

/// /Publieke werkgroep URI-standaard/

////////////////////////////////////
Datum: 16/01/2017

Locatie: VAC Gent, Jakob Van Artevelde

Aanwezig: Dirk De Baere, Ziggy Vanlischout, Laurens De Vocht, Koen Van Daele (Erfgoed Vlaanderen), Paul Hermans (ProXML), Bert Van Nuffelen (Tenforce), Wim Van Acker (Cipal Schaubroek), Jelle De Vreese (Cipal Schaubroek), Dirk Vermeylen (HPE), Bart Salen (Geosolutions), Aad Verstedden (Tenforce).

Afwezig: Tom Van Gulck, Katrien Desmet, Liesbet D'Hondt, Wim Devroey, Raf Buyle.

Bijlagen:

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

1 AGENDA

- Verwelkoming, praktische indeling werkgroep
- Contextschets: Vlaamse URI-standaard en richtlijn voor toepassing van de URI-standaard voor Informatie Vlaanderen
- Overlopen Vlaamse URI-standaard met vraag en antwoord

2 VERLOOP

2.1 VERWELKOMING

De sessie wordt gestart met een verwelkoming en enkele praktische mededelingen over het verloop van de sessie. Vervolgens stellen de deelnemers zich kort even voor.

2.2 CONTEXTSCHETS VLAAMSE URI-STANDAARD

Dirk De Baere start dit deel met een verwijzing naar [het huidige document URI-strategie gepubliceerd op het](#)

[Vlaams Open Dataportaal](#) en licht het ontstaan toe van 2 nieuwe documenten, nl.:

- **Vlaamse URI-standaard:**

Doelgroep: vlaamse overheidsdiensten in eerste instantie, maar ook federale overheid en andere internationale overheden moeten zich hier in kunnen terugvinden.

- **URI richtlijn voor toepassen URI-standaard, voor Informatie Vlaanderen:**

Apart document gemaakt over de toepasbaarheid van de URI-standaard. Doel is niet meteen om dit document voor andere instanties verplicht navolgbaar te maken maar het kan wel als hulpbron gebruikt worden bij vragen over de vereisten in de URI-standaard of bij het toetsen van de eigen URI's aan de nieuwe Vlaamse URI-standaard.

2.3 OVERLOPEN VLAAMSE URI-STANDAARD MET VRAAG EN ANTWOORD

Bedoeling is nu om in deze sessie alle aanwezigen de kans te geven om inhoudelijke feedback te geven op het document vanuit een specifieke visie ingegeven door de dagelijkse praktijk waarin de deelnemers aan de sessie aan de slag zijn.

Laurens start de bespreking van het URI-standaarddocument met een toelichting van de structuur van het document.

Inleiding

- Resource
- URI's
- Domeinnaam
- Namespaces

Vormregels

- Domein
- Type
- Concept
- Referentie

Uitzonderingen

URI-standaard Compliancy Checklist

Hierna volgt een verslag van de belangrijkste discussietopics, telkens weergegeven bij het onderdeel van het

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

document waar ze werden besproken.

Inleiding

Resources:

2 categorieën: informatie resources en niet-informatie resources

Vormregels

Belangrijkste onderdeel van de URI en uitgangspunt voor wat er in de richtlijn van Informatie Vlaanderen is vermeld.

Dit zijn de belangrijkste regels:

Regel 1: Het HTTPS of HTTP URI-schema moet gebruikt worden als basis voor definitie van URI's.

Regel 2: Alle URI's moeten gevormd worden volgens het patroon:

http(s)://{*domein*}/{*type*}/{*concept*}/{*referentie*}*

Regel 2.1: $\{domein\}$, $\{type\}$ en $\{concept\}$ zijn **verplichte** onderdelen van een URI.

Regel 2.2: *Referenties zijn **optioneel** en er kunnen er meerdere zijn naargelang de situatie.*

Regel 2.3: De basiscombinatie $\{domein\}/\{type\}/\{concept\}$ bij afwezigheid van een $\{referentie\}^*$ mag slechts naar één resource verwijzen (=zgn uniek op het web) en moet onveranderlijk zijn (daar de $\{referentie\}$ optioneel is).

Regel 2.4: De $\{referentie\}^*$ in combinatie met $\{domein\}/\{type\}/\{concept\}$ moet uniek op het web en onveranderlijk zijn. Dat wil zeggen dat eenmaal de $\{referentie\}$ is toegekend, mag de referentie zelf als de semantiek van de gerefereerde resource niet nooit meer wijzigen. Zo is het domein best niet afhankelijk van tijdsgebonden elementen zoals namen van toepassingen, merken, ... Voorbeeld uit eigen huis: `data.vlaanderen.be` en **niet** `data.informatievlaanderen.be`

Vormregeln: **Domein**

OPMERKING: erfgoed.net -> .net is internationaal domein. Valt op deze manier ‘buiten’ vlaanderen. Je kan ook ‘same as’ gebruiken. Net omdat de organisatiennaam al een paar keren gewijzigd is in de loop der tijd is gekozen voor erfgoed.net, naar analogie van Groot-Brittannië, waar met heritage.org gebruikt. **AANVULLING:** Boodschap is vooral om niet iets te kiezen als domein waarvan je weet dat het zeer vatbaar is voor wijzigingen. Koen geeft het voorbeeld van bij Erfgoed waar de organisatiennaam in de loop van het laatste decennia al een

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

aantal keren veranderd is en waar vervolgens domein migraties moesten worden uitgevoerd.

Vormregels: Type

Type beschrijft de mogelijke resources. Er zijn 3 categorieën die elk een naam kregen: id, doc, ns.

VRAAG: Wat is de meerwaarde om ID af te splitsen van het DOC? Om maw een aparte identifier te hebben voor het document, naast de id van het ding zelf. De opsplitsing is immers een cruciaal element in de URI-standaard. **ANTWOORD:** omdat aan het document ook metadata gekoppeld kan worden. Vraag wordt gesteld hoe je een factuur in deze logica dient te interpreteren: enerzijds het factuur als object, afgedrukt op briefpapier van een onderneming, versus de representatie van een factuur op het web. Ander voorbeeld dat het belang van het onderscheid beter duidt is bijvoorbeeld de DUBLIN CORE informatie over het schilderij 'De Mona Lisa': Auteur van het schilderij, versus de auteur van het document dat de Mona Lisa beschrijft. Er worden soms attributies gemaakt tav wetenschappelijke beschrijving van een kunstobject. Is heel gebruikelijk in de kunst en erfgoedwereld.

De nieuwe Vlaamse URI-standaard geeft beide methodes voor redirect weer omdat ze ook beiden toegelaten worden. De exacte keuze hangt af van het scenario en de mogelijkheden voor de specifieke implementatie. De standaard heeft een lichte voorkeur voor de 303 redirect ipv de #ID. Wel niet door elkaar gebruiken. Ook niet de intentie om nog een andere methode te introduceren of te ondersteunen. Als je nieuw begint wordt de 303-redirect methode aanbevolen, als je al de #id gebruikt dan mag deze verder gebruikt worden. We komen hier nog op terug in het onderdeel uitzonderingen.

Vormregels: Concept

Concept is een soort van classificatie van de resource. Wordt geïllustreerd aan de hand van waterweg. **VRAAG** naar de redenering hierachter? **ANTWOORD:** we willen vermijden dat concepten zich op hetzelfde niveau gaan manifesteren als subdomeinen. Zorgen dat er geen letterlijke overlap is, hoewel er semantisch een overlap kan zijn. Het is hier puur om het uitzicht. Unicité blijft. Wat hier nu in de standaard staat is al een afzwakking van een vroegere, strengere regel.

BEMERKING: Dit principe kan bij erfgoed wellicht soms voor technische problemen zorgen... er wordt later op teruggekomen.

Vormregels: Referentie

Referentie kan één of meerdere keren voorkomen. Maar belangrijk is dat ze naar 1 resource verwijzen. Je kan ze op 2 manieren invullen: basisresource met of zonder versie. Als je versie gebruikt wel aangeven over welke versie je het hebt. Versies gaan specifieke URI krijgen. Op niveau van de standaard wordt de keuze vrij gelaten.

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

Een referentieversie moet steeds volgen op een referentiepagina.

BEMERKING: verschillende versies van een document waarbij je telkens de laatste 'latest' opvraagt. 'Latest' kan op zich niet uniek zijn maar in combinatie met domein en referentie wel.

VRAAG: Kan je referentiestuk en versie van de resource wel nog makkelijk uit elkaar houden in de URI?

ANTWOORD: hoe het binnen een bepaald domein zal werken moet – in toepassing van de standaard – heel duidelijk omschreven worden want je zal het wellicht niet zo uit het gedrag kunnen afleiden. In de richtlijnen van Informatie Vlaanderen wordt hier een specifieke interpretatie voor gegeven.

Uitzonderingen

1. Als het gaat over services zijn de vormregels niet van toepassing, hoewel dit strikt gezien een resource zijn.

Men mag afwijken van de vormregels wanneer het gaat over services, want meestal niet wenselijk wegens toepassingspecifiek en volatiel. OPGELET: bij webservices die data ontsluiten zal de te ontsluiten data moeten voldoen aan de vormregels uit de URI-standaard om persistentie te bekomen. Bepaalde function calls zullen dan moeten gemapped worden op een persistente URI die een doorverwijzing doet.

Alle http en https-uri's die persistent moeten zijn moeten maw steeds behouden blijven. Als iemand al persistente uri's heeft moeten die niet veranderen. Het gaat hier dus hoofdzakelijk over nieuwe, of over services die nog niet het http of https protocol volgen.

VRAAG: als er voorheen webservices zijn die niet voldoen aan de persistentievoorwaarden (bvb resource van een ftp-link) zou het dan niet verstandig zijn om een 'same as' te definiëren voor een http? **ANTWOORD:** Een legacy systeem kan zijn regels aanhouden, nieuwe systemen volgen de nieuwe regels. Deze strategie moet je zien als een vorm van uitdoofbeleid.

OPMERKING: Technisch is het niet altijd handig om een set-up voor redirects halverwege aan te passen, dwz voor een deel van de dataset de oude regels en voor een ander deel de nieuwe regels. Beter om alles ineens te migreren. De standaard staat de migratie volgens eigen keuze nu niet toe. Als de standaard ooit verplicht wordt, dan komen we hier in de problemen. Daarom zinniger om te zeggen dat de legacy systemen de regels aanhouden tot er een volledige migratie komt en dat de nieuwe systemen bij aanvang de (nieuwe) regels volgen. Oude systemen zouden dan wel geen uri's van de oude stijl meer mogen genereren. **OPMERKING:** zelfs al wordt er een nieuw systeem opgezet, de oude al persistente URI's moeten ook blijven bestaan en gegarandeerd worden.

BESLISSING: passage onder uitzondering 1 in de standaard aanpassen zodat migratie volgens eigen keuze (planning) wel toegelaten wordt en dat de ‘oude’ systemen die reeds persistente URI’s aanbieden deze kunnen blijven behouden.

VOORBEELD: kathedraal Antwerpen zal hopelijk over 100 jaar nog persistent zijn, id van de kathedraal kan wel wijzigen. Persistentie is een nobel streven maw maar alles is per definitie wel gedoemd om eindig te zijn.

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

2. Het is toegestaan voor legacy-systemen om nieuwe, niet-conforme URI's te produceren als er daarnaast ook conforme URI's worden voorzien.
3. Elementen met als type "ns" mogen fragment-identifiers gebruiken om alle termen uit een bepaald vocabularium in één enkel webdocument, elk in hun namespace, te definiëren.
4. Een alternatieve manier voor regel 4.2, om het onderscheid tussen id/doc te maken, is door de introductie van een fragment-identifier (het gedeelte vanaf het hekje of hash symbool "#" in een URI). Wanneer een gebruiker deze URI zal resoven, wordt de fragment-identifier genegeerd (als onderdeel van het HTTP protocol) en zal in feite zo een andere resource opvragen. Deze andere resource stelt dan de informatie resource voor. Deze aanpak is enkel toegestaan als er voldoende redenen kunnen aangehaald worden om de aangeraden redirect aanpak uit regel 4.2 niet te volgen.

Afwijking van de redirect regel aan de hand van de fragment identifier kan er ook naar het concept verwezen worden maar het is niet uitgesloten dat indien iemand graag met een redirect werkt in het geval van een namespace, dan mag er met een 303-redirect met doorverwijzing gewerkt worden.

Voor vocabularia wordt een #ID vaak gebruikt maar zonder moet ook kunnen, om welke reden dan ook. Vanuit de standaard is er de voorkeur om met een 303-redirect te werken, maar een #ID kan ook als alternatief mits motivering van de reden waarom.

OPMERKING: In het document de voorkeur voor 303 expliciet maken. Ook vermelden dat ID een gereserveerde terminologie wordt binnen dit domein. Zo kan het niet bijvoorbeeld in een vocabularium gebruikt worden en voor verwarring zorgen. In principe zal er bij gebruik van het vocabularium een verwijzing staan in de namespace en zou deze verwarring niet mogelijk moeten zijn. Beter hier afschermen voor DOC.

URI-standaard Compliancy Checklist

Een herhaling van alle regels uit de URI-standaard maar dan in vraagvorm zodat je aan de hand van dit overzicht je overeenstemming met de standaard te kunnen inschatten.

VRAAG: Wordt ergens gespecificeerd wat een legacy systeem is? **ANTWOORD:** er wordt niet gespecificeerd wat een legacy systeem is in deze context. **VRAAG:** niet beter om dat dan te definiëren om te vermijden dat er aan de standaard ontsnapt wordt? **ANTWOORD:** Als je nog geen linked data publiceert dan val je onder de standaard als je vanuit dat systeem wil beginnen om linked data te publiceren. Als je wel al linked data publiceert dan val je onder de definitie van een legacy systeem. Deze voorwaarde zit al voor een stuk vervat in de verwoording van de pre-check, niet 100 expliciet, maar wel voldoende om dit te counteren. **CONCLUSIE:**

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

We passen document niet verder aan.

OPMERKING: Het voornaamste probleem met de nieuwe Vlaamse URI-standaard in zijn huidige bewoordingen zijn, volgens Koen, de DOC-URI's. Een DOC-URI moet op een apart domein zitten. Maar bij erfgoed is er een splitsing gemaakt: URI's voor concepten en aan de andere kant is er ook een document voor dat op een ander domein gehost is. De reden daarom is bepaald door de look and feel van de organisatie en de site. Het grote menselijke web en het dataweb wordt dus niet gesplitst. Nu lijkt het dat alle documenten moeten gemigreerd worden, waar er weinig 'buy-in' voor te vinden zal zijn binnen de organisatie van erfgoed. URI is trouwens al eens gewijzigd in het verleden. Toen is de persistente URI is dezelfde gebleven. Mensen die persistent willen koppelen, verwijzen naar id.erfgoed.net. Bij erfgoed is er geen strikte scheiding tussen de business systemen en de systemen die linked data publiceren. Daardoor zorgt de DOC-URI in de standaard voor problemen.

BESPREKING:

- Extra redirect toevoegen zou een optie kunnen zijn om dit op te lossen maar is technisch niet zo handig en wijkt ook wel af van de cool uri standaard omdat je een abstracte uri voor een DOC gaat toevoegen.
- Makkelijkste manier voor Erfgoed Vlaanderen om dit op te lossen is verduidelijking in de tekst van de standaard dat de manier van werken bij erfgoed vlaanderen toestaat; nl. de situatie tolereren waarbij doc en id op een verschillend domein zit waarbij je met de 303 redirect werkt. Voor de DOCs hebben ze bij erfgoed vlaanderen verschillende domeinen.
- Vraag is of id/doc/ns expliciet aanwezig zijn, als ze optioneel zijn dan kunnen ze desgevallend weggelaten worden. Als alternatief voor het toevoegen van dat type aan de namespace. We kunnen dit best in de standaard beschrijven als 2 alternatieve vormen.
- Persistente uri's en business toepassingen zitten bij erfgoed nog steeds op een apart domein. Daarom de vraag of het toegelaten is om een simpele redirect toe te voegen? **BESLISSING:** Dit hoeft misschien niet in de standaard maar wel in het verslag van deze vergadering. Akkoord voor opname in verslag.

Meerdere redirects zijn toegestaan om dan uiteindelijk de opgevraagde doc te kunnen resoluten. Probleem bij Erfgoed Vlaanderen wordt niet opgelost met extra redirect maar wel met extra URI.

DISCUSSIE of Concept niet moet worden weggelaten. Doel is niet helemaal duidelijk van deze clausule. Je kan er ook makkelijk rond werken. Doel is vooral om semantische afstemming te bereiken tussen partijen eens ze data gaan uitwisselen. De regel over concept roept eigenlijk meer vragen op dan dat hij duidelijkheid schept. Beter te verhuizen naar de richtlijn.

- **BESLISSING:** Aanpassing van de tekst onder Concept. De beperking van de letterlijke overlap wordt geschrapt.

DISCUSSIE: Versionnering op de URI? Is dit nodig? Wat bij de komst van een volgende versie van deze

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

standaard. Moet de URI dit uitdrukken? Weinig buy-in om dit door te voeren. **BESLISSING:** we doen dit voorlopig niet.

3 BESLUIT EN ACTIES

Beslissingen:

Ivm Uitzonderingen op de URI-standaard:

- **BESLISSING:** passage onder uitzondering 1 in de standaard aanpassen zodat migratie volgens eigen keuze (planning) wel toegelaten wordt en dat de 'oude' systemen die reeds persistente URI's aanbieden deze kunnen blijven behouden.

lvm checklist

- We passen document niet verder aan voor verduidelijking van wat een legacy systeem is.
- Beslissing om in document een simpele redirect toe te laten.
- Aanpassing van de tekst onder Concept. De beperking van de letterlijke overlap wordt geschrapt.
- Geen versionnering op URI expliciteren.

Volgende stappen / To do's:

- Aanpassen standaard document en op Google Drive zetten als versie 0.10 (Laurens De Vocht)
- Volgende sessie (februari 2017): overlopen aanpassingen + overlopen **URI-richtlijn** van Informatie Vlaanderen ter info.

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