

URI-beleid voor Linked Data

v 1.1.2

Datum:	21 april 2017
Auteur:	Kirsten Hulsker, Roosmarijn de Groot, Enno Meijers, René Voorburg
Versie:	1.1.2
Status:	definitief - ter accordering

Distributie:

Versie	Datum	Naam:	Functie:
vóór 0.9	-	Ruben Verborgh	extern adviseur, iMinds
vóór 0.9	-	Roland Cornelissen	extern adviseur, Metamatter
0.9.0	21 juni 2016	Roland Cornelissen	extern adviseur, Metamatter
0.9.1	24 juni 2016	Inge Hofsink, Meta van der Waal-Gentenaar, Lian Wintermans, Martijn van der Kaaij	PS / Metadatabeheer
0.9.1	27 juni 2016	Jeffrey van der Hoeven, Steven Claeysens	business analist, dataservices
0.9.1	27 juni 2016	Juliette Lonij, Theo van Veen, Willem Jan Faber	I&O Onderzoek
1.0		Karin Lodder	opdrachtgever
1.1	20 maart 2017	Inge Hofsink, Meta van der Waal-Gentenaar, Marieke Moolenaar, Martijn van der Kaaij, Miriam Wagenaar, Dennis Eijsten, Steven Claeysens, Lian Wintermans, Juliette Lonij, Theo van Veen	
1.1.2	21 april 2017	Karin Lodder	opdrachtgever

Versiegeschiedenis:

Versie	Datum	Wijzingen:
0.9.1	24 juni 2016	diverse, mn. taalkundig
0.9.2	7 juli 2016	feedback Jeffrey van der Hoeven verwerkt, mn. verduidelijkingen in de tekst
1.0	14 juli 2016	feedback Theo van Veen, Juliette Lonij, Meta van der Waal-Gentenaar, Inge Hofsink en Martijn van der Kaaij verwerkt, mn tav opvragen van serialisaties dmv URL (-parameters)
1.0	oktober 2016	accordering door MT P&B
1.1.0	17 maart 2017	<p>Wijzigingen betreffende de <i>vorm</i> van de URI om beter aan te sluiten bij partijen in netwerk:</p> <ul style="list-style-type: none"> • wijziging basisdomein naar data.bibliotheek.nl • wijziging vorm tav onderscheid tussen URI van entiteit en URI van pagina over entiteit • nieuw: vereiste URI's te representeren als QName • nieuw: vorm van URI's voor termenlijsten / ontologie. <p>Verder:</p> <ul style="list-style-type: none"> • nieuw: aanzet richtlijnen voor versiebeheer • verwijderd: voorbeelden redirect van access-omgeving naar resolver weggehaald (betere oplossing: beschrijving geeft resolverlink als een property).
1.1.1	31 maart 2017	Term 'catalogus' (bijlage V) vervangen door 'catalogiseersysteem' en opmerking over het 'PPN' verwijderd.
1.1.2	21 april 2017	Mede nav feedback van Frank Bergsma en Elsbeth Kwant: domeinnaam van URI's veranderd naar data.bibliotheek.nl. Domein 'data.bibliotheek.nl' lijkt organisatorisch gezien duurzaamheidsrisico's met zich mee te dragen. Dit speelt niet bij 'data.bibliotheek.nl', wat bovendien in lijn is met voorbeelden als 'data.archieven.nl' (Nationaal Archief).

Samenvatting

URI's, of specifiek http-URI's, vormen de basisbouwstenen voor Linked Data. Een URI is een *identifier* die een entiteit representeert. Zo'n entiteit kan een publicatie zijn, een auteur, een concept, et cetera. Binnen voorliggend URI-beleid wordt zowel de vorm van de door de Koninklijke Bibliotheek (KB) te creëren Linked Data-URI's bepaald als ook het gedrag van het informatiesysteem bij opvraging van de door deze URI's gepresenteerde entiteiten. De kern van het beleid is gebaseerd op internationale *best practices*, toegepast op de bestaande situatie bij de KB. Het betekent in hoofdlijnen dat de KB-*resolver* haar bestaande functionaliteit behoudt. Voor URI's van andersoortige objecten zal de opbouw <http://data.bibliotheken.nl/{type}/{verzameling}/{referentie}> als sjabloon gebruikt worden. URI's van door de KB te beheren termen uit vocabulaires of ontologieën gebruiken patroon <http://data.bibliotheken.nl/def> als basis.

1. Inleiding

URI's zijn *identifiers* voor dingen, entiteiten of ook wel objecten of *resources* genoemd. HTTP-URI's zijn URI's die gebruikt kunnen worden om met die URI via `http://` of via `https://` een representatie van de geïdentificeerde entiteit op te vragen¹. HTTP-URI's vormen de basisbouwstenen van Linked Data. Zonder HTTP-URI's is er geen Linked Data mogelijk.

Het onderhavige URI-beleid heeft betrekking op enerzijds de *vorm* van de gehanteerde HTTP-URI's als ook op het *gedrag* van het informatiesysteem wanneer de HTTP-URI's gebruikt worden om een object op te vragen. URI-beleid kan helpen om de zichtbaarheid van de KB te vergroten binnen de Linked Data-wolk. URI-beleid helpt de duurzame toegankelijkheid van de objecten in de collectie te vergroten. Het vastleggen van het gewenste gedrag van URI's bij het opvragen van de geïdentificeerde entiteit is nodig om een betrouwbare en consistente toegangsomgeving voor Linked Data te kunnen creëren. URI-beleid biedt hiermee een kader voor het genereren en publiceren van Linked Data binnen de Koninklijke Bibliotheek.

URI-beleid kan niet helemaal los gezien worden van metadata-beleid, maar bovenal is het zo dat het elkaar raakt, niet overlapt. In het kader van URI-beleid worden bijvoorbeeld geen uitspraken gedaan over te hanteren metadatamodellen², -processen of te gebruiken vocabulaires.

Het URI-beleid is in essentie een uitwerking, selectie en interpretatie van bestaande *best practices*³, toegepast op de situatie van de Koninklijke Bibliotheek. Het beleid wordt in hoofdstuk 2 puntsgewijs gepresenteerd. Daarbij wordt waar nodig verwezen naar onderbouwing en uitwerking die in de bijlagen te vinden is.

¹ Niet alle URI's zijn HTTP-URI's. NBNs zijn bijvoorbeeld URI's maar geen HTTP-URI's. Als in dit document over URI's gesproken worden dan worden daar HTTP-URI's mee bedoeld.

² Afgezien van het fundamentele onderscheid tussen 'dingen', 'beschrijvingen van dingen' en 'representaties van dingen', zie hiervoor punt 1 van hoofdstuk 2, verdere keuzes in de modellering van metadata hebben geen gevolg voor het URI-beleid.

³ Zie met name *W3C Interest Group Note 03 December 2008, Cool URIs for the Semantic Web* (<https://www.w3.org/TR/cooluris/>) maar ook *Platform Linked Data Nederland: Aanzet tot een nationale URI-Strategie voor Linked Data van de Nederlandse overheid* (<http://www.pilod.nl/wiki/Boek/URI-strategie>) en *Cabinet Office: Designing URI sets for the UK public sector* (<https://www.gov.uk/government/publications/designing-uri-sets-for-the-uk-public-sector>).

2. Eisen aan URI's binnen de KB

1. De KB maakt onderscheid tussen vier groepen type entiteiten en weerspiegelt dit in de vorm van de door haar beheerde URI's:

1. digitale representaties van objecten, bijvoorbeeld van een aflevering van een krant,
2. objecten die onafhankelijk van een informatiesysteem bestaan, bijvoorbeeld een persoon zoals de auteur van een boek,
3. beschrijvingen van objecten zoals vallende onder 2, bijvoorbeeld een beschrijving van een auteur,
4. vocabulaires of ontologieën en specifieke termen (danwel *classes* of *properties*) daarbinnen (feitelijk een verbijzondering van 2).

Justificatie:

*Deze indeling doet recht aan internationale best practices (2, 3 en 4), zonder de binnen de KB reeds bestaande praktijk voor digitale objecten los te laten (1.). Deze indeling is zinvol omdat juist de best practices voor het gedrag van informatiesystemen hier sterk mee samenhangen. Voor een nadere uitleg over deze indeling zie **bijlage I**.*

2. Voor URI's voor de digitale representaties van objecten (2.1.1) worden *resolver*-URI's gebruikt.

Justificatie:

*De KB publiceert al jaar en dag URI's via de metadata in KB-MDO. Hoewel de semantiek van dat wat deze URI's representeren nooit expliciet is vastgelegd gaat het grosso modo om digitale representaties van objecten. Er is geen noodzaak deze functionaliteit te veranderen. Zie **bijlage II** voor een nadere beschrijving van de resolver in relatie tot een te realiseren toegangsomgeving voor Linked Data.*

3. Het <http://data.bibliotheken.nl/>-domein wordt gebruikt voor URI's van entiteiten anders dan digitale representaties van objecten.

Justificatie:

*De resolver vormt een kernsysteem binnen de KBGA. Om complexiteit van dit bedrijfskritische systeem niet onnodig te laten toenemen (architectuurprincipe 'functionele componenten') zal de nieuwe voor Linked Data gewenste functionaliteit in een aparte omgeving gerealiseerd worden. Zie ook **bijlage II**.*

Het 'data'-subdomein wordt breed gebruikt voor het Linked Data URI's. Keuze voor domein 'bibliotheek.nl' volgt de best practice om geen organisatienamen in URI's op te nemen, verhoogt daarmee de potentiële duurzaamheid

4. De <http://data.bibliotheken.nl/>-URI's voor niet-informatieobjecten (2.1.2) en de beschrijvingen daarvan (2.1.3) zijn opgebouwd volgens een vaste systematiek:
<http://data.bibliotheken.nl/{type}/{verzameling}/{referentie}>

Justificatie: Hoewel de vorm van de identifiers bij gebruik door derden als betekenisloos beschouwd moet worden is het uit oogpunt van beheer nodig om een vaste systematiek te volgen. Dit is bijvoorbeeld nodig om te verzekeren dat URI's uit verschillende systemen uniek zijn.

Het expliciet aangeven van het 'type' sluit aan bij een breed gedragen praktijk ondersteunt op zichtbare wijze het ook via het '303 See other'-redirect mechanisme (zie 6.) benadrukte semantische onderscheid.

*Zie **bijlage III** voor een nadere invulling van de vorm van de toegepaste <http://data.bibliotheken.nl/>-URI's.*

5. Entiteiten zoals ontologieën, vocabulaires of termen die daar een onderdeel van uitmaken (genoemd onder 2.1.4) kennen een URI die begint met <http://data.bibliotheken.nl/def> (ofwel hebben een URI met 'def' als waarde voor '{type}' - conform 2.4).

Justificatie: In deze fase is er nog geen goed beeld in welke mate er behoefte zal zijn aan KB-eigen ontologieën en dergelijke. Duidelijk is wel dat aan dergelijke URI's andere eisen gesteld worden dan aan de URI's voor reguliere objecten.

- *wenselijk is dat URI's omgevormd kunnen worden tot betekenisvolle QNames,*
- *als een ontologie niet stabiel is dan zal het wenselijk zijn om in de URI een aanduiding voor de versie op te nemen (bijvoorbeeld met een datum).*

6. De KB implementeert http-'content negotiation' en HTTP-redirects conform *best practices*, zowel binnen de <http://data.bibliotheken.nl/> -omgeving als ook binnen de resolver.

Justificatie:

Via content negotiation kan de gebruiker op gestandaardiseerde wijze kiezen in welk formaat een beschrijving geleverd wordt, bijvoorbeeld RDF/XML of JSON-LD. Redirects conform best practices helpen de server-respons te interpreteren en verhelderen het onderscheid tussen entiteiten en de beschrijving van die entiteiten.

*Ondersteuning voor content-negotiation is de enige aanpassing van de resolver die benodigd is⁴. Via content-negotiation worden de representaties van entiteiten gekoppeld aan de beschrijvingen in de <http://data.bibliotheken.nl/>-omgeving Zie verder **bijlage IV** voor voorbeeld-scenario's waarin content negotiation en redirects toegepast worden.*

⁴ Zoals er binnen <http://data.bibliotheken.nl/> in verband met gebruikersgemak URLs komen voor specifieke serialisaties van de data zo is het ook zinvol om binnen de resolver-omgeving ondersteuning voor de 'role'-functionaliteit breed en consequente te implementeren.

7. De KB zet zich in om de duurzaamheid van gehanteerde URI's te waarborgen

Justificatie:

*Omdat URI's de bouwstenen vormen voor Linked Data moeten ze gebruikt kunnen worden om een stevig, duurzaam fundament te bouwen. Zie verder **bijlage V**.*

8. Beheer van URI's geschiedt zoveel mogelijk in bronsystemen

Justificatie:

*URI's zijn in een Linked Data-wereld het belangrijkste 'kenmerk' van een object. Het is zowel vanuit het perspectief van beheerbaarheid als ook vanuit perspectief van duurzaamheid wenselijk dat de gegevens van een object bij voorkeur in één, bronsysteem beheerd worden. Dit is bijvoorbeeld van belang voor de overgang van GGC naar WorldShare. Zie verder ook **bijlage V**.*

Bijlage I

URI's, dingen, beschrijvingen en representaties

Voor een goed begrip van de strekking en implicaties van het URI-beleid is het van belang om vooraf de betekenis van enkele kernbegrippen helder te krijgen. Het gaat hierbij om het onderscheid tussen de termen URI's, dingen, objecten of entiteiten, beschrijvingen en representaties. Bovendien wordt ook de Engelstalige begrippen *non information object* en *real world object* verhelderd.

Begrip van het onderscheid tussen de verschillende soorten met URI's te identificeren zaken is specifiek van belang omdat het gedrag van URI's conform *best practices* hier sterk mee samenhangt.

Het inleidende hoofdstuk stelt dat URI's *identifiers* zijn die een 'entiteit', een zeker ding of object kunnen identificeren. Deze termen worden hier uitwisselbaar gebruikt. Zo'n entiteit kan van alles zijn, een bestaande of fictieve persoon, een boek in de zin van een abstract 'werk' of juist het specifieke gematerialiseerde boek dat op mijn bureau ligt, een filosofisch concept, een webpagina, enzovoort.

Er zijn geen beperkingen ten aanzien van wat URI's kunnen identificeren. URI's identificeren niet alleen digitale dingen of dingen op het internet, maar ook dingen daarbuiten, uit de 'echte wereld'. Dit soort zaken, die niet per sé fysiek of tastbaar van aard hoeven te zijn, worden ook wel *real world objects* genoemd.

Binnen de *best practices* ten aanzien van het gedrag van informatiesystemen voor Linked Data wordt ook onderscheid gemaakt tussen het ding zelf, de beschrijving van het ding en de representatie van het ding in het informatiesysteem. Ieder van die zaken, de representatie, de beschrijving en het ding zelf, kan een eigen URI hebben.

Ter illustratie kan het schilderij 'De Nachtwacht' van Rembrandt een URI toebedeeld krijgen. In een informatiesysteem is nooit het schilderij 'De Nachtwacht' zelf te bekijken, hooguit een digitale representatie met een eigen URI⁵. De (metadata-) beschrijving van 'De Nachtwacht' zou tot slot ook met een eigen URI geïdentificeerd worden.

Hoewel ieder ding beschreven kan worden, is er een klasse van *real world objects* die binnen een informatiesysteem niet gerepresenteerd kunnen worden. Anders dan 'De Nachtwacht' kan een persoon wel beschreven worden maar de essentiële kenmerken⁶ van een persoon kunnen niet binnen een informatiesysteem gerepresenteerd worden. In de Engelstalige literatuur wordt een persoon dan een *non information object* of *non information resource* genoemd. Een *non information object* kan dus zelf een URI hebben en ook een URI voor een beschrijving maar nooit voor een representatie.

⁵ En omdat er meerdere representaties mogelijk zijn, zijn er hier ook meerdere URI's mogelijk.

⁶ Zie W3C Interest Group Note 03 December 2008, Cool URIs for the Semantic Web (<https://www.w3.org/TR/cooluris/>)

Een dikwijls moeilijk te beslechten discussie betreft de vraag wanneer gesteld kan worden dat de essentiële kenmerken⁷ van een object binnen het informatiesysteem gerepresenteerd kunnen worden. Voor bijvoorbeeld een fiets of een levend persoon is duidelijk dat ze in essentie niet te representeren zijn in een informatiesysteem maar over de representeerbaarheid van 'De Nachtwacht' valt te twisten. Deze discussie wordt hier gepasseerd door het onderscheid terug te brengen tot enerzijds dingen die bestaan los van een informatiesysteem' (zoals een concept of een gebouw) en anderzijds dingen die in essentie een document in een informatiesysteem zijn (zoals digitale representaties van objecten of digitale beschrijvingen van objecten⁸).

Bij een ontologie of vocabulaire komen beide niveaus bij elkaar. In essentie is een ontologie een conceptueel model dat gebruikt kan worden als een modelmatige kapstok om te beschrijven objecten een plaats binnen te kunnen geven. Een concept is een entiteit buiten het informatiesysteem maar juist voor concepten uit een ontologie of vocabulaire is het essentieel dat ze ook door het informatiesysteem beschreven kunnen worden en dat er binnen het informatiedomein naar de concepten zelf verwezen kan worden.

⁷ <https://www.w3.org/TR/cooluris/>: "we have a Web document (there called *information resource*) if *all its essential characteristics can be conveyed in a message*. Examples are a Web page, an image or a product catalog."

⁸ Deze indeling lijkt goed aan te sluiten bij <http://lists.w3.org/Archives/Public/www-tag/2003Jul/0377>

Bijlage II

Gebruik *resolver*-URI's

II.1. De relatie tussen de *resolver* en <http://data.bibliotheek.nl/>

De KB publiceert al jaar en dag URI's met de metadata in KB-MDO. De objecten achter deze URI's kunnen met de *resolver* opgevraagd worden dankzij <http://resolver.kb.nl/> URI's. De *resolver* levert dan het gevraagde object in één standaard representatie. Naast *resolver*-URI's is enige tijd geleden een begin gemaakt om <http://data.kb.nl/> te gebruiken⁹ voor het leveren van beschrijvingen in één (RDF-) formaat.

Hoewel het voor de semantiek van de *resolver*-URI's nooit expliciet is bepaald, kan op basis van de manier waarop de *resolver* nu ingezet wordt, grosso modo gesteld worden dat de *resolver*-URI's representaties van digitale objecten identificeren. Omdat de *resolver* geen HTTP-*content negotiation* ondersteunt gaat het daarbij specifiek om representaties in één standaard formaat.

Er zou voor gekozen kunnen worden om de *resolver* zo uit te breiden dat deze ook ondersteuning biedt voor URI's van *real world objects*¹⁰, en voor het leveren van beschrijvingen in andere formaten. Dit vereist dat de *resolver* aanvullende URI-patronen moet gaan ondersteunen en dat uitgebreide voorzieningen voor *content negotiation* ingebouwd moeten worden. Dit zou een relatief grote aanpassing betekenen in een voor de KB-infrastructuur zeer kritische component.

Een eenvoudiger te realiseren oplossing die de infrastructuur bovendien minder belast is om een nieuwe omgeving onder <http://data.bibliotheken.nl/>¹¹ in te zetten voor URI's voor beschrijvingen en *real world* objecten. De <http://data.bibliotheken.nl/> kan ook veel eenvoudiger geschikt gemaakt worden om te functioneren als Linked Data-portal.

- Domein resolver.kb.nl wordt gebruikt voor URI's van digitale objecten. Bestaande *resolver*-URI's worden geïnterpreteerd als URI's van digitale objecten¹².
- Domein data.bibliotheken.nl wordt gebruikt voor URI's van overige entiteiten, te weten beschrijvingen¹³ van objecten en van *real world* objecten, waaronder ook ontologieën en vocabulaires.

⁹ Te zien als voorloper voor data.bibliotheek.nl, zie versie 1.0 van dit document.

¹⁰ Juist *real world objects* vormen een ideale verbindende schakel voor Linked Data. Het is daarom zinvol dat de KB ook eigen URI's inzet voor real world object. Denk hierbij bijvoorbeeld aan URI's die auteurs die in de thesaurus beschreven staan identificeren.

¹¹ Uit oogpunt van persistentie is het zinvol om de al gepubliceerde data.kb.nl-URI's door te laten verwijzen naar het nieuwe data.bibliotheek.nl-equivalent.

¹² Dit maakt het mogelijk bestaande functionaliteit van de *resolver* in hoofdlijnen in stand te houden. Alleen voor de overige entiteiten zal in een aparte omgeving -data.bibliotheek.nl- functionaliteit bijgebouwd moeten worden.

¹³ Beschrijvingen, hier itt representaties.

De samenhang tussen representaties en beschrijvingen, tussen de *resolver* en <http://data.bibliotheken.nl/> zal ook functioneel tot uiting moeten komen. Aan de zijde van de *resolver* wordt dit gerealiseerd door HTTP-*content negotiation* te implementeren. Op deze wijze kan van een *resolver*-object via de *resolver*-URI bijvoorbeeld de beschrijving in RDF/XML opgevraagd worden. De *resolver* zal hierbij een *redirect* geven naar de overeenkomstige <http://data.bibliotheken.nl/>-URI. Zie verder **bijlage IV** voor scenario's die de relatie tussen <http://data.bibliotheken.nl/> en de *resolver* illustreren.

II.2 De vorm van *resolver*-URI's

De uiterlijke verschijningsvorm van *resolver*-links (URI's) wordt hier enkel in hoofdlijnen besproken. Vermeld dient te worden dat in de loop van de jaren er onwenselijk veel variatie is ontstaan. Dit bemoeilijkt beheer en vertraagt softwareontwikkeling. Aanbevolen wordt de regels voor *resolver*-links te documenteren en het navolgen van de bestaande *best practices* af te gaan dwingen.

Niet alle mogelijke en gebruikte URLs binnen het resolver.kb.nl-domein zijn te beschouwen als volwaardige URI's. Zo kan bijvoorbeeld een afbeelding uit GVN opgevraagd worden met

- <http://resolver.kb.nl/resolve?urn=urn:gvn:NFA05:JES-950-14&role=thumbnail>
- <http://resolver.kb.nl/resolve?urn=urn:gvn:NFA05:JES-950-14&role=large>
- <http://resolver.kb.nl/resolve?urn=urn:gvn:NFA05:JES-950-14&role=tiny>
- <http://resolver.kb.nl/resolve?urn=urn:gvn:NFA05:JES-950-14>

Alleen de variant zonder aanvullende url-parameters beschouwen we als de 'officiële' ('canonical') URI. Deze laatste geeft nu standaard een *redirect* naar een overeenkomstige pagina in de website Geheugen van Nederland. Voorstel is om de *resolver* via een /robots.txt-constructie zo aan te passen dat alleen de canonical URI's in de Google-zoekindex opgenomen worden.

Bijlage III

<http://data.bibliotheken.nl/> -URI's

1. Domein

Als domein voor de URI's van objecten, beschrijvingen van die objecten én voor (termen uit) termenlijsten en ontologieën zal data.bibliotheken.nl gehanteerd worden. De keuze voor domein 'bibliotheken.nl' volgt de best practice om geen organisatienamen in URI's op te nemen, verhoogt daarmee de potentiële duurzaamheid. De domeinnaam sluit bovendien goed aan bij het omvatte informatiedomein en is herkenbaar en aanspreken.

Een entiteit zal zo mogelijk niet meer dan één URI in het domein toebedeeld krijgen en URI's zullen ook nimmer hergebruikt worden.

2. Objecten en beschrijvingen

Binnen het data.bibliotheken.nl-domein hanteren we abstracte URI's om de vorm van de URI niet te verwarren met de betekenis van dat wat met de data achter de URI beschreven wordt. Conform een veel voorkomende praktijk geeft de URI wél expliciet aan of het een object betreft, de beschrijving van een object of een term of groep termen uit ontologie of vocabulaire. Het gehanteerde model voor URI's van objecten of beschrijvingen is:

```
http://data.bibliotheken.nl/{type}/{verzameling}/{referentie}
```

Het '{type}' wordt ingevuld met een aanduiding van entiteitstype, waarbij 'id' voor niet-informatieobjecten staat (volgens 2.1.2) en 'doc' voor beschrijvingen (volgens 2.1.3). De string '{verzameling}' kan intern gelezen worden als een aanduiding voor het bronsysteem van de er door gegroepeerde objecten en dient ook om te verzekeren dat de combinatie '{verzameling}' en '{referentie}' een unieke waarde oplevert. Als uitgangspunt voor de invulling van '{verzameling}' geldt dat een abstracte en dus ogenschijnlijk betekenisloze string gekozen wordt teneinde de persistentie van de URI's maximaal te kunnen verzekeren¹⁴,

De string '{referentie}' bevat in de kern de identifier uit het bronsysteem maar om ook als deel van een QName¹⁵ te kunnen functioneren wordt het zo nodig met een prefix tot een geldige NCName¹⁶ gemaakt. Als prefix kan gedacht worden aan een string die de aard van de gebruikte identifier benadrukt bijvoorbeeld 'ppn' voor 'ocn'.

¹⁴ Hiermee wordt afgeweken van de dubbelzinnig aanbeveling uit "Aanzet tot een nationale URI-Strategie voor Linked Data van de Nederlandse overheid" om hier een voor menselijke lezers betekenisvolle term te gebruiken die tegelijkertijd 'zonder semantische waarde' zou moeten zijn.

¹⁵ <https://www.w3.org/TR/REC-xml-names/#NT-QName>

¹⁶ <https://www.w3.org/TR/REC-xml-names/#NT-NCName>

3. Serialisaties

Beschrijvingen van objecten zullen via *content negotiation* in verschillende formaten beschikbaar gesteld worden, zoals bijvoorbeeld RDF/XML of JSON-LD. Het is handig als deze specifiek representaties van de beschrijvingen ook direct via een URI opgevraagd kunnen worden¹⁷. Hiervoor worden URI's gebruikt met de volgende vorm:

```
http://data.bibliotheken.nl/{type}/{verzameling}/{referentie}.{formaat}
```

Uitgangspunt is om, zowel via de '{formaat}'-extensie als ook via *content negotiation*, de meest gangbare serialisaties te ondersteunen, zie de hiervoor de volgende tabel.

.{formaat}:	Serialisatie:
.html	HTML
.rdf	RDF/XML
.nt	n-triples
.ttl	Turtle
.json	JSON-LD

4. Ontologie en termen

Termen binnen een termenlijst, of meer formeel, classes en properities binnen een ontologie, zullen een URI gebruiken die begin met:

```
http://data.bibliotheken.nl/def
```

Verder invulling wordt in deze fase niet vastgelegd. Er zou hier op verschillende manieren gebruik van gemaakt kunnen worden. Bijvoorbeeld losse termen zonder een al te formele samenhang zouden direct als een hash-uri onder /def gepubliceerd kunnen worden, bijvoorbeeld:

```
http://data.bibliotheken.nl/def#ppn
```

In gevallen waarin een ontologie bepaald is met daarin termen dan zal het in vele gevallen een goede praktijk zijn om een versie-aanduiding (bijvoorbeeld met een wijzigingsdatum) in de URI op te nemen. Een beschrijving van een ontologie zou dan de volgende URI kunnen hebben:

```
http://data.bibliotheken.nl/def/objectmodel/20170320
```

¹⁷ Deze URI's worden bij *content negotiation* meegegeven in de zgn 'Content-Location'-header.

Conform eerder aangehaald best practices krijgt in dit voorbeeld de ontologie zelf (als concept) deze URI:

<http://data.bibliotheken.nl/def/objectmodel/20170320#>

Termen van classes of properties binnen deze ontologie krijgen dan URI's als bijvoorbeeld:

<http://data.bibliotheken.nl/def/objectmodel/20170320#signatuur>

<http://data.bibliotheken.nl/def/objectmodel/20170320#Telefoonboek>

Bemerk, conform standaardpraktijken wordt een class aangeduid met een term met een beginkapitaal.

Bijlage IV

Voorbeeldscenario's met *content negotiation* en *redirects*

De in deze bijlage gepresenteerde voorbeeldscenario's dienen ter illustratie van een implementatie conform *best practices* van de relatie tussen objecten, representaties en beschrijvingen enerzijds en de data.kb.nl- en de resolver-omgeving anderzijds.

IV.1. Opvragen *real world* object

In dit voorbeeld wordt een *real world* object opgevraagd van data.kb.nl. Dit leidt tot een '303'-*redirect* naar een beschrijving. Deze beschrijving wordt geleverd in het via *content negotiation* verlangde formaat.

- verzoek om "text/html" van <http://data.bibliotheken.nl/id/001/ppn123456789>
- 303-*redirect* naar <http://data.bibliotheken.nl/doc/001/ppn123456789>
- 200 ('Content-Location' wijst naar <http://data.bibliotheken.nl/doc/001/ppn123456789.html>)

IV.2. *Deprecated* URI's en 301 *redirects*

Het kan voorkomen dat een URI niet meer actief ondersteunt wordt. Dit zou bijvoorbeeld kunnen gebeuren als eventueel bij overgang van GGC naar Worldshare URI's in plaats van op PPNs op OCNs gebaseerd gaan worden. Het is dan wenselijk dat de oude URI's een zinvolle respons blijven geven en zo mogelijk *redirecten* naar een equivalente URI. Dit kan met een 301-*redirect*. Feitelijk wordt hiermee een owl¹⁸:sameAs relatie aangegeven.

In het uitgewerkt voorbeeld worden de nu aangeboden 'thesaurus' URI's vervangen door een URI met een andere prefix (in dit geval het abstracte '003'). Omdat het hier een *real world* object betreft volgt ook een 303-*redirect*.

- verzoek om "rdf/xml" van <http://data.kb.nl/thesaurus/1234>
- 301-*redirect* naar http://data.bibliotheken.nl/id/{nog_in_te_vullen}/ppn1234
- 303-*redirect* naar http://data.bibliotheken.nl/doc/{nog_in_te_vullen}/ppn1234
- 200 ('Content-Location' wijst naar http://data.bibliotheken.nl/doc/{nog_in_te_vullen}/ppn1234.rdf)

IV.3. Relatie tussen *resolver* en data.bibliotheken.nl

Onderstaande voorbeelden illustreren de samenhang tussen object, representatie en beschrijving ofwel de samenhang tussen data.bibliotheken.nl en de resolver..

¹⁸ <http://www.w3.org/2002/07/owl#>

Dit eerste voorbeeld illustreert hoe in de bestaande situatie de *resolver* een standaard representatie van een object uit Geheugen van Nederland uitlevert. Dit is ook de wenselijke situatie, hiervoor is dus geen aanpassing van de *resolver* nodig.

- verzoek om "text/html" van <http://resolver.kb.nl/resolve?urn=urn:gvn:NFA05:JES-950-14>
- 302-redirect naar <http://www.geheugenvannederland.nl/?nl/items/NFA05:JES-950-14>
- 200 "OK"

Als nu in voorgaande situatie niet om het standaard formaat "text/html" maar bijvoorbeeld om "rdf/xml" gevraagd wordt zal de *resolver* anders moeten gaan reageren. In dit voorbeeld wordt veronderstelt dat de abstractie prefix "003" gebruikt wordt voor items die een MDO-*identificer* hebben.

- verzoek om "rdf/xml" van <http://resolver.kb.nl/resolve?urn=urn:gvn:NFA05:JES-950-14>
- 302-redirect naar http://data.bibliotheken.nl/doc/{nog_in_te_vullen}/gvn:NFA05:JES-950-14
- 200 ('Content-Location' wijst naar http://data.bibliotheken.nl/doc/{nog_in_te_vullen}/gvn:NFA05:JES-950-14.rdf)

Op een bijna geheel identieke wijze verloopt het opvragen van een beschrijving in een ander formaat, bijvoorbeeld als turtle:

- verzoek om "text/turtle" van <http://resolver.kb.nl/resolve?urn=urn:gvn:NFA05:JES-950-14>
- 302-redirect naar http://data.bibliotheken.nl/doc/{nog_in_te_vullen}/gvn:NFA05:JES-950-14
- 200 ('Content-Location' wijst naar http://data.bibliotheken.nl/doc/{nog_in_te_vullen}/gvn:NFA05:JES-950-14.ttl)

Voorgaande route van data.bibliotheken.nl naar de *resolver* werkt net wat anders als er niet om een standaard representatie van een object gevraagd wordt (dus een formaat anders dan "text/html"). De *resolver* levert immers alleen de standaard representatie, dus via *content negotiation* zal de *resolver* weer terugverwijzen naar de data.kb.nl-omgeving:

- verzoek om "rdf/xml" van <http://data.kb.nl/003/gvn:NFA05:JES-950-14>
- 301-redirect naar <http://resolver.kb.nl/resolve?urn=urn:gvn:NFA05:JES-950-14>
- 302-redirect naar <http://data.kb.nl/003/gvn:NFA05:JES-950-14/>
- 200 ('Content-Location' wijst naar <http://data.kb.nl/003/gvn:NFA05:JES-950-14/rdf>)

Bijlage V

Duurzaamheid URI's

Een URI is een *identifier*. Ten opzichte van een 'gewone' *identifier* representeert de *identifier* niet alleen het betreffende object maar kan een representatie of beschrijving van het object ook met die *identifier* over het world wide web (dus via het http://- of https://- protocol) opgevraagd worden. Om als bouwsteen voor Linked Data-toepassingen gebruikt te kunnen worden is het van belang dat een URI voldoende duurzaam is en het gebruik van de URI bij bevraging van het informatiesysteem zinvolle informatie blijft geven.

Een duurzame *identifier* is vrij van implementatiedetails en zo vrij van semantiek als praktisch mogelijk¹⁹. Vooral afhankelijkheid van technische implementatiedetails moet met oog op de duurzaamheid vermeden worden. Bovenal wordt duurzaamheid van *identifiers* zoals URI's verkregen doordat de KB als organisatie het benodigde conservatisme betracht bij aanpassingen aan informatiesystemen die nieuwe URI-patronen introduceren of – erger – al gehanteerde URI's onbruikbaar maken. Dit vormt een belangrijk argument voor het behouden van de de *resolver*-URI's voor de digitale objecten beheerd door de KB.

Naast de *resolver*-URI's voor de digitale objecten publiceert de KB sinds enige tijd ook URI's onder het data.kb.nl-domein. Daarmee wordt een domeinnaam benut die aansluit bij Linked Data-praktijken van andere nationale bibliotheken. Omdat duurzaamheid gediend is bij een domeinnaam waarin geen organisatiename in voorkomt zal binnen de publicatieomgeving als domeinnaam data.bibliotheek.nl gehanteerd worden. Dit domein zal als basis dienen voor de URI's van beschrijvingen en van niet-informatieobjecten, zoals bijvoorbeeld ook ontologieën. Hiermee wordt data.bibliotheek.nl ook de basis voor de KB-portal voor Linked Data.

De duurzaamheid van het voorgestelde URI-model (zie **bijlage III**) is ten dele afhankelijk van de duurzaamheid van de in de URI's gebruikte sub-*identifiers*. Omdat de KB niet zelf de duurzaamheid van alle gehanteerde *identifiers* kan garanderen is dit een risico.

De overgang van een PPN als *identifier* naar een OCN als *identifier* hoeft geen probleem te betekenen. Sommige URI's zullen eenvoudigweg gefundeerd zijn op OCNs, andere op PPNs. Dit kan goed gerealiseerd worden als de URI in het bronsysteem opgeslagen en beheerd wordt, in dit geval dus GGC / Worldshare.

Er zijn risico's die met deze aanpak nog niet opgelost worden. Als een object in de catalogus een ander PPN krijgt dan zal dit zonder bijzondere maatregelen tot de onwenselijke situatie leiden dat de oude URI 'verdwijnt' en er een nieuwe URI voor in de plaats komt. Een vergelijkbare situatie doet zich voor als een titel geheel uit de catalogus verdwijnt. Een ander probleem zou zijn als een PPN hergebruikt wordt.

Er zijn verschillende benaderingen denkbaar die er minstens toe leiden dat 'oude' URIs blijven werken en doorverwijzen naar de nieuwe URI. Ontwerp of keuze voor een dergelijke oplossing vallen buiten de reikwijdte van dit document.

¹⁹ Bij gebruik van een *identifier* van een derde partij moet afgezien worden van semantische aannames ten aanzien van de opbouw van de *identifier* (anders dat wat standaard geldt voor alle HTTP-URI's).

Hoewel het voorkomen moet worden zal er in sommige gevallen behoefte kunnen gaan ontstaan om een entiteit een nieuwe URI te geven. Een voorbeeld daarvan is het corpus van de huidige experimentele <http://data.kb.nl/thesaurus/> URI's. Het is wenselijk om die te blijven ondersteunen hoewel achterliggende entiteiten een nieuwe definitieve URI krijgen volgens het hier voorgestelde model.

De 'oude' URI's zullen benaderbaar kunnen blijven via een standaard HTTP-301-doorverwijzing.

Het URI-beleid heeft een relatie met metadata-beleid, specifiek het beleid ten aanzien van het gebruik van generieke *persistent identifiers* zoals bijvoorbeeld NBN's, DOI's of Handles. De implementatie van *persistent identifiers* is in het verleden niet eenvoudig gebleken. Gebruik van *persistent identifiers* als onderdeel van URI's zou echter wel tot de mogelijkheden kunnen behoren. De zeer generieke aard van Linked Data maakt een dergelijke toepassing van *persistent identifiers* echter nog een factor lastiger dan bij de domein-specifieke toepassingen voor meer traditionele metadata. In het onderhavige document wordt daarom wel uitgewerkt wat er nodig is om URI's te verduurzamen zonder hierbij expliciet gebruik te maken van de oplossingen die gebruikelijk onder de term '*persistent identifiers*' geschaard worden. We volgen daarmee de praktijk van het merendeel van de andere nationale bibliotheken²⁰. Uiteraard is het een aanbevelenswaardige praktijk om binnen de Linked Data beschrijvingen zelf (de RDF) eventuele al bestaande '*persistent identifiers*' als een kenmerk van het object op te nemen. Dit staat los van gebruik van *persistent identifiers* in de URI's. In de toekomst is het mogelijk om alsnog URI's te baseren op mechanismen voor *persistent identifiers*. Uiteraard zal dan ook de moeite genomen moeten worden om oude URI's correct door te laten verwijzen naar de nieuwe URI.

Afsluitend enige richtlijnen voor versiebeheer. Voorgesteld wordt om voor objecten zelf *geen* onderscheid te maken tussen versies. Van beschrijvingen van objecten kunnen *wel* verschillende versies bestaan. Indien er behoefte bestaat om een URI voor een nieuwe versie van een object te creëren is dit een aanwijzing dat de gehanteerde semantische *class* van het object te ruim gekozen is voor beoogde doel. Ter illustratie: als er van een boek een geheel nieuw geredigeerde versie uitgegeven wordt dan vraagt de achterliggende ontologie van het catalogiseersysteem om een nieuwe record aan te maken en niet om een nieuwe versie van een bestaand record te produceren.

Voorstel is dat hier voorgestelde sjabloon voor URI's van beschrijvingen in ieder geval gebruikt wordt om de meest recente versie aan te duiden. Voor het opvragen van specifieke oudere versies lijkt met Memento-protocol²¹ een praktisch zeer bruikbare voorziening te bieden.

Zoals in III.4 ook aangegeven zou het voor een ontologie of termenlijst wel zeer nuttig kunnen zijn om een creatie-datum of versie-aanduiding in de URI op te nemen.

²⁰ Zie inventarisatie Roland Cornelissen:

<https://docs.google.com/document/d/1GXRRSzMCC9-GHwrQadBXBLRELxn19PbogCO0bc67VQ>.

²¹ Zie <https://tools.ietf.org/html/rfc7089>