

URI-RICHTLIJNEN VOOR DATA.VLAANDEREN.BE

Versie /// 1.0

Publicatiedatum /// 18 december 2017

Auteurs:

Naam	Rol	Organisatie
Dirk De Baere	Reviewer	Informatie Vlaanderen
Geraldine Nolf	Reviewer	Informatie Vlaanderen
Erik Mannens	Reviewer	imec - UGent - IDLab
Dieter De Paepe	Auteur	imec - UGent - IDLab
Laurens De Vocht	Auteur	imec - UGent - IDLab
Mathias Van Compernelle	Auteur	imec - UGent - MICT

Datum aanmaak: 23 maart 2017

Datum afdruk: 18 december 2017

Interne bestandsnaam: URI-RichtlijnenVoorDataVlaanderenBe_v1.docx

<p>Informatie Vlaanderen</p> <p>Hoofdzetel Boudewijnlaan 30, 1000 Brussel +32 (0)2 553 72 02</p> <p>Regionale zetel Koningin Maria Hendrikaplein 70, 9000 Gent +32 (0)9 276 15 00</p> <p>informatie.vlaanderen@vlaanderen.be</p>
--



Documenthistoriek:

Revisie	Datum	Medewerker(s)	Commentaar
0.1	09/11/2015	L. De Vocht, D. De Paepe	Generiek maken document, buiten MIR/GDI context situeren, afstemmen op ISA en W3C best practices, Nieuw uitgewerkte voorbeelden, integreren voorbeeldsectie met sectie rond richtlijnen. Placeholder bijlage: resolvable van URI's en inhoudelijke invulling van URI's (RDF)
0.2	09/12/2015	L. De Vocht, D. De Paepe, G. Nolf, D. De Baere	Updates en herformuleringen. Nieuwe voorbeelden verder toelichten.
0.9	14/03/2016	L. De Vocht, D. De Paepe, G. Nolf, D. De Baere	Updates en herformuleringen. Nauwkeurig preciseren te maken keuzes i.v.m. richtlijnen voor Vlaanderen. URI Strategie -> URI-standaard. Technische bijlagen toegevoegd met nuttig advies en pointers m.b.t. implementaties van de URI-standaard.
0.9.1	25/05/2016	L. De Vocht, D. De Paepe	Omgaan met legacy (reeds bestaande) URI's. Verduidelijking versies.
0.10	09/06/2016	L. De Vocht, D. De Paepe	Toepassen styling template. Extractie van services naar bijlage. Verduidelijking versies, verwijderen "latest" versie regel.
0.11	01/07/2016	L. De Vocht, D. De Paepe	Toelaten hash URI's. Verduidelijking types.
0.12	09/08/2016	L. De Vocht, D. De Paepe	Toevoegen kenmerken persistente URI's. Betere afstemming met ISA termen.
0.13	05/10/2016	L. De Vocht M. Van Compernelle	Opsplitsen normerend en richtinggevend gedeelte.
0.14	31/10/2016	L. De Vocht M. Van Compernelle	Feedback werkgroep 11/10/2016 verwerken.
0.15	18/11/2016	L. De Vocht M. Van Compernelle	Feedback werkgroep 8/11/2016 verwerken.
0.9	13/12/2016	L. De Vocht M. Van Compernelle	Laatste call voor interne feedback 30/11/2016 verwerken
0.91	23/03/2017	L. De Vocht M. Van Compernelle	Feedback publieke werkgroep 12/02/2017 verwerken
1.0	26/10/2017	Dieter De Paepe Dirk De Baere	Revisie nav interne opmerkingen



Lijst van afkortingen

Term	Definitie
IANA	Internet Assigned Numbers Authority
ISO	International Standardisation Organisation
LOD	Linked Open Data
OGC	Open Geospatial Consortium
RDF	Resource Description Framework
REST	Representational State Transfer
UML	Unified Modeling Language
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
URN	Uniform Resource Name
UUID	Universally unique identifier
W3C	World Wide Web Consortium

Lijst van symbolen gebruikt in een URI-patroon

Symbool	Betekenis
{ }	URI-parameter
()	Selector om een geheel aan te duiden binnen een URI-patroon
*	0 of meerdere keren het voorgaande geheel
+	1 of meerdere keren het voorgaande geheel
?	0 of 1 enkele keer het voorgaande geheel



INHOUD

Inhoud.....	5
1 Inleiding.....	7
1.1 Doel.....	7
1.2 Scope en toepassingsgebied.....	7
2 Richtlijnen voor vervaardigen van persistente uri's voor data.....	9
2.1 Domein.....	9
2.2 Type.....	10
2.3 Concept.....	12
2.4 Referentie.....	14
2.4.1 Referentie-basis.....	14
2.4.2 Referentie-versie.....	14
3 Use Cases: toepassing bij Informatie Vlaanderen.....	17
3.1 Vlaamse Open Data Portaal.....	17
3.2 Services.....	17
3.3 OSLO Vocabulary.....	18
3.4 Data URIs.....	18
4 Bijlage A: Implementatie.....	19
4.1 URI's laten resolvable.....	19
4.2 Representatie van resources als Linked Data.....	19
4.3 Omgaan met verschillende bestandsformaten en talen in geval van Linked Data.....	20
5 Bijlage B: De ISA-kenmerken van goede persistente URI's.....	22



1 INLEIDING

Dit document bevat de richtlijnen van Informatie Vlaanderen, voor het beheer van persistente URI's voor data, opgesteld volgens de "Vlaamse URI-standaard voor data". Het is een begeleidend document voor de implementatie van de URI-standaard voor data door Informatie Vlaanderen. Het bevat richtlijnen en voorbeelden voor het aanmaken van goede persistente URI's voor data, adviezen inzake implementatie, en enkele voorbeelden bij bestaande URI's van Informatie Vlaanderen.

1.1 DOEL

Het doel van dit document, "URI-richtlijnen voor data.vlaanderen.be", is het beschrijven hoe informatie het best kan ontsloten worden volgens de "Vlaamse URI-Standaard voor data". "URI" staat voor "Uniform Resource Identifier". Een URI is een formele manier om een *resource* te identificeren. De identificatie kan geschieden door gebruik te maken van een Uniform Resource Locator (URL). Als een URL gebruik maakt van het HTTP(S)-protocol, dan verwijst de identificator naar een webdocument van of over een resource. Elke URL is een URI (van een resource die bestaat op het web)¹.

Specifiek in deze richtlijnen ligt de focus op hoe resources uniek worden geïdentificeerd en hoe de vormregels om de identificatie te vervaardigen, zoals ze opgelijst zijn in de Vlaamse URI-standaard voor data, dienen toegepast te worden.

In dit document dat de URI-richtlijnen bevat, staat de toelichting en verduidelijking van de toepassing van de vormregels uit de Vlaamse URI-standaard voor data. Een aantal resources moeten een duurzame identificatie (de zogenaamde persistentie) krijgen, door middel van een URI, waarbij de vormregels de bedoeling hebben om conflicten en problemen te voorkomen zodat een lange-termijn informatie-infrastructuur kan bestaan voor overheden in Vlaanderen. Onder duurzame identificatie wordt hier een "benaming" van een resource verstaan - dit kan bijvoorbeeld een (hash)code of een databank-sleutel zijn - die gegarandeerd gedurende lange tijd (meerdere decennia) geldig is en kan gebruikt blijven. Gebruikers (zowel organisaties, diensten als systemen) van URI's worden er namelijk afhankelijk van, gezien de manier waarop de URI's zijn opgebouwd als verwijzing naar resources.

Daarenboven zijn deze richtlijnen van belang voor de medewerkers van Informatie Vlaanderen, gezien de specifieke rol van het agentschap, m.n. het goede beheer van hulpmiddelen die een beschrijvende, duidende of ondersteunende functie ten aanzien van andere diensten en beleidsdomeinen. Dit kan o.a. betrekking hebben op de basisregisters, andere bestaande catalogi of (eigen) informatie-registers.

1.2 SCOPE EN TOEPASSINGSGBIED

We verduidelijken de vormregels uit de Vlaamse URI-standaard voor data.

¹ RFC3986 - <http://www.ietf.org/rfc/rfc3986.txt>



Onderstaande zaken zijn in scope van deze URI-richtlijnen:

- Een verduidelijking van de richtlijn(en) om een unieke identificatie op te stellen, te beheren en de verklarende motivatie (hoofdstuk 2);
- Voorbeelden van implementaties aan de hand van use cases (hoofdstuk 3). De use cases illustreren hoe de URI's er in concreto uit kunnen zien.
- Aanbevelingen voor de implementatie (bijlage A)
- De ISA kenmerken van goede persistente URI's (bijlage B)

Deze richtlijn geldt voor de technologische systemen, producteigenaren, themabeheerders en data-eigenaren die hun data persistent wensen te ontsluiten via data.vlaanderen.be.



2 RICHTLIJNEN VOOR VERVAARDIGEN VAN PERSISTENTE URI'S VOOR DATA

Om de basisconcepten van de URI-standaard – namespaces, identifiers, versies – op een uniforme wijze te kunnen vormen, worden een aantal richtlijnen bepaald die men hoort te volgen.

Bij deze richtlijnen is steeds vermeld op welke vormregels ze gebaseerd zijn en verwijzen we naar andere standaarden waarop ze geïnspireerd zijn.

De vormregels hieronder hebben de bedoeling om bindend te zijn voor het identificeren van informatie en het ontsluiten van **data**.

2.1 DOMEIN

Richtlijn 1: Een URI-domein hoort idealiter in het beheer of onder de bevoegdheid van de organisatie die resources aanbiedt.

Er kan via een subdomein gewerkt worden bij elke URI. Het is belangrijk dat de keuze van de (sub)domeinnaam niet specifiek gebonden is aan een naam van een instantie: de naam moet toekomstige veranderingen kunnen doorstaan. Domeinen of subdomeinen mogen dus niet overeenstemmen met bijvoorbeeld namen van organisaties of de beleidsdomeinen. Al kan het wel zijn dat de benaming van een (sub)domein overeenkomt met een instantie of een beleidsdomein op een bepaald moment, mits deze term generiek genoeg is. Via subdomeinen kan het beheer van de URI's voor een bepaald thema uitbesteed worden aan andere diensten en/of organisaties.

Richtlijn 2: Het top-level domein voor informatie gepubliceerd door Informatie Vlaanderen is **data.vlaanderen.be**, andere domeinen kunnen gebruikt worden wanneer het beheer van het domein doorgegeven kan worden indien de originele beheerder het domein niet meer kan/wil beheren.

Richtlijn 3: Omwille van consistentie raden we aan om via het top-level domein te garanderen dat *domein/subdomein* doorverwijst naar *subdomein.domein*. Het is aangewezen om waakzaam te zijn dat deze doorverwijzing gebeurt.

Voorbeeld

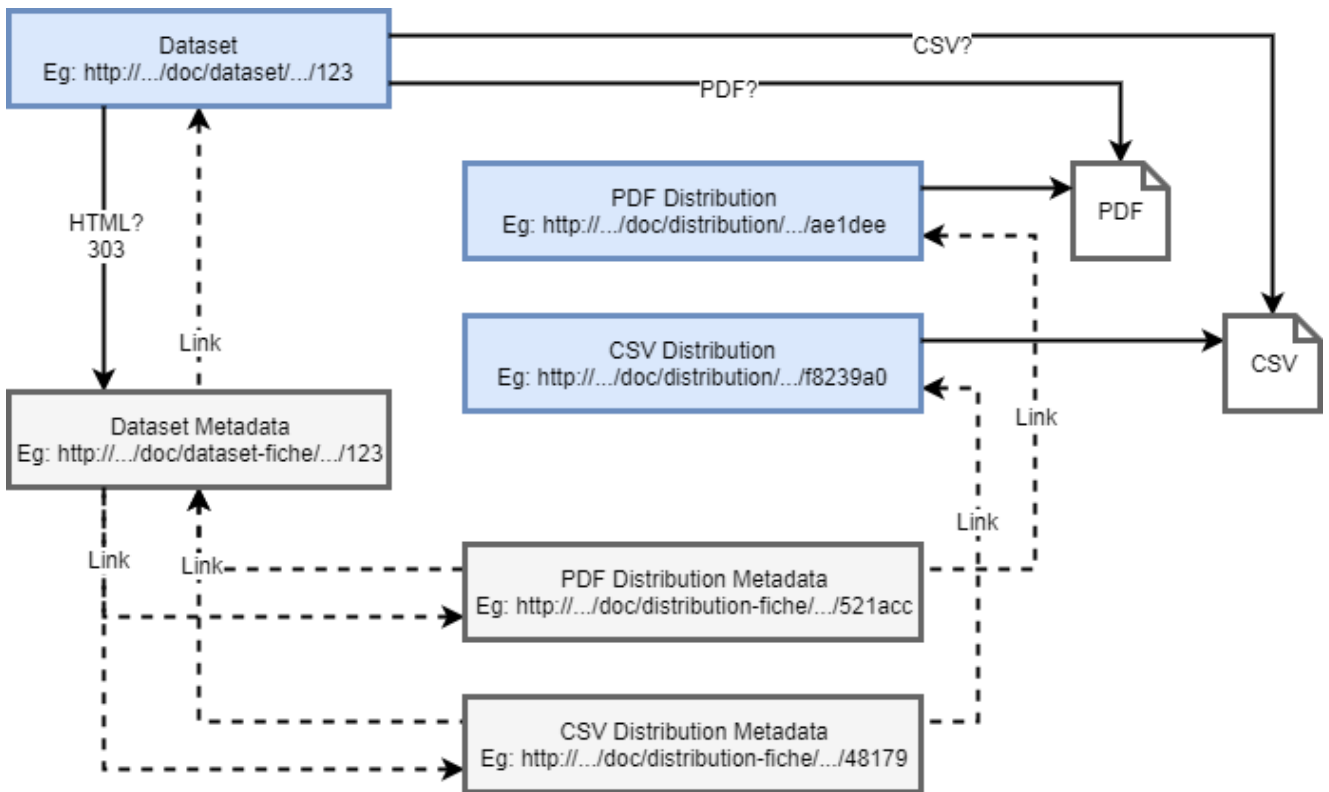
data.vlaanderen.be/wegen en *wegen.data.vlaanderen.be* verwijzen naar dezelfde informatie

We raden af om met meer dan één subdomein-niveau te werken onder data.vlaanderen.be

Voorbeeld

wegen.data.vlaanderen.be kan geen verdere subdomein-onderverdeling krijgen





In dit diagram stellen de blauwe rechthoeken de dataset en zijn distributies voor. Daarnaast hebben we ook een beschrijvende pagina voor elk, voorgesteld met de grijze rechthoeken, en de ruwe data van de dataset, voorgesteld door de 2 documenten aan de rechterkant.

Aangezien een dataset een information resource is, verwachten dit onder *doc*. We verwachten ook dat we de dataset kunnen downloaden via deze URI. Omdat er echter meerdere representaties van de data beschikbaar zijn, is het niet eenduidig welke representatie een client zou willen. Een client kan dit te kennen geven via *content negotiation*, op de tekening voorgesteld door de volle pijlen. Indien de client expliciet om een HTML representatie vraagt, of het voor de server niet duidelijk is welke versie de client wil, kan de server de client doorsturen naar een beschrijvende pagina van de dataset. Op deze beschrijvende pagina zullen dan HTML links aanwezig zijn die naar de aparte distributies linken, voorgesteld door de stippellijnen. Merk op dat ook op de metadata in verschillende vormen zou kunnen opgevraagd worden, bijvoorbeeld HTML, JSON of RDF.

Voorbeelden van URIs

We geven een voorbeeld van elk type, telkens in hetzelfde thema, om het onderscheid duidelijk te maken. We gebruiken voor de volledigheid hier reeds de volledige URI, die later in het document verder verklaard wordt. De voorbeelden hier zijn vooral educatief bedoeld en zijn niet optimaal qua structuur/naamgeving opgebouwd.

id/waterweg/schelde

is een identificatie voor het echte wereld object “de Schelde”. Dit object kan niet op het web bestaan, maar de referentie ernaar kan wel op het web bestaan, zo kunnen andere toepassingen of data naar het echte wereld object verwijzen.



doc/waterweg/schelde

verwijst naar een document (bv. HTML pagina) die uitleg geeft over de Schelde. Dit document kan gestructureerde data bevatten, maar dit is geen vereiste. Als er meerdere representaties beschikbaar zijn (HTML, JSON, RDF...) kunnen deze via *content negotiation* bereikt worden.

id/waterweg

Is een identificatie voor het abstracte concept “de verzameling van waterwegen”. Dit concept is een niet-informatieresource en krijgt dus een *id* URI. Het resoluten van deze URI zal resulteren in een 303 redirect naar het webdocument dat dit concept beschrijft: *doc/waterweg*.

doc/waterweg

Is een document dat het concept “de verzameling van waterwegen” beschrijft. Het kan bijvoorbeeld een uitleg geven van het concept en verwijzen naar een lijst/verzameling van waterwegen (*id/waterweg/schelde*, *id/waterweg/dender*, ...). Als er meerdere representaties beschikbaar zijn (HTML, JSON, RDF...) kunnen deze via *content negotiation* bereikt worden.

doc/meting/diepte/waterweg/schelde

is de ruwe data die het domein ter beschikking stelt omtrent dieptemetingen over “De Schelde”. Dit zal resoluten naar het effectieve bestand met de ruwe data.

ns/waterweg

is een namespace die kan gebruikt worden in een vocabularium dat over waterwegen handelt. Bv. *ns/waterweg#diepte* of *ns/waterweg/diepte* om de betekenis van de diepte van een waterweg aan te geven. Vaak zullen namespaces worden overgenomen vanuit andere domeinen. In dit voorbeeld zou bv. <http://linkedgeodata.org/ontology/depth> kunnen herbruikt worden.

2.3 CONCEPT

Zoals gesteld in de URI-standaard voor data <https://overheid.vlaanderen.be/OSLO-URI-standaard> stelt het *{concept}* als onderdeel van het URI-patroon de categorie van de resource voor.

Richtlijn 5: Een *{concept}* krijgt zijn betekenis binnen het (sub)domein en dient als volgt geïnterpreteerd te worden: “de resource is (een) *{concept}* *{type}* met als onderwerp *{referentie}*”.

Voorbeelden:

- *id/waterweg/schelde* is een waterweg identificatie met als onderwerp de Schelde.
- *doc/waterweg/schelde* is een waterwegen-metadata-document met als onderwerp de Schelde.
- *doc/meting/diepte/waterweg/schelde* is een ruwe data meting met als onderwerp de diepte van de Schelde.

Het concept wordt steeds in het enkelvoud en Nederlandstalig geformuleerd. Acroniemen, afkortingen en verkortingen van woorden zijn toegestaan.



Merk op dat er een onderscheid bestaat tussen het (sub)domein en het *{concept}*:

- Het is de bedoeling om binnen een (sub)domein afspraken te maken over het beheren en benoemen van de mogelijke concepten en hun toepassingsgebieden. Vervolgens volgt elk *{concept}* de gemaakte afspraken binnen het (sub)domein.
- Een *{concept}* is gebonden aan de strikte “is een” relatie in combinatie met het *{type}*, terwijl een (sub)domein dit niet is. Een duidelijk voorbeeld van een subdomein is “milieu”, dit zal niet kunnen voldoen aan de “is een” relatie, zoals onderstaand voorbeeld duidelijk maakt.

Voorbeeld:

Correct:

milieu.data.vlaanderen.be/id/waterweg/schelde is een waterweg-identificatie met als onderwerp: de Schelde.

data.vlaanderen.be/milieu/id/waterweg/schelde is een equivalente waterweg-identificatie met als onderwerp de Schelde.

Verkeerd:

data.vlaanderen.be/id/milieu/waterweg/schelde is een milieu-identificatie met als onderwerp de Schelde. In dit laatste voorbeeld wordt ‘milieu’ opgevat als een domein, in plaats van categorie van een resource.

Elke (sub)domein-beheerder moet waken over de eigen *{concept}*-termen. Het wordt aangeraden voor deze aanduidingen:

- te opteren voor een zo beknopt mogelijke en herbruikbare lijst van concepten;
- de betekenis van elk concept goed te documenteren.

De verantwoordelijkheid om de concepten en hun beschrijvingen te onderhouden, ligt bij de (sub)domeinbeheerder(s) in overleg met de resource-eigenaars.

Een *{concept}* heeft een specifieke semantiek in de context van een project/product/merk dat op een bepaald domein wordt gecodeerd. Dit kan ook een subdomein zijn, maar hoeft niet. Ideaal gezien overlappen de web-domeinen met de conceptuele domeinen (bijvoorbeeld waterwegen, spoorwegen, onderwijs, informatie e.d.)

Een *{concept}* kan geen actie of werkwoord zijn. Het is typisch een afkorting of een zelfstandig naamwoord. Ook zal het *{concept}* nooit een bestandsformaat zijn, het kan wel de benaming van een specificatie zijn.

Voorbeelden van concepten:

- register: registers
- organisatie: organisaties
- applicatie: applicaties
- specificatie: specificaties
- schema: schema’s
- ...

Voorbeelden:

<https://data.vlaanderen.be/id/organisatie/informatievlaanderen>

<https://data.vlaanderen.be/id/register/adres>



2.4 REFERENTIE

In deze sectie bespreken we hoe Informatie Vlaanderen de *referentie-basis* en *referentie-versie* interpreteert.

Referentie-basis: Dit deel van de *referentie* duidt de *resource* aan. Dit kan een lokale aanduiding zijn, bvb. afkomstig uit het achterliggende systeem, zolang deze samen met de combinatie *domein*, *type*, *concept* en eventueel voorgaande *referenties* tot een globaal unieke identificatie leidt.

Referentie-versie: Verwijst naar een bepaalde versie of timestamp. Dit kan een vast moment voorstellen (bvb 2016-01-01) of een relatief moment (bvb *vandaag*). De versie *latest* heeft een speciale betekenis en stelt de meest recente versie voor.

2.4.1 Referentie-basis

Richtlijn 6: De *{referentie-basis}* is een code die de identificatie voor de resource voorstelt. Dit kan een lokale identificatie zijn indien deze samen met de combinatie domein, concept, type en eventueel voorgaande *{referentie}*s tot een globale unieke identificatie leidt.

De resource dient eerst grondig geanalyseerd te worden, zo moet de identificatie, levensloop en precieze definitie gekend te zijn. Achteraf kan er een URI worden toegewezen. Aangezien de *{referentie-basis}* van zeer groot belang is voor de URI, is het van essentieel belang dat de resource-beheerder de grootst mogelijke zorg besteedt aan de persistentie en duurzaamheid ervan.

Een vaak voorkomend probleem is het automatisch genereren van (incrementele) volgnummers: daarbij is het niet evident om te garanderen dat bij het opnieuw genereren van de URI's op basis van dezelfde input data, datzelfde volgnummer zal gegenereerd worden. Er kunnen dus enkel (incrementele) volgnummers gebruikt worden als *{referentie-basis}* wanneer er kan gegarandeerd worden dat exact dezelfde nummers opnieuw zullen gegenereerd worden bij latere aanpassingen of herpublicaties. We volgen hierin de Europese strategie geformuleerd door ISA inzake incrementele opvolgnummers (zie ook boven sectie 3.8).

Indien er geen sleutel beschikbaar is die als identificatie kan dienen, kan een genormaliseerde versie van een tekstueel label of titel gebruikt worden. Merk op dat het gebruik van een naam die in de toekomst nog kan veranderen op zich afgeraden blijft.

Voorbeeld:

De URI <https://data.vlaanderen.be/id/event/ovl/aalter/bb2a5e00-c9d1-11e2> heeft als onderdelen:

- type: id
- concept: event
- referenties: ovl, aalter en bb2a5e00-c9d1-11e2

2.4.2 Referentie-versie

Richtlijn 7: De *{referentie-versie}* als onderdeel van de *{referentie}* duidt de versie voor één instantie van een document/resource aan, dit is dan ofwel een bepaalde revisie van het document/de resource ofwel een andere versie van de resource zelf (op een bepaald tijdstip).

////////////////////////////////////

Richtlijn 8: Bij het gebruik van jaartallen/datums/tijd als waarde voor de *referentie-versie* wordt het W3C Date Time² formaat gebruikt (volgens ISO 8601³).

Een resource kan een levenscyclus hebben, d.w.z. de resource kan veranderen: ofwel door een verandering in de echte wereld, ofwel door een verandering in de beschrijving. Indien elke verandering in de echte wereld wordt bijgehouden, heeft het eerstgenoemde sowieso het laatstgenoemde als gevolg.

Het bijhouden kan gebeuren aan de hand van een van onderstaande oplossingen of een combinatie ervan:

1. het genereren van een nieuwe *instantie* van de resource die de nieuwe versie voorstelt;
2. documenteren in de metadata die leeft buiten de resource (bijvoorbeeld tijds codering (timestamping) of versienummering).

Het al dan niet implementeren van versies voor resources hangt af van het doel van die resources en wordt dus uitgewerkt in de ontwerpfasen. Indien men met een URI wil kunnen verwijzen naar 1 specifieke versie van een resource (om een onderscheid te maken tegenover een andere versie van diezelfde resource), is de *referentie-versie* nodig, om zo de aparte versies als volwaardige resources te benaderen.

Indien versies worden geïmplementeerd, is het wellicht ook nuttig om naar de meest recente of huidige versie te kunnen verwijzen. Een aan te raden aanpak hiervoor is deze te mappen op de URI zonder de *referentie versie*. Bijvoorbeeld: <http://.../fiche/123/2010-01-01> wijst naar een specifieke versie, en <http://.../fiche/123> wijst naar de huidige versie. Alternatief kan men een aparte versie definiëren hiervoor, bijvoorbeeld: <http://.../fiche/123/latest>.

Het voordeel van aanduiding van levenscyclus in externe metadata is dat deze metadata in catalogi kunnen worden opgeslagen die dan op een gestandaardiseerde wijze kunnen worden ondervraagd⁴.

Wanneer is het refereren van verschillende versies van een resource nuttig? Onderstaande checks worden het best behandeld tijdens de ontwerpfasen:

1. Er kan iets veranderen in de echte wereld en er moet zowel naar de nieuwe als de oude resource kunnen gerefereerd worden/blijven.

Voorbeeld: Verandering in de echte wereld

De grenzen van een gemeente worden gewijzigd d.m.v. een Wet.(de oude grensinformatie blijft toegankelijk).

2. Om een onderscheid te kunnen maken tussen de levensloop van echte-wereld objecten en hoe ze beschreven zijn (hun representatie).

Voorbeeld: Verandering in de beschrijving

De ligging (geometrie) van de grenzen van een gemeente wordt accurater opgemeten (de oude en de meer accurate nieuwe geometrie informatie blijven toegankelijk).

3. Archivering: Als zaken (bv. beslissingen) gebaseerd zijn op een toestand van een resource in het verleden, dan moet die toestand uit het verleden kunnen opgevraagd worden vanuit het "archief".

² <https://www.w3.org/TR/NOTE-datetime>

³ <http://www.iso.org/iso/home/standards/iso8601.htm>

⁴ OGC Catalog Service standaard, <http://www.opengeospatial.org/standards/specifications/catalog>



Voorbeeld: Archivering

Onderstaand voorbeeld beschrijft de jaarlijkse update van de staat van de vervuiling van de Schelde, waarover jaarlijks een document wordt gepubliceerd.

Er zijn twee methodes: ofwel wordt er enkel gewerkt met een 'live' - domein, ofwel met zowel een 'live'-domein als met een 'archieef'-domein.

In de eerste variant wordt enkel met een live domein met versienummering zonder archiefdomein gewerkt. In de tweede variant wordt er gewerkt met een live-domein zonder versienummering en een archief-domein; zowel op het live- als op het archief-domein kan er met versienummers gewerkt worden, maar dit hoeft niet.

Stel dat er besloten is dat versies moeten worden bijgehouden, dan zijn er dus twee mogelijkheden:

1. Met versienummering op het live domein

In dit geval houdt de live-server de verschillende versies bij. Elke URI verwijst dus naar één enkele versie (snapshot) en zal nooit veranderen. Er kan een URI zijn die naar de meest recente versie doorverwijst.

Laatste versie:

http://{live_domein}/doc/meting/vervuiling/latest/waterweg/schelde/

Hierbij stelt het gebruik van "latest" als versie een stabiele URI naar de meest recente versie (in dit geval 2015) voor.

Vorige versies:

http://{live_domein}/doc/meting/vervuiling/2015/waterweg/schelde

http://{live_domein}/doc/meting/vervuiling/2014/waterweg/schelde

http://{live_domein}/doc/meting/vervuiling/2013/waterweg/schelde

2. Zonder versienummering op het live-domein

In dit geval stelt de URI de huidige (meest up to date) versie van de resource voor. Wanneer de resource geüpdatet wordt (en de oude versie moet behouden blijven), kan deze naar een archief verhuisd worden, waar er dan een nieuwe, bijkomende (archieef) URI wordt toegewezen aan de snapshot. De live-URI verwijst conceptueel nog steeds naar dezelfde resource, namelijk de meest recente versie van de meting, ook al is de achterliggende data veranderd.

Indien het nodig is dat men direct naar de nieuwe versie kan verwijzen (als snapshot), dan moet men bij elke update direct de nieuwe versie ook archiveren, aangezien enkel het archief snapshots bevat.

Vorige versies komen terecht op een archief domein:

http://{archieef_domein}/doc/meting/vervuiling_2015/waterweg/schelde

http://{archieef_domein}/doc/meting/vervuiling_2014/waterweg/schelde

...

Of

http://{archieef_domein}/doc/meting/vervuiling/2015/waterweg/schelde

http://{archieef_domein}/doc/meting/vervuiling/2014/waterweg/schelde

...



3 USE CASES: TOEPASSING BIJ INFORMATIE VLAANDEREN

Deze sectie beschrijft het gebruik van HTTP-gebaseerde URI's bij enkele producten van Informatie Vlaanderen. Dit is geen exhaustieve lijst maar is een weergave van enkele kenmerkende voorbeelden van wat de impact is op de URI's van enkele services en resources bij Informatie Vlaanderen.

3.1 VLAAMSE OPEN DATA PORTAAL

Voorbeelden van bestaande URIs:

1. Overzicht van de datasets: <http://opendata.vlaanderen.be/dataset>
2. Overzicht van datasets die van stad Gent over vogels: <http://opendata.vlaanderen.be/dataset?q=vogel&organization=stad-gent>
3. Metadata van 1 bepaalde dataset: <http://opendata.vlaanderen.be/dataset/vogeltelling-bourgoyen-waarnemingf900b>

Hoewel het data portaal als een service kan gezien worden (en de URI standaard dus niet van toepassing is), voldoet hij bijna aan alle vormregels. De query parameters in (2) vormen geen probleem, indien men de vorm van de parameters gegarandeerd stabiel houdt, of indien men de parameters op een bepaalde, machine-interpreteerbare manier kan beschrijven.

Ideale URIs:

1. <http://opendata.vlaanderen.be/doc/dataset>
2. <http://opendata.vlaanderen.be/doc/dataset?q=vogel&organization=stad-gent>
3. <http://opendata.vlaanderen.be/doc/dataset-fiche/vogeltelling-bourgoyen-waarnemingf900b>

Indien het open data portaal reeds bedoeld zijn om persistente URIs aan te bieden, zou de uitzonderingsregel voor legacy toelaten de huidige URIs te behouden.

Indien het open data portaal persistente URIs zou moeten aanbieden in de toekomst, moet men kunnen garanderen dat de aanwezige dataset identifiers blijven bestaan, waardoor een strenge governance nodig zal zijn.

3.2 SERVICES

Voorbeelden van bestaande URIs:

1. Geopunt zoekfunctie: <https://metadata.geopunt.be/zoekdienst/apps/tabsearch/index.html?hl=dut>
2. Geopunt TMS: <http://tile.geopunt.be/ws/raadpleegdiensten/tms/layer@CRS>
3. WSDL beschrijving van CRAB: <http://crab.agiv.be/wscrab/wscrab.svc?wsdl>

Gezien het hier over services gaat, vallen deze buiten de scope van de URI standaard.



3.3 OSLO VOCABULARIUM

Voorbeelden van bestaande URIs:

1. De definitie van een Belgisch adres: <http://data.vlaanderen.be/ns/adres#Adres>
2. De definitie van de geplande start van een zitting: <http://data.vlaanderen.be/ns/besluit#geplandeStart>

Deze URIs voldoen aan alle vormregels en richtlijnen voor het definiëren van termen uit een vocabularium (regel 4.5 van de URI-standaard).

3.4 DATA URIS

Voorbeelden van bestaande URIs:

1. Een adres: <http://data.vlaanderen.be/id/adres/3785195>
2. De documentatie van een adres: <http://data.vlaanderen.be/doc/adres/3785195>
3. <http://mir.agiv.be/data/IMKL/v2.1/Aansluiting/vmw-be:67564444:v2>

Voorbeeld 1 en 2 voldoen aan alle regels en richtlijnen.

Voorbeeld 3 heeft enkele tekortkomingen:

- Het gebruik van de organisatiernaam “agiv” in de domeinnaam is een schending van regel 3 van de standaard.
- Indien deze URI een fysieke aansluiting (niet-informatieresource) zou voorstellen, is het gebruik van de “v2.1” wellicht problematisch. Een verhoging van de versie zou immers betekenen dat alle verwijzingen naar het object veranderen, wellicht is dit niet bedoeld.
- Indien de versie eerder betrekking heeft op het soort antwoord dat de server zou geven (bijvoorbeeld: het datamodel dat gebruikt zal worden) is de aanwezigheid van de “v2.1” verdedigbaar, al zijn er betere alternatieven: *content negotiation* over het gewenste formaat of het plaatsen van het formaat achteraan de *referentie* keten.
- Het laatste deel van de URI stelt wellicht een lokale id voor. Door de “v2” af te splitsen als een *referentie versie*, wordt het intuïtiever om op de URI voor de meeste recente versie te vinden.

Ideale URI voor voorbeeld 3:

- Indien bedoeld als echte wereld object: <http://data.vlaanderen.be/id/aansluiting/imkl/vmw-be/67564444/v2>
- Indien bedoeld als beschrijving: <http://data.vlaanderen.be/doc/aansluiting/vmw-be/67564444/v2/IMKL-v2.1>



4 BIJLAGE A: IMPLEMENTATIE

In dit hoofdstuk worden enkele opmerkingen met betrekking tot de implementatie van de URI-standaard bij Informatie Vlaanderen behandeld.

4.1 URI'S LATEN RESOLVEN

- ‘Resolven’ of ook wel ‘dereferencing’ van een URI betekent dat de resource waarnaar verwezen wordt, wordt opgehaald aan de hand van zijn URI via het HTTP(S) (GET) protocol. *“Obtain the representation of the resource that it (the URI) identifies”*⁵. Dit betekent evenwel ook dat een resource op meer dan één manier kan voorgesteld worden.
- Een goede manier om een redirect service te implementeren is via *rewrite rules* die ondersteund worden door standaard web servers (bijvoorbeeld zoals <http://purl.org>⁶ of <http://w3id.org>⁷). Meer informatie kan gevonden worden via de beschrijving van “A central redirection service for persistent URI’s”⁸.

4.2 REPRESENTATIE VAN RESOURCES ALS LINKED DATA

Het is aanbevolen dat de weergegeven representatie, bij het resolveren van de URI’s, volgende elementen voorziet:

- **labels** die de resources benoemen. Bij elke opgevraagde resource (van eender welk type) is het belangrijk om ergens een eenduidig label te vermelden. Dit kan bij de HTML in de meta-tags⁹ of rechtstreeks als gestructureerde data; bijvoorbeeld via *rdfs:label*¹⁰;
- **types** die de soort resource omschrijven net zoals de *labels* via bijvoorbeeld *rdf:type*¹¹;
- **referenties** naar andere resources waar relevant;
- een maximaal hergebruik van de W3C aanbevolen **herkomst terminologie** (‘provenance’)¹² in de metadata (‘doc’ type van resources) om de herkomst en de historiek van (de acties op) een bepaalde resource aan te duiden;

⁵ Zoals gesteld in “URI-Resolution” in RFC3986, <https://www.ietf.org/rfc/rfc3986>.

⁶ Garantie voor persistentie wordt geboden door The Internet Archive.

⁷ <https://github.com/perma-id/w3id.org>

⁸ <https://joinup.ec.europa.eu/community/semic/news/persistent-uri-taskforce-experiments-central-uri-redirection-service>

⁹ <https://www.w3.org/TR/rdfa-primer/#the-first-steps-adding-machine-readable-hints-to-web-pages>

¹⁰ https://www.w3.org/TR/rdf-schema/#ch_label

¹¹ https://www.w3.org/TR/rdf-schema/#ch_type

¹² <https://www.w3.org/TR/prov-o/>



- een ondersteuning voor **content-negotiation**¹³ voor elke URI met een representatie in HTML(+RDFa) en ten minste één andere gestructureerd formaat (bijvoorbeeld JSON of XML). En daarbij liefst de W3C Resource Description Framework (RDF) Recommendation¹⁴ te volgen (bijvoorbeeld JSON-LD of RDF/XML).

4.3 OMGAAN MET VERSCHILLENDE BESTANDSFORMATEN EN TALEN IN GEVAL VAN LINKED DATA

Content-negotiatie start in de eerste plaats bij de proxy regels die het domein afhandelt. Om een zo harmonieus mogelijk gedrag te bekomen beschrijven we in deze sectie een aantal implementatie-keuzes die best toegepast worden. Zo wordt een eenduidig verwachtingspatroon gecreëerd naar de ontwikkelaars en is het voor de toekomstige gebruikers (developers) duidelijker welk stramien te volgen om nieuwe data te incorporeren (bestaande) toepassingen.

Content-negotiatie veronderstelt een gedrag op basis van de gegeven waarde voor de “Accept”-header¹⁵ in een HTTP(S) request. Daarnaast moet er voor elk formaat een document-extensie gekozen worden om met een URI-referentie data op te kunnen halen in een bepaald formaat¹⁶. De minimaal ondersteunde formaten worden in onderstaande tabel weergegeven.

Formaat	Accept header	Extensie
HTML	text/html	.html
RDF/XML	application/rdf+xml	.rdf .xml
TURTLE	text/turtle	.ttl
NTRIPLES	text/ntriples	.nt
JSON-LD	application/ld+json	.jsonld .json

¹³ <https://www.w3.org/TR/cooluris/#implementation>

¹⁴ <https://www.w3.org/TR/rdf-schema/>

¹⁵ HTTP Accept Headers: <https://www.w3.org/Protocols/rfc2616/rfc2616-sec14.html>

¹⁶ Ter ondersteuning van het html link element `mijn adres in RDF/XML formaat`
<http://philarcher.org/diary/2013/uripersistence>



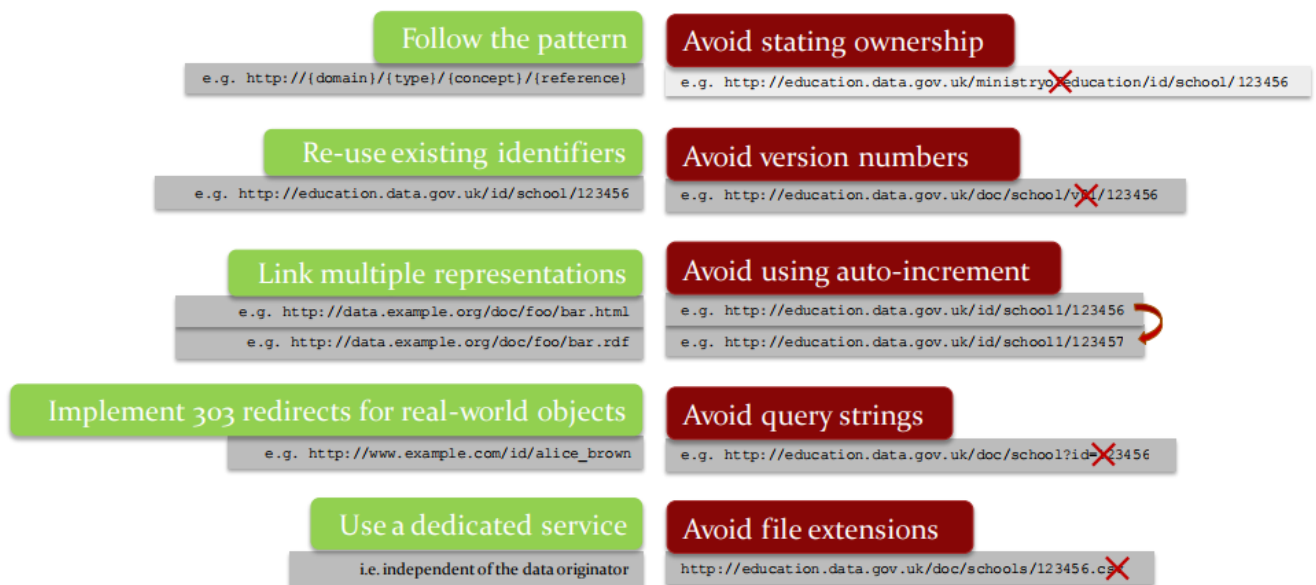
HTTP(S) ondersteunt ook meertaligheid in de content-negotiatie. Namelijk hiervoor wordt de 'Accept-Language'-header¹⁷ gebruikt. Op dit moment wordt er nog geen aanbeveling gemaakt rond het gebruik en het gewenste gedrag in verband met meertaligheid. Het is afgeraden om een verschillend domein per taal te gebruiken voor de persistente data-URI's.

¹⁷ HTTP Accept Language: <https://www.w3.org/Protocols/rfc2616/rfc2616-sec14.html#sec14.4>



5 BIJLAGE B: DE ISA-KENMERKEN VAN GOEDE PERSISTENTE URI'S

In 2012 heeft ISA, de Europese *Interoperability Solutions for Public Administrations*, een studie over persistente URI's uitgevoerd¹⁸, waarbij ze de "best practices" hebben geïdentificeerd. Hierin staan een aantal kenmerken opgelijst van persistente identificaties. Gebruik makend van deze lijst kan men beslissen in welke mate reeds gepubliceerde URI's voldoen aan de kwaliteitsnormen. Bij de opmaak van de Vlaamse URI-standaard voor data en de URI-richtlijnen voor data.vlaanderen.be, werden deze voorschriften gevolgd.



10 regels voor persistente URI's (bron: ISA¹⁹)

In deze bijlage bespreken we de richtlijnen die gebruikt kunnen worden bij het vormgeven van nieuwe URI's of bij de evaluatie van reeds bestaande URI's.

Wanneer men voor het eerst persistentie URI's publiceert of wanneer reeds bestaande URI's niet voldoen, dienen de vormregels uit deze bijlage gebruikt worden om correcte URI's te genereren.

Voor reeds bestaande URI's die niet voldoen aan bovenstaande kenmerken raden we aan om via redirects te verwijzen van de oude naar de nieuwe URI's, zodat ook de oude URI's geldig blijven. Nieuwe URI's dienen enkel via de nieuwe structuur aangemaakt.

¹⁸ ISA URI best practices, <https://joinup.ec.europa.eu/sites/default/files/D7.1.3%20-%20Study%20on%20persistent%20URIs.pdf>

¹⁹ ISA URI best practices, <https://joinup.ec.europa.eu/sites/default/files/D7.1.3%20-%20Study%20on%20persistent%20URIs.pdf>



Volg het patroon²⁰ <http://{domein}/{type}/{concept}/{referentie}>

Hergebruik bestaande identificaties²¹

Wanneer er reeds unieke (niet-URI) identificaties bestaan, kunnen deze gebruikt worden in de URI als sleutel. Een voorbeeld hiervan zijn de ISO landcodes.

Link meerdere representaties²²

Informatie kan verschillende formaten aannemen, bijvoorbeeld: HTML voor mensen, RDF of JSON voor machines. Het wordt aangeraden om voor elke informatie resource een formaat voor mensen en voor machines (bij voorkeur RDF) te voorzien, en deze te linken in elke representatie.

Gebruik 303-redirect voor niet-informatie resources²³

Wanneer men de URI resolvet van een niet-informatie resource (deze kan niet doorgestuurd worden als een serie van bytes), dan moet men via een HTTP 303-redirect doorverwezen worden naar een beschrijvend document van die resource.

Gebruik een toegewijde service²⁴

Gebruik mechanismen zodat het beheer niet gebonden is aan de eigenaar van de data en eenvoudig kan doorgegeven worden. Er is immers veel kans dat doorheen de tijd het beheer zal evolueren.

Vermijd de naam van de eigenaar²⁵

Aangezien instanties, projecten of diensten vlug evolueren doorheen de tijd, vermijdt men best de naam van de eigenaar van de resource in de identificatie.

Vermijd versienummers²⁶

Het gebruik van versienummers om updates aan te duiden aan een ontologie, taxonomie of resources wordt best vermeden om achterwaartse compatibiliteit te garanderen. In de vormregels hieronder beschrijven we hoe versies wel mogen gebruikt worden om resources te identificeren waarvan verschillende versies effectief verschillende resources voorstellen.

Vermijd auto-increment²⁷

Het meervoudig gebruik van auto-increment om sleutels te genereren voor resources is gevaarlijk als dit niet reproduceerbaar is. Er moet immers gegarandeerd worden dat de achterliggende betekenis van de URI niet verandert doorheen de tijd.

²⁰ <http://philarcher.org/diary/2013/uripersistence/#recURIform>

²¹ <http://philarcher.org/diary/2013/uripersistence/#reuseEx>

²² <http://philarcher.org/diary/2013/uripersistence/#desBuildMulti>

²³ <http://philarcher.org/diary/2013/uripersistence/#imp303>

²⁴ <http://philarcher.org/diary/2013/uripersistence/#dediService>

²⁵ <http://philarcher.org/diary/2013/uripersistence/#avoidOwn>

²⁶ <http://philarcher.org/diary/2013/uripersistence/#avoidVer>

²⁷ <http://philarcher.org/diary/2013/uripersistence/#avoidAuto>



Vermijd query strings²⁸

Query strings duiden vaak op een directe link met de onderliggende technologie, die wellicht zal evolueren over de tijd.

Vermijd bestand-extensies²⁹

Extensies duiden vaak op een directe link met de onderliggende technologie, die wellicht zal evolueren over de tijd.

²⁸ <http://philarcher.org/diary/2013/uripersistence/#avoidQuery>

²⁹ <http://philarcher.org/diary/2013/uripersistence/#avoidFileEx>

