INFORMATIE VLAANDEREN

/// Report

URI STANDARD FLANDERS

Programme /// OSLO²

Version /// 0.9

Date of Publication /// 01/07/2017

Vlaamse overheid

www.vlaanderen.be/informatievlaanderen

Informatie Vlaanderen

Hoofdzetel
Boudewijnlaan 30, 1000 Brussel
+32 (0)2 553 72 02

Regionale zetel
Koningin Maria Hendrikaplein 70, 9000 Gent
+32 (0)9 276 15 00

oslo@kb.vlaanderen.be

Contents

1 ABOUT

1.1 REVISION HISTORY

Current status: draft

Revision	Date	Contributor(s)	Comments
0.13	05/10/2016	L. De Vocht M. Van Compernolle	Dividing the normative and guiding section
0.14	31/10/2016	L. De Vocht M. Van Compernolle	Processing feedback working group 11/10/2016
0.15	18/11/2016	L. De Vocht M. Van Compernolle	Processing feedback working group 8/11/2016
0.9	13/12/2016	L. De Vocht M. Van Compernolle	Processing last call for internal feedback 30/11/2016

1.2 AUTHORS

Name	Role	Organisation
Dirk De Baere	Reviewer	Informatie Vlaanderen
Geraldine Nolf	Reviewer	Informatie Vlaanderen
Erik Mannens	Reviewer	Imec – UGent – IDLab
Dieter De Paepe	Author	Imec – UGent – IDLab
Laurens De Vocht	Author	Imec – UGent – IDLab
Mathias Van Compernolle	Author	Imec – UGent - MICT
Ziggy Vanlishout	Reviewer	Informatie Vlaanderen
Raf Buyle	Reviewer	Informatie Vlaanderen

Note of the authors:

- Explanation of the used symbols can be found in the back of this document
- This document focuses on the guidelines for defining the URIs and illustrates the exceptions.

2 MANAGEMENT SUMMARY

This document describes the Flemish URI standard for the disclosure of resources, information and data, by the Flemish authorities and has been specified by Information Flanders. A checklist is included at the end of this document, which can be used as a self-evaluation instrument for the evaluation of URIs.

The guidelines concerning the governance of persistent URIs were developed in the context of the Open Standards for Linked Organizations program (OSLO²). OSLO² is an interoperability program in the Region of Flanders. The program brings together expertise from different business domains and governmental levels, independent of a specific thematic project. The Flemish Government developed a domain model in line with international standards including ISA and INSPIRE[1] enriched by data extensions to comply with the local context. The formal specification is published at data.vlaanderen.be[2].

The standard defines the minimal requirements[3] to which the Flemish authorities should comply in order to disclose resources in a uniform manner via the web. In addition, this document describes how 'Uniform Resource Identifiers' (URIs) should be defined in order to identify these resources in a persistent way. This enables back-office systems to evolve while the URI remains stable. This document also applies to data disclosed via services, but does not necessarily apply to the services themselves, given that persistence for services is hard to realise and it is inefficient.

All persistent URIs must be formed according the following pattern:

`http(s)://{domain}/{type}/{concept}(/{reference})*`

The parameters are placed between **{}**; ***** indicates '0 or multiple times'; **()** indicates a group.

Domain: This part of the URI pattern is formed by the *hostname* and is possibly preceded by a *subdomain*. The guidelines concerning the management of subdomains are outside the scope of the URI standard but are part of the guidelines for the application of the URI standard. Domains must use neutral, organisation-independent and timeless terms to cope with changes of names.

Type: This term describes the nature of the underlying resource. For the *type*, as part of the URI pattern, a classification is proposed with at least the following terms: **id** (*identification*), **doc** (*document*) and **ns** (*namespace*). The aim is to make a distinction between the representation on the web and the actual concept or object in the real world.

[1] http://inspire.ec.europa.eu/

[2] http://data.vlaanderen.be/ns/

[3] Note that there are no language or content guidelines to which the terms, as part of the URI, must comply. Every public service or organisation is free to develop a policy according to their own needs, make arrangements and document this.

Concept: This term represents the category of the resource (e.g. by means of a hierarchical classification). The *resource-category* obtains its meaning in the context of a *domain* and should be interpreted as follows: {*resource*} is (a) {*concept*}-{*type*} (e.g. .../**id/waterway/schelde** is a **waterway-identification**)

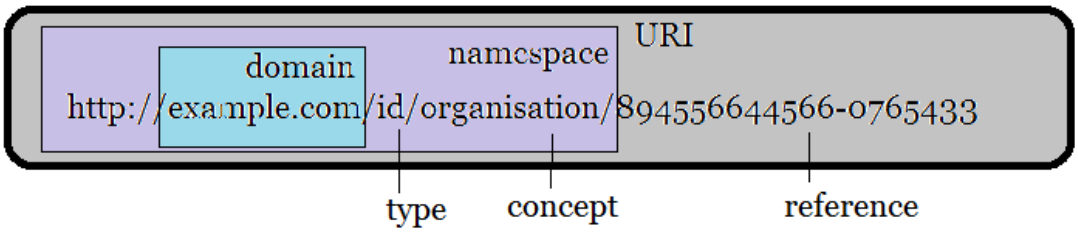
Reference: refers to one particular instance of a resource and is formed by the following sub-pattern: {*reference-basis*}(/{*reference-version*})?

3 INTRODUCTION

In the digital society, it is expected from governments to disclose databases with public sector information over the web between governments and third parties following a widely supported methodology. To ensure structured and identifiable processes, for both computer, machines and humans, URIs provide a solution. To ensure a common understanding, we will first clarify the following concepts: *resource*, *domain name*, *namespace* and *URIs*. Furthermore, this document describes the minimum requirements a persistent Flemish URI should comply with in order to be resolvable by the World Wide Web (=web). In this way, it will work anywhere in the world.

Note that there are no language or content guidelines to which the terms, as part of the URI, must comply. Every public service or organisation is free to develop a policy according to their own needs, to develop agreements and document these.

The example below shows the composition of a URI, containing a domain, namespace and a reference.



In the example above:

- The domain: example.com
- The namespace: http://example.com/id/organisation/
- The URI: http://example.com/id/organisation/894556644566-0765433

Type, concept and reference are explained more in detail in Chapter 4.

3.1 RESOURCE

The term 'resource' was first introduced to refer to specialized pages and content within a website (URL)[4], more specifically to the destinations of UR**L**s (e.g. the contact page of the website of a company), but later the definition was expanded to cover everything to which a UR**I** can refer[5] (e.g. the address with the geolocation in Google Maps from the same contact page).

For a clear understanding of these terms, we distinguish two groups of resources: information resources and non-information resources.

1. **Information resources:** objects that (may) have a digital representation, such as
    a. **Data**: representation of objects or things stored in any electronic format on a storage medium;
    b. (web)-**services**: allow access to data, taking actions with data (= modifying the data. (web)-services are services that enable a machine to provide access to data and allow to perform actions on and with the data;
    c. **Ontologies** (or vocabularies): terms, their definitions and the relationship between the terms to describe data;
    d. **Documents**: describing (non)-information resources on the web;
2. **Non-information resources:** things, objects or events from the 'real' world without representation on a computer, but for which the description can have a representation on a machine. The description of non-information resources are included in **documents**.

    E.g. the report of a meeting is a document (= the report) describing a non-information resource (= the meeting).

3.2 URIS

A URI is a formal way to refer to a resource. The most known format of a URI is a URL, which can be seen as an equivalent of a web address. An URL is a URI that makes use of the well-known HTTP URI scheme. Every URL is a URI (of an existing resource on the web)[6].

[4] RFC3986. https://tools.ietf.org/html/rfc3986

[5] RFC3987. https://tools.ietf.org/html/rfc3987

[6] RFC3986 - http://www.ietf.org/rfc/rfc3986.txt

Further, in this standard, it will turn out that a URI must be 'resolvable' to be able to retrieve the resource. This is achieved by a service that interprets the requested URI, and replies either with a description of the content or refers to another service that can describe this URI. All resources should be uniquely identifiable.

3.3 DOMAIN NAME

A domain name (or domain) is a name in the Domain Name System (DNS), used for the identification of computers such as web servers, services and applications. The domain name is needed to retrieve things on the web provided via computers. This term of the URI pattern is formed by the host name, possibly preceded by a subdomain.

3.4 NAMESPACES

Namespaces are a way to define, among others, concepts and variables and to categorise them in order to be unique in the group they belong to. They are constructed based on a domain name, to which one or more additional clarifying terms can be added.

It is very common to define a namespace via the HTTP(S) URI scheme. The HTTP(S) URI scheme is a particular scheme format to shape a Uniform Resource Identifier (in short, URI[7]). This format is very popular as URIs are managed by a certain organisation, thus they are well differentiated among each other.

The table below[8] shows the difference between the construction of a URI based on a domain name and a namespace, each time for the example 'data', 'service', 'non-information resource' and 'ontology'.

Type of resource	Description	Domain name	Namespace	URI
Data	List of institutions of the EU in Brussels	demo.thedatatank.com	http://demo.thedatatank.com/brussels	http://demo.thedatatank.com/brussels/european_institutions
Service	Flemish Open Data Portal	opendata.flanders.be	http://opendata.flanders.be	http://opendata.flanders.be/dataset?q={searchterm}
Non-information resource	The Atomium, a monument in Brussels	dbpedia.org	http://dbpedia.org/resource/	http://dbpedia.org/resource/Atomium

[7] RFC3986, https://tools.ietf.org/html/rfc3986

[8] These examples refer to existing URIs as illustration and determine how they are structured at the moment

Ontology	The 'Organization' ontology	www.w3.org	http://www.w3.org/ns/org#	http://www.w3.org/ns/org#Organization
Document	Page describing the Atomium, a monument in Brussels	dbpedia.org	http://dbpedia.org/page/	http://dbpedia.org/page/Atomium

4 RULES

In order to define the basic concepts of the URI standard - namespaces, identifiers, versions - in a uniform manner, a number of guidelines should be followed.

The context and the source for these guidelines on which the guidelines are based or inspired are always mentioned.

Rule 1: the HTTPS or HTTP URI scheme must be used as the basis for the definition of the URIs.

Rule 1 is based on international Internet standards and practices. The Request for Comments (RFC recommendations) of the Internet Engineering Taskforce (IETF) recommend the URIs to follow[9] the HTTPS[10] or HTTP scheme[11]. These URIs can be requested via the HTTP(S) GET protocol[12]. In recent years, there is a growing consensus within various standardisation communities regarding the use of HTTPS or the HTTP URI scheme[13] (OGC, INSPIRE, W3C). The HTTPS or HTTP URI scheme[14] is one of the possible URI schemes allowed by IANA (Internet Assigned Number Authority)[15].

The pattern for forming the URI is based on the guidelines from INSPIRE[16] and ISA[17].

Rule 2: All URIs must be defined following this pattern:

[9] https://www.w3.org/TR/ld-bp/#HTTP-URIS

[10] RFC2818 - https://tools.ietf.org/html/rfc2818#section-2.4

[11] RFC7230 - http://tools.ietf.org/html/rfc7230

[12] HTTP status code 302 "Found" can also be used here.

[13] See INSPIRE "Recommendation 8" (in INSPIRE Guidelines for encoding of spatial data (D2.7), version 3.2)

[14] RFC3986 - https://tools.ietf.org/html/rfc3986

[15] See IANA http://www.iana.org/assignments/uri-schemes/uri-schemes.xhtml

[16] INSPIRE "Recommendation 10" (in INSPIRE Guidelines for encoding of spatial data (D2.7), version 3.2) states: "Every Member State and the European Commission should develop, document and maintain a URI scheme for their resources". INSPIRE Guidelines for encoding of spatial data (D2.7), version 3.2

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

http(s)://{*domain*}/{*type*}/{*concept*}(/{*reference*})*

This convention has the advantage that URIs can be defined consistently. It is important that the URIs are described in a formal way, hence the following guidelines:

Rule 2.1: {*domain*}, {*type*} and {*concept*} are **mandatory** parts of the URI.

Rule 2.2: *References* are **optional** and there can be more than one depending on the situation.

Rule 2.3: The basic combination {*domain*}/{*type*}/{*concept*} in the absence of a (/{*reference*})* may only refer to one resource (=unique on the web) and must be invariable (since {*reference*} is optional).

Rule 2.4: The (/{*reference*})* in combination with {*domain*}/{*type*}/{*concept*} must be unique and constant on the web. This means that once the {reference} is assigned, the reference itself as well as the semantic description of the resource cannot be changed.

4.1 DOMAIN

Rule 3: The name of the domain and all potential subdomains, must all be independent of the organisation, product, brand and time.

The domain cannot contain a name that may seize to exist. After all, we want to avoid using a name in the domain name that can disappear or go out of use, because the domain name as a whole must always keep on existing to maintain persistence.

4.2 TYPE

Rule 4: the {type} as part of the URI pattern is mandatory and is the main classification for resources. The {type} makes a distinction between (1) the actual object/concept, (2) the digital or web-representation and/or (3) a term belonging to a vocabulary or ontology.

The {type} as part of the URI pattern says something about the nature of the resource description and follows a classification which contains at least the following terms to make a clear distinction.

1. **id**: *identifier* is a reference to an object from the real world or an abstract concept;
2. **doc**: *document* is a representation on the web or a description of real-world objects or abstract concepts. It deals with general descriptive information (web documents).
3. **ns**: *namespace* of a taxonomy, ontology or vocabulary[18].

The parts in bold – the abbreviations – must be used as the '{*type*} component' in the URI pattern. Additional types are possible and should be implemented following the agreed guidelines that apply to the domain.

[18] This is equivalent with "def" that is used in other places (e.g. the UK URI-strategy, p.7: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/60975/designing-URI-sets-uk-public-sector.pdf).

Rule 4.1: The URI of a non-information resource with type *id* is required to refer with redirections (303 HHTP-redirect[19] or #id fragment identifier[20]) to a digital web representation with the type *doc*.

Examples

id/waterway/schelde

is an identifier for the real world object 'Schelde'. This object cannot exist on the web, but it can be referred to from other applications or data.

Remark: if the waterway (real world object) 'Schelde' ever changes its name or would no longer exist, its URI must remain by either referring to the old object 'Schelde' or by referring to the same object with the new name.

doc/waterway/schelde

- refers to a document (e.g. HTML page) containing an explanation about the river 'Schelde'. This document can have structured data, but this is not a requirement.
- describes meta-information such as the history, versions, source, and details of the actions made on all representations of {*type*}//*waterway*/*schelde*.

ns/waterway

Is the namespace that can be used in the vocabulary dealing with waterways.

E.g. *ns/waterway#depth* or *ns/waterway/depth* refers to the term *depth* from the namespace *ns/waterway* and serves to indicate the depth of the waterways.

4.3 CONCEPT

Rule 5: The {*concept*}, as mandatory part of the URI pattern, represents the category of the resource. The categorisation can be done by means of an agreed (hierarchical) classification or list, but this is not mandatory. Overlaps with terms used for type and reuse of the terms from the domain name are not allowed.

Examples

id/waterway/schelde

id/school/s_7838483

id/address/9039439430

id/property+45a-750bk-ZB85h-dz385

~~http://www. waterweg en.be/id/waterweg/schelde~~

[19] https://www.w3.org/TR/cooluris/#r303gendocument

[20] https://tools.ietf.org/html/rfc3986#section-3.5

http://www. waterweg en.be/id/rivier/schelde

4.4 REFERENCE

Rule 6: the (/{*reference*})* as part of the URI pattern refers to one particular resource. The {*reference*} element can appear multiple times to create a hierarchical structure.

Rule 7: Every {*reference*} can be completed in two different ways: {*reference-basis*} or {*reference-basis*}/{*reference-version*}.

The *reference-basis* is the part of the reference that refers to the resource. The *reference-version* is the version of this resource that is managed. There should always be a *reference-basis* in the reference, but not always a *reference-version*. The standard says nothing about how these components are implemented. The interpretation of the *reference-basis* and *-version* is organisation dependent (but also data management dependent) and the potential URI guidelines within organisations will clarify how these terms are implemented. Rule 7 thus leaves the choice open to use only the *reference-basis* or both the *reference-basis and the reference-version*.

5 EXCEPTIONS

One **can** (but is not obliged to) deviate from the rules in the following situations:

1. When it comes to **services**, the guidelines are not applicable, although they are strictly speaking a resource. If the data coming from services must be persistent, they have to follow all the rules from the URI standard. Therefore it makes little sense for services that are disclosing data, to additionally enforce persistence and/or enforce the rules. Moreover, persistence for existing services is not evident: existing services can possibly not be adapted anymore to the URI standard, or existing URIs can already be constructed in a well-considered manner (according to their own rules). For new services and services currently in development, it makes little sense to enforce persistence and the guidelines. After all, in these cases where persistence is required, the disclosed data already have to comply with the rules from the URI standard.
2. HTTPS or HTTP URIs that are already published (**legacy**)and were created with the aim to be persistent, must remain preserved. This does not mean that it is allowed for legacy-systems to produce new, non-conforming URIs. The only goal of this exception is to guarantee the persistence of the URIs already in use by third parties.
3. Components with type '**ns'** may use fragment identifiers[21] in order to define all terms from one particular vocabulary into one web document, each in their namespace. The fragment identifier

[21] https://tools.ietf.org/html/rfc3986#section-3.5

component of the URI allows adding an indirect identification after the '**#**' of the resource described in the corresponding resource identified before the '**#**'.

Example: fragment identifier

/ns/artefact

Refers to the namespace in which the artefacts are described

/ns/artefact#Relic

Refers to the specific term in which the fragment identifier *Relic* is used in order to refer to it within the namespace *ns/artefact*

/ns/artefact#Relic is an alternative for */ns/artefact/Relic* that (i) exists on its own; or (ii) executes a 303-redirect to the namespace */ns/artefact* or (iii) redirects to a document describing e.g. */doc/artefact/Relic*.

4. An alternative way of guideline 4.2 to make the distinction between *id/doc* is by introducing a fragment identifier[22] (the part after the square bracket '[' or hash '#' in the URI). When a user resolves this URI, the fragment identifier will be ignored (as part of the HTTP protocol) and will request another resource[23]. This other resource represents the information resource. **This approach can only be used in the case of sufficient reasons not to follow the recommended approach of rule 4.2**. In this case, the {*type*} *doc* is used in the URI structure instead of *id,* and the URI must be followed by the fragmented identifier '#id'. **It is prohibited to produce URIs following both the redirect and the hash manner for resources of the same type.**

Example: /doc#id - /doc vs. /id - /doc

Resolving */doc/person/2da46001b3#id* results in the descriptive document */doc/person/2da46001b3*.

This is an alternative for the method from guideline 4.2:

/id/person/2da46001b3 that performs a 303-redirect to */doc/person/2da46001b3*.

[22] https://tools.ietf.org/html/rfc3986#section-3.5

[23] https://www.w3.org/TR/cooluris/#hashuri

6 THE URI STANDARD COMPLIANCY CHECKLIST

This section describes a framework to evaluate URIs. A URI is 100% compliant with the Flemish URI standard if all questions can be answered with 'yes'.

PRECHECK: In case the URIs are provided via a legacy service that guarantees the persistence via the domain and uses the http(s) protocol, then the existing URIs should not be compliant with the following checklist. This checklist can help to gain insights into how to improve the format of URIs. All other and new URIs should follow this checklist.

Rules
Does the URI scheme make use of the http(s) protocol?
Does the URI scheme follow the structure: {*domain*}/{*type*}/{*concept*}(/{*reference*})*?
Domain
Is the existing (sub)domain name independent of **organisation, product, brand** or **time**?
Is the domain name guaranteed persistent: can it be guaranteed that the domain name will last forever and will never change?
Type
Is the ***{type}*** included as part of the URI pattern **AND** is there at least a distinction between representation, the actual object/concept and a term belonging to a vocabulary, thesaurus or ontology?
Does the ***{type}*** of all URIs in the domain and subdomain follow the same strict classification including at least **id**, **doc** and **ns**?
Concept
Is the ***{concept}*** as part of the URI pattern unique, meaning that there is no literal overlap with the terms appearing in the domain, type or references?
Reference
Except in the type "**ns**', no fragment identifiers are used.
Does (/{***reference***}) as part of the URI pattern refer to one particular instance of a resource?
Can the {***reference***} be completed in one of the following ways: {*reference-basis*} or {*reference-basis*}/{*reference-version*}?
Is the assigned {***reference***} persistent and is the whole in combination with {*domain*}/{*type*}/{*concept*} unique on the web?

7 WORD LIST

Term	Definition
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 Modelling Language
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
URN	Uniform Resource Name
UUID	Universally Unique Identifier
W3C	World Wide Web Consortium

8 LIST OF SYMBOLS USED IN THE URI PATTERN

{ }	URI parameter
( )	Selector to indicate a section within an URI pattern
*	0 or multiple times the foregoing section
+	1 or multiple times the foregoing section
?	0 or once the foregoing section