> <artikelcode> <controlecijfer> en bij ons breed bekend als de barcode. Ook locaties zijn identificeerbaar met de GTIN in de vorm van een GLN [Global Location Number](#) van 13 posities voor fysieke en digitale locatie resp. organisatie. Het basisprincipe is dat de uitgifte van de codes door de maker van de GTIN-code zelf wordt gedaan en bijgehouden. Daarbij worden de organisatiecodes centraal uitgegeven en geregistreerd.

2.7 RPUID-code systeem OGC

In het voorjaar van 2021 heeft het [OGC Open GIS Consortium](#) aangekondigd een *Unique Geospatial Real Property Identifier (RPUID)* te gaan vormgeven. Het lijkt een toepassing om verzekeraars in de Verenigde Staten een unieke identificatie van verzekerde objecten te bieden. Tegelijk lijkt de RPUID al 15 jaar eerder voor de [Departement of Defence DoD in het RPI initiatief](#) en bij enkele staten voor [belastingheffing](#) op onroerend goed geïmplementeerd te zijn. Ze is vastgelegd in de [USACE ECB 2013-24 Directive](#).

¹ GEPiR = Global Electronic Party Information Registry



Hoofdstuk 3

Technische oplossingen UOI-code-stelsel

In dit hoofdstuk worden de gevonden technische oplossingen voor de UOI-code-stelsel beschreven..

3.1 Vinden van wegen naar de databronnen

Binnen de context van het UOI-code-stelsel zullen de objecten behorend bij de UOI codes in verschillende registraties opgeslagen zijn. Het vinden van de weg naar de databron (dataset) kent twee niveaus:

1. Het vinden van de registratie waar de data zich bevindt
2. Het binnen de registratie vinden van de specifieke data van het informatie-object

Vinden van een specifieke registratie

Door het toekennen van vindbaarheids-metadata en publiceren daarvan worden datasets vindbaar. Voorbeelden hiervan zijn bijvoorbeeld het nationaal georegister voor geodata (<https://www.nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/home>), zoals te vinden op PDOK, of het dataregister van de Nederlandse overheid (<https://data.overheid.nl/>).

Verondersteld wordt dat metadata over databron gepubliceerd zijn. In dat geval kun je een webcrawler inzetten om deze bronnen (datasets) te vinden.

Vinden van het specifieke informatieobject

Daar waar de weg naar de databronnen veelvuldig gebruikt , wordt, hebben de bronhouders van die registraties vaak koppelentiteiten benoemd waarmee de informatieobjecten direct aan elkaar gerelateerd kunnen worden. Dat kunnen 1:1 tot N:M relaties zijn. Veel gebruikte koppelentiteiten in het geo-domein zijn het adres, pand, verblijfsobject uit de BAG (via de [BAG-ID](#)) en de basistopografie uit de BGT (via de [BGT-ID](#)). Deze dekken evenwel op dit moment niet alle objecten in de gebouwde omgeving. In situaties waar er een koppelentiteit is vastgelegd, is de weg naar de domeinregistratie bekend. Deze is te vinden in op het internet dan wel in catalogi zoals [PDOK](#).

Er zijn ook andere manieren denkbaar, waarmee de weg naar de databronnen te vinden is. We geven een kort overzicht zonder te detailleren. Er zijn daarbij twee voorkomende manieren van benaderen:

- A. De identificatie van het object is bekend <zoeken naar waar gegevens over het object met de UOI-code gevonden kunnen worden>
- B. De identificatie van het object is onbekend <zoeken naar de relevante UOI-code en daarna stap A>

Ad A. De identificatie van het object is bekend (resolver)

Wanneer de UOI-code bekend is wordt het centrale UOI-code-register of een federatief stelsel van aangesloten registraties bevraagd om vast te stellen welke bron deze gegevens heeft. Het centrale register kan daarbij naar de bron verwijzen of deze gegevens zelf bevatten en verstrekken. In het andere geval wordt na het verkrijgen van de bronlocatie de bron geraadpleegd en zo de gegevens verkregen. Let wel het gaat hier om het vinden dan de gegevens van een informatie-object dat gegevens over het object uit de gebouwde omgeving bevat. Over een object in de gebouwde omgeving zullen in meerdere registraties informatieobjecten bestaan. Zij zullen allemaal verschillende UOI-codes bezitten. De UOI-code wordt immers aan het informatieobject gehangen. Voor domein-overstijgende vragen moet je dus meerdere UOI-codes kennen die bij elkaar horen omdat ze verwijzen naar informatie-objecten die elk iets weten over een



fenomeen in de gebouwde omgeving. Zie in dit verband ook de definitie van een [resolver](https://identificer.overheid.nl) op <https://identificer.overheid.nl>.

Ad B. De identificatie van het object is onbekend (bevragen registraties)

In dat geval is de vrager op zoek naar objecten op basis van eigenschappen van die objecten die in de registratie zijn vastgelegd. Meestal gaat de vrager opzoek naar de locatie van het gezochte object in de gebouwde omgeving en kiest dat als selectiemechanisme. Als bekend is welk soort informatie-object gezocht wordt kan deze geografische selectie in de desbetreffende gegevensverzameling worden uitgevoerd, mits deze zo toegankelijk is gemaakt. Voor domein-overstijgende vragen moet de geografische selectie herhaald worden in elk van de relevante domeinregistraties.

Wanneer een geografische selectie niet mogelijk is omdat het een informatie-object betreft dat nog geen geo-referentie heeft (een generiek ontwerp voor een woning: een ontwerp dat meerdere keren wordt gebruikt), zal een administratieve koppeling/selectie uitkomst moeten bieden. Dit vraagt om een mogelijkheid van relateren vooraf dan wel op het moment van bevragen. Dat kan bijvoorbeeld een lokale project-identificator of ontwerpnummer zijn.

Zoals zojuist onder B is beschreven worden relaties tussen instanties van informatieobjecten gezocht ofwel via een koppelentiteit dan wel via locatie.

3.2 Ontsluiten van de data uit databronnen

Het ontsluiten van data uit een databron kan verricht worden via het bevragen van de bron via een REST-API, SOAP of een webservice. Daarop kan een informatiebeveiligings-regime (autorisatielaag) gelden dat de veilige geautoriseerde toegang verzekert. Deze dataportalen zijn er in alle soorten en maten. Deze zijn zowel toegewijd aan een domein of speciale verzameling van sectorale registraties als in de vorm van generieke datahubs. Deze kunnen zowel een niet-commercieel als commercieel dan wel een overheidskarakter hebben. In onderstaande tabel zijn een aantal voorbeelden opgenomen. We verwijzen hier tevens naar het [Programma Regie op Gegevens](#) en de [Data Sharing Coalition](#) alwaar over dit soort platforms nader onderzoek is verricht. Linked data is ook een techniek voor een API waarbij men met SPARQL de API beschrijft.

Naam Platform	Focusgebied	Karakter
IShare	Logistiek	Datadeelplatform
Data Sharing Coalition	Afspraken voor Datadelen	Afsprakenstelsel
PDOK	Ruimtelijke data	Geoplatform INSPIRE
Ruimtelijke Plannen.nl	Ruimtelijke wetten en regels	Ontsluiten ruimtelijke plandata
DSO.NL	Ruimtelijke wetten en regels	Ontsluiten ruimtelijke plandata
KLIC.NL	ondergrondse infrastructuren	Ontsluiten data over ondergrondse infrastructuren
i4Sociaal	Persoonsgebonden gegevens uit sociale domein	Persoonsgericht raadplegen
Medmij	Persoonsgebonden gegevens uit medische domein	Persoonsgericht raadplegen
Edumij	Persoonsgebonden gegevens uit onderwijs domein	Persoonsgericht raadplegen

NB. Bij al deze platformen is digitaal bevraging als uitgangspunt gehanteerd.



3.3 Semantisch duiden

Zonder te pretenderen volledig te zijn beperken we ons hier tot het benoemen van de belangrijkste principes en technieken. We verwijzen naar de desbetreffende plekken waar mee nadere informatie gevonden kan worden. De centrale boodschap is en blijft dat vandaag de dag via een combinatie van [linked data](#), [knowledge graphs](#) (kennisgraaf) en semantische zoektaken zoals [SPARQL linksets](#) kunnen verkregen worden, die semantisch interpreteerbare relaties tussen informatieobjecten kunnen vastleggen. Met deze linksets en onderliggende semantische (informatie-) modellen ([ontologie](#)) kunnen domein-overstijgende vragen beantwoord worden.

We schetsen kort de technieken achter:

- Linked data
- Ontologies
- Knowledge graphs (kennisgraaf)
- Linksets

Linked data

Linked data is een techniek om gegevens uit verschillende bronnen op basis van kenmerken aan elkaar te relateren en de uitkomst van deze activiteit vast te leggen in zogeheten linksets. Linked data maakt daarbij onder andere gebruik van de verwantschappen (relaties op model niveau) die entiteiten in registraties op semantisch niveau met elkaar hebben.

Linked data is een aanpak om bestaande registraties met unieke identificaties te uniformeren door er een gestandaardiseerd geautomatiseerde machineleesbare interface omheen te bouwen. Je ontsluit de bestaande registraties als Linked-Data (https://en.wikipedia.org/wiki/Linked_data). Hierdoor ontstaat als het ware een infrastructuur om databronnen te ontsluiten.

De vier basisprincipes van Linked-Data zijn:

1. "Use URIs as names for things". (Deze URIs zijn te construeren met als basis al bekende identificatie in de lokale registratie).
2. "Use HTTP URIs so that people can look up those names"
3. "When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)"
4. "Include links to other URIs, so that they can discover more things"

We lichten de stappen die de Linked-Data-aanpak volgt nader toe.

1. Unieke identificatie. (Er wordt gebruik gemaakt van een unieke URI (Unique Resource Indicator)
2. Vindbaar en toegankelijk maken van de data (datasets zijn gepubliceerd en in catalogi opgenomen, API's geven toegang tot de data).
3. Afspraken maken over de begrippen (vocabulaire) en betekenis van entiteiten. Daarmee krijgen de gegevens geïdentificeerd door de UOI-code betekenis (semantiek)
4. Opnemen van verwijzingen zorgt voor een samenhangend geheel waarin objecten door de registraties (en dus door de levenscyclus heen) gevolgd kunnen worden.

Ontologie

"In de [informatica](#) en de [logica](#) is een **ontologie** het product van een poging een uitputtend en strikt [conceptueel schema](#) te formuleren over een bepaald domein. Een ontologie is typisch een datastructuur die alle relevante [entiteiten](#) en hun onderlinge [relaties](#) en regels binnen dat domein bevat, zoals bij een **domeinontologie** het geval is. Op het terrein van de [kunstmatige intelligentie](#) wordt het begrip ontologie gebruikt als aanduiding voor een door computers interpreteerbare beschrijving van de werkelijkheid (kennisrepresentatie). In het [semantisch web](#) wordt dezelfde aanduiding gebruikt. Binnen het semantisch web moet een computer de betekenis van tekst en metadata kunnen afleiden en op basis van die betekenis kunnen redeneren en gevolgtrekkingen maken.



Een ontologie wordt wel gedefinieerd als: 'Een strikt en uitputtend schema voor een bepaald onderwerpsdomein, meestal in een hiërarchische structuur, die alle relevante grootheden en hun relaties bevat, alsmede de regels waaraan die grootheden en relaties binnen dat domein voldoen.'^[1] Het gebruik van het woord binnen de informatica is afgeleid van het woord [ontologie binnen de filosofie](#).

Een ontologie verschilt van een [databank](#) doordat een ontologie niet alleen feiten bevat maar ook regels, gevat in logische formules. Uit dergelijke regels kan men nieuwe feiten afleiden met een [automatisch redeneerprogramma](#).

Een ontologie die niet verbonden is aan een specifiek domein, maar dat algemene entiteiten probeert te beschrijven, wordt een [foundation ontology, top level ontology of upper ontology](#) genoemd. Een meer gespecialiseerde tussenvorm wordt een [core ontology](#) genoemd en probeert een specifiek maar breed toepassingsgebied te beschrijven. Over het algemeen dienen gespecialiseerdere en domeinspecifieke schema's te worden gecreëerd om de [data](#) geschikt te maken voor beslissingen in de werkelijkheid." (Bron Wikipedia)

Wij spreken in dit onderzoek van verwantschappen wanneer we het hebben over de relaties tussen informatiemodellen op model niveau (deze zijn domein-overstijgend) binnen de gebouwde omgeving. Deze verwantschappen zijn niet te verwarren met zogeheten [upper of foundation ontologies](#) (die op topniveau in abstracte termen generieke begrippen met elkaar gebruiken en in principe domein onafhankelijk zijn).

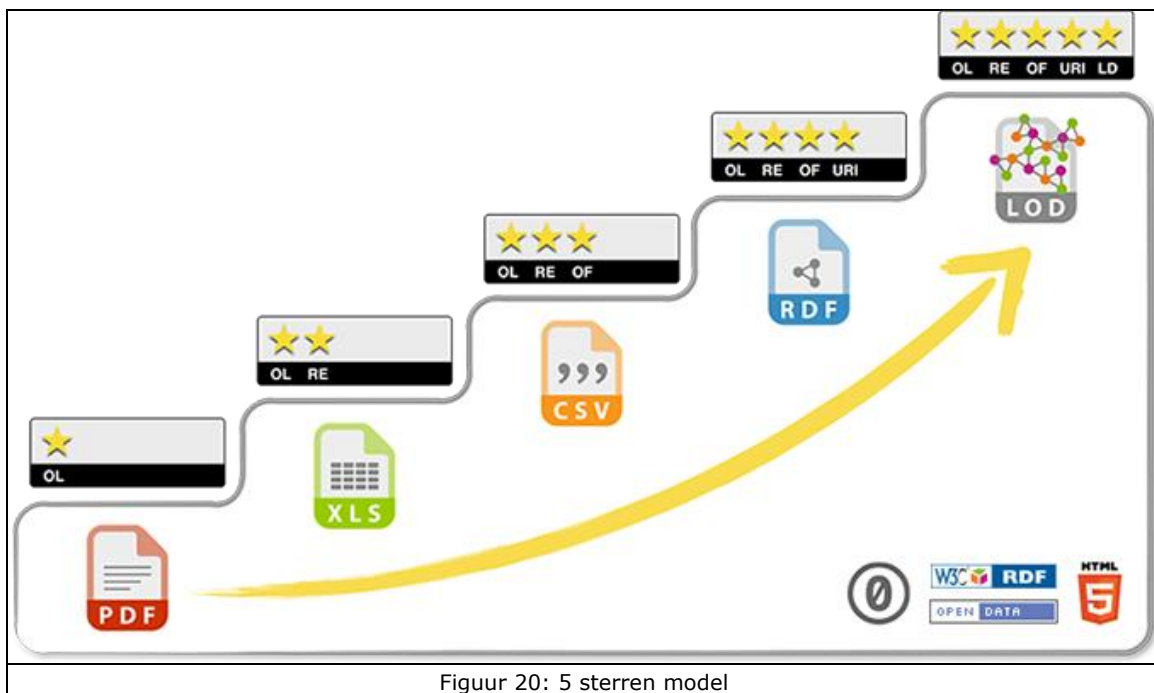
Feitelijk is de verwantschapsontologie een soort kern-ontologie op het niveau van de gebouwde omgeving en wel zo dat de concepten (entiteiten) uit onderliggende domein-ontologieën onderling begrepen kunnen worden. Een informatie-model is een domein-ontologie.

Kennisgraaf

De kennisgraaf vertegenwoordigt een verzameling onderling verbonden beschrijvingen van entiteiten - objecten, gebeurtenissen of concepten. Kennisgrafieken plaatsen gegevens in context via links en semantische metadata en bieden op deze manier een basis voor gegevensintegratie, unificatie, analyse en delen van data.

Een kennisgraaf wordt veelal gerepresenteerd in RDF, ook wel bekend als de standaard voor Linked Data. In principe is dan ook een dataset die beschikbaar is als Linked Data (bv. de BAG) ook al een (mini) kennisgraaf. In de communicatie noemen we Linked Data nu veelal kennisgraaf (Knowledge Graph), omdat beter aansluit bij de terminologie die gehanteerd worden door o.a. Gartner. Toch reserveren we de term kennisgraaf vaak wel voor een dataset overstijgend karakter. Ook wordt een kennisgraaf vaak opgesteld voor een specifiek doel / use case. Terwijl we Linked Data in de communicatie meer gebruiken voor het aanbod perspectief. Een voorbeeld hiervan: "Kadaster publiceert de BAG als Linked Data. En samen met de BRT en BGT kan de gebruiker hem toepassen in de Leefomgeving Knowledge Graph." De onderliggende technologie is gelijk.

De URI is het belangrijkste concept dat universele linking tussen entiteiten mogelijk maakt. **Een kennisgraaf wordt daarom ook wel uitgelegd als de 5^e (maximale) ster in het (open) data herbruikbaarheidsmodel**, waarbij de 5^e ster wordt toegekend nadat URIs met elkaar verbonden zijn, en wordt dan ook gevisualiseerd met een plaatje dat een graaf voorstelt. In de praktijk is het belangrijk dat de regels voor het toekennen en omgaan (zoals persistentie) goed worden nageleefd, en dat tevens zaken zoals een raadpleegservice voor URI's zijn ingeregeld.



Figuur 20: 5 sterren model

Alle entiteiten in een dataset hebben URI's, zowel op instantie als op model niveau; het linken in een Knowledge Graph kan dan ook op verschillende niveaus. Vaak worden concepten uit data modellen aan elkaar gerelateerd. Dit gebeurt wanneer verschillende werelden met elkaar verbonden moeten worden om gewoon om ergens te beginnen. Neem bijvoorbeeld een ontologie uit de bouw wereld, en een ontologie uit de geo-wereld, beide bevatten noties over plaatsen/locaties en panden/gebouwen, dan zijn deze twee ontologieën via een bovenliggende (upper) ontologie aan elkaar te relateren. Dit wordt veel toegepast ondanks dat er ook wel kwaliteitsproblemen aan deze aanpak kleven; de zogenaamde "same as" problematiek.

Ook al hebben nu we nu extra kennis, we weten nog steeds niet of pand88 hetzelfde is als gebouw99 in de andere dataset. Daarvoor zullen we ook relaties op instantie niveau moeten vastleggen. Dit klinkt misschien complex, maar in de praktijk is dit vaak te automatiseren op basis van 'business rules'; en kan vastgelegd worden in 'triples'. Deze kunnen opgeslagen worden in 1 van de 2 te linken datasets, of in een losstaande linkset. Regime (inhoudelijke regels) hiervan is een belangrijk aspect daardoor heeft opname in 1 van de bestaande datasets, die al een regime heeft ingeregeld, voordelen.

Als voorbeeld: in eerste instantie waren de BAG, BGT en BRT niet aan elkaar gerelateerd. Vervolgens werden deze los staande silo's als Linked Data gepubliceerd. Een soort van losstaande mini kennisgrafen. Vervolgens werden er losstaande linksets gecreëerd en niet-officieel gepubliceerd, zodat toch een grote kennisgraaf ontstond, maar er geen regime was op de linksets. Inmiddels zitten we in een vervolg stadium (mede dankzij ontwikkelingen in de SOR) waarin de links direct zijn opgenomen in de basisregistraties, waardoor de linksets overbodig zijn geworden, en direct onder de regimes van de basisregistraties valt. Het ligt in de lijn der verwachting dat deze trend zich doorzet en minimaal het stelsel van geo-basisregistraties op deze manier als kennisgraaf beschikbaar komt, en waarschijnlijk nog veel breder: van geo-basisregistraties, naar geo-datasets, naar locatie gebaseerde data.

Geodata kan ook een impliciete link hebben op basis van de locatie. In de praktijk worden deze vaak geëxpliciteerd door deze relaties administratief vast te leggen.



Dit is eenvoudig te doen door middel van een [GEOSPARQL](#) query uit te voeren en het resultaat op te slaan in 'triples' (data). Er kunnen wel kwaliteitsissues van toepassing zijn als locaties/ruimtelijke volumes niet overeenkomen.

De [kennisgraaf](#) (model + instanties) is op zichzelf ook weer vergelijkbaar met een dataset, en kan ontsloten worden op de bekende manieren die ook voor datasets gelden. Maar daarnaast zijn er aanvullende mogelijkheden zoals redenering over data, inclusief het afleiden van nieuwe feiten die niet rechtstreeks in de data staan.

<https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph/>

Linksets

Een [linkset](#) is een verzameling RDF links tussen twee datasets. Het zijn dus koppelparen in een specifiek RDF-formaat (triples) gegoten om zo het zoeken via SPARQL op deze linksets mogelijk te maken. Een voorbeeld waar Linked Data via linksets van Nederlands Geo-registraties bevraagd wordt, is te vinden op: <https://www.pdok.nl/sparqlviewer>



Bijlage 1

Onderstaand vindt u samenstelling van het onderzoeksteam.

Samenstelling onderzoeksteam

Naam	Rol / Organisatie
Eric van Capelleveen	Projectleider / procesadviseur (Geonovum)
Paul Janssen	Modelleringsdeskundige (Geonovum)
Wilko Quak	Modelleringsdeskundige (Geonovum)
Frank Kooij	Deskundige basisregistraties (kadaster)
Erwin Folmer	Deskundige Linked Data (kadaster)