Towards a national URI-Strategy for Linked Data of the Dutch public sector

Contents

About this document

Authors

- Hans Overbeek (KOOP) [1]
- Linda van den Brink (Geonovum) [2]

Translation

- Nancy Christiaans, Nijmegen

This is a translation. The original text [29] is in Dutch.

Realization

This article was written further to the Pilot Linked Open Data (PiLOD) on the part of Geonovum in 2012 and 2013 [3]. It describes the insights as acquired by the ‘Working group URI-strategy' within the action line ‘Technology’. The working group was established for the purpose of examining whether a national URI-strategy for Linked Data of the Dutch public sector is useful and, if so, which ingredients should make up part of such a strategy. The members of the working group conclude that there is a need at numerous implementing authorities for guidelines for formulating URIs. Linked Data is a new field; it involves quite some complexity and includes cross-domain problems. Both technical as well as organizational. This document is not a URI-Strategy as yet, but it does provide an inventory of the issues and a few recommendations.

The most important sources of knowledge that the working group used, in order to come to this initial impulse to a national URI-strategy, are:

- The Inspire directive, which dictates a national strategy for URIs for geo-information, with the recommendation to link this geo-strategy to a generic national strategy [4]
- Designing URI sets for the UK Public Sector. A recommendation of the British government, who are considered frontrunners in the area of the publication of Linked Open Government-Data [5]
- 10 Rules for persistent URIs. A comprehensive EU report with comparable initiatives and a valuable overview of the most recent best-practices [6]

Numerous international experts in the sphere of Linked Data were present during the W3C workshop 'Open Data on the Web' [7]. We presented [9] a paper [8] during this workshop that contained our insights and questions. This resulted in valuable feedback, particularly on the part of Jeni Tennison [10], who contributed to the Linked Data guidelines of the British government. This alone demonstrates the importance of keeping up with European and other international developments. By participating in European and international programmes, we gain access to international knowledge and expertise, we stay abreast of the most recent insights and future developments and we can even influence the development of standards so that these develop themselves in keeping with the Dutch needs.

The working group URI-strategy has attempted to summarise the knowledge from the sources referred to above and to apply this knowledge to the Dutch situation. We developed different views in the working group with respect to certain aspects, which were in part confirmed by the experiences that have meanwhile been acquired in the UK.

Continuation

This article has been translated into English so that we can also request international feedback. Afterwards, it will remain available as a stable version and end product of the PiLOD. At the most, errors will be corrected.

Any ideas, comments, feedback is appreciated and can be sent to hans.overbeek@koop.overheid.nl.

We hope that this initial impulse is to be continued in the form of an actual URI-strategy which is supported by policy frameworks. During the final meeting of the PiLOD, this article was presented to Nico Westpalm van Hoorn, chairman of the Forum Standaardisatie[1] [11] along with the request to have the policymakers pay attention to the developments in the field of Linked Data and the importance of a URI strategy that is supported throughout the government.

Working group

The authors wish to thank all of the participants in the Working group URI-strategy, without whom this initial impulse to a national URI-strategy could not have come about:

- Thijs Brentjens (Geonovum)
- Wilko Quak (Geonovum)
- Paul Hermans (ProXML)
- Hayo Schreijer (KOOP)
- Michel Grothe (Geonovum)
- Marcel van Mackelenbergh (Tax authorities)
- Jan Jelle Boomgaardt (Digimelding)
- Rob van Dort (Mapplica)
- Arjen Santema (Land Registry)
- Bart van Leeuwen (Fire Brigade Amsterdam-Amstelland)
- Marco Brattinga (Land Registry)

and other participants in the PiLOD.

[1] Forum Standaardisatie is the standardisation body for the Dutch government.

Background

Linked Data

Much has already been said about the added value of Linked (Open) Data for the government and the Dutch business community. The most important advantages include a more efficient operational management on the part of the government (cost reduction and prompter action); reliability and transparency of the government (clarity and accountability), and the economic value of data.

In practice, however, the yield often falls short of expectations or promises. An often-heard complaint is that the data is difficult to find, and that it cannot be easily linked to other data. These problems are counteracted by copying data of others, involving heavy expenses for collecting, converting and synchronising the data, or by building expensive national provisions. The result is an abundance of copies and much doubt regarding the authenticity of the information.

A better solution would be to make the authentic data permanently available so that everyone can use it. This requires equipping the data with a reliable form of identification, so that you can refer to the data and can also understand the references of others. These direct references to authentic data result in more coherence and improved traceability and render copying and synchronising redundant. This method is referred to as Linked Data. [12] The Basisregistraties[2] have already been set up to serve as an authentic source for the reference-data. What is lacking is a good strategy for opening up these sources such that they are also machine-readable. There is a need for a kind of Velcro that can be used to join datasets together without effort.

The aim of the Pilot Linked Open Data (PiLOD) [13] of Geonovum was mainly to examine what is needed in order to convert 'Open Data' into useful 'Linked Data'. The PiLOD showed the need for a strategy for the identification of authentic government data, starting with the base registrations. An initial impulse to a strategy of this kind resulted from the Working group 'URI-strategy'. URI stands for 'Uniform Resource Identifier', a standard for the identification of objects and concepts ('everything is a resource') of the W3C [14]. The base registrations and other authentic sources for reference data, such as the collections for legislation and regulations on *overheid.nl*, for example, are at the centre of this strategy. After all, identifications of authentic reference data make up the 'links' that are used to generate Linked Data.

Example

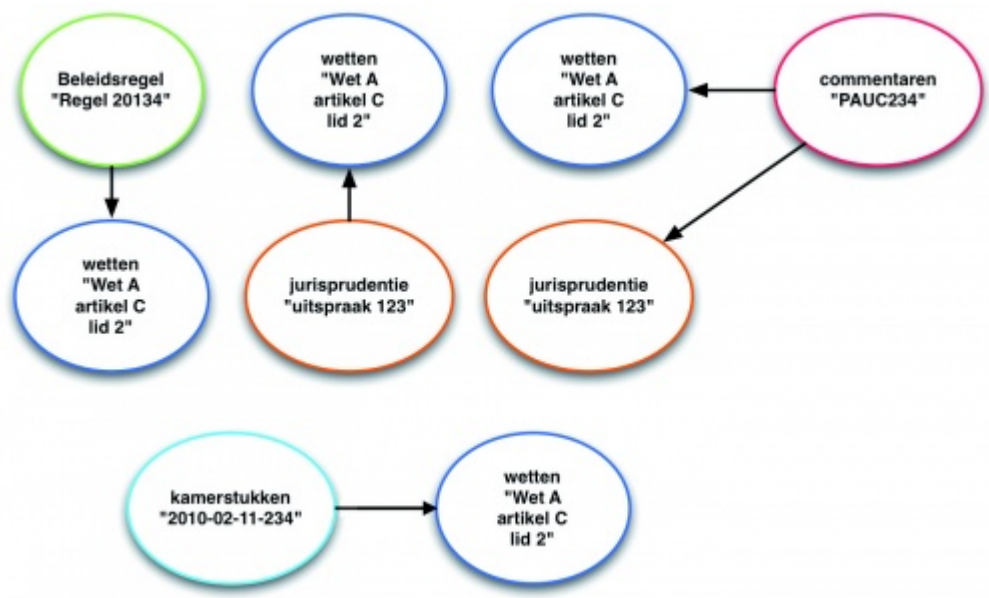
This is illustrated in the example below:

- Take the case in which an implementing organisation formulates a policy guideline (in Dutch: beleidsregel) for the implementation of policy. These policy guidelines have a

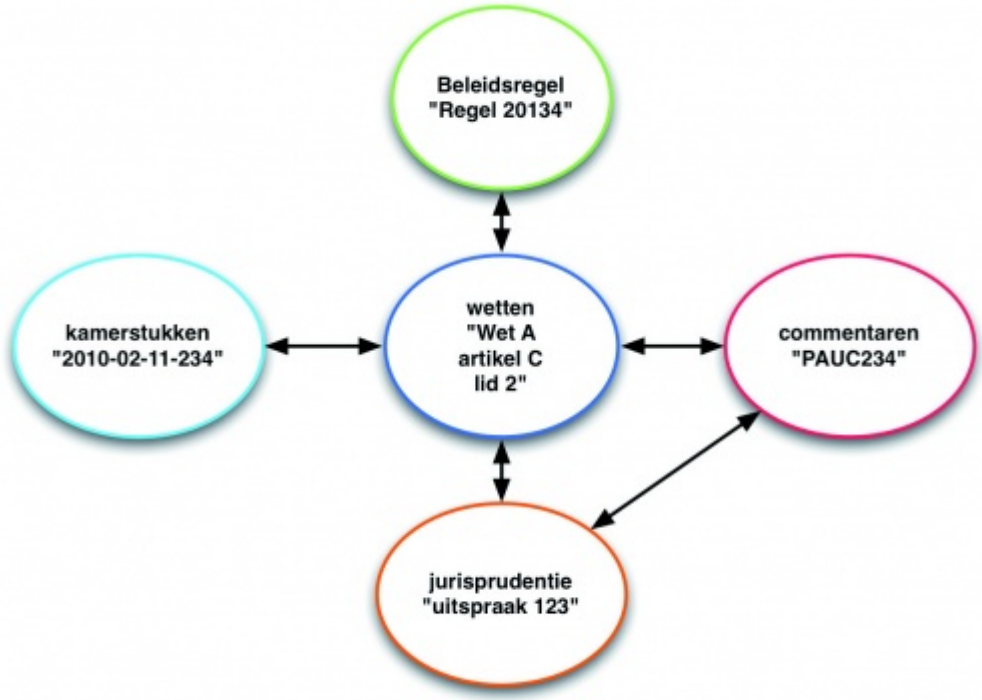
[2] The Basisregistraties (Base registrations) are 13 national registers for reference objects such as cadastral objects, companies and organisations, inhabitants, etc. which are defined authentic resource for specific data by law.

legal basis for which the implementing organisation refers to the appropriate paragraph in an article of a law (in Dutch: wet). The policy guideline is published on their website.

- A judge delivers a judgment (in Dutch: uitspraak) in a lawsuit and bases the judgment on that same article. The decision is published on *rechtspraak.nl*.
- A jurist writes a comment (in Dutch: commentaar) further to this decision and refers to the same article of a law. Her comment is published in a magazine on legislation.
- Finally, the article of the law is discussed in the Parliament and the proceedings (in Dutch: kamerstukken) are published on the parliamentary website.



By standardizing the reference to the article in the law, the collections of policy guidelines, judgments, comments and parliamentary documents can be easily linked to one another. We can now easily relate a number of documents from other collections to each of these information objects.



The example demonstrates that a uniform identification of the reference-objects (the URIs) - articles in laws and judgments in this example - is essential in order to be able to create the links between the information objects. These make up the hooks and loops of the Velcro that is used to link the collections of information to one another.

We can make another comparison, namely to the IBAN-numbers in the financial world. After years and years of problems further to international and inter-bank monetary transactions, the financial institutions are finally implementing a standard for the identification of bank accounts. This is a lengthy and costly operation. Similar problems threaten databanks: links are being created between collections of data at numerous locations and each time each link is devised from scratch. We, on the other hand, have the opportunity to timely develop a strategy with which to give the objects from our authentic databanks (registers) a standard 'bank account number', so that we can solve problems at an early stage, or even stay ahead of problems, and prevent costly restructuring later on.

Added value of a URI-strategy

A clear-cut strategy, formulated in consultation with the stakeholders, must ensure that the parties that wish to set to work with Linked Data can make the sound choices that are needed to generate Linked Data-solutions. Ideally, the URI-strategy, oriented towards the technical implementation of the identification of authentic data, should be embedded in a broader Linked Data Strategy, in which organisational aspects are considered as well.

A good strategy offers added value for numerous projects that are already underway.

- Stelselcatalogus 2.0 [15] has abandoned the idea of a single umbrella model for all of the base registrations and opts to evaluate the – inevitable – differences that exist between the models of the base registrations respectively. Linked Data is found to be

more suitable for this approach compared to the traditional methods for data modelling.

- OWMS, the metadata standard for Dutch public sector information on the web supports both text-values (labels, for example creator='Utrecht'), as well as the use of identifiers (pointers to more information, for example creator=http://standaarden.overheid.nl/owms/terms/Utrecht_(gemeente)). The use of identifiers yields references that are much more accurate and offers more possibilities than labels.
- Data-collections of the public sector [16] that are published as Linked Data can be linked to other datasets and are therefore put to better use and used more often than datasets to which no links are possible. The latter must be copied in order to be reused.
- In the Linked Data Public Sector (LiDO) project at KOOP, laws and regulations are used a binder for Linked Data. This increases the coherence and the traceability of public sector information, making the integration of content easier.

The URI-strategy

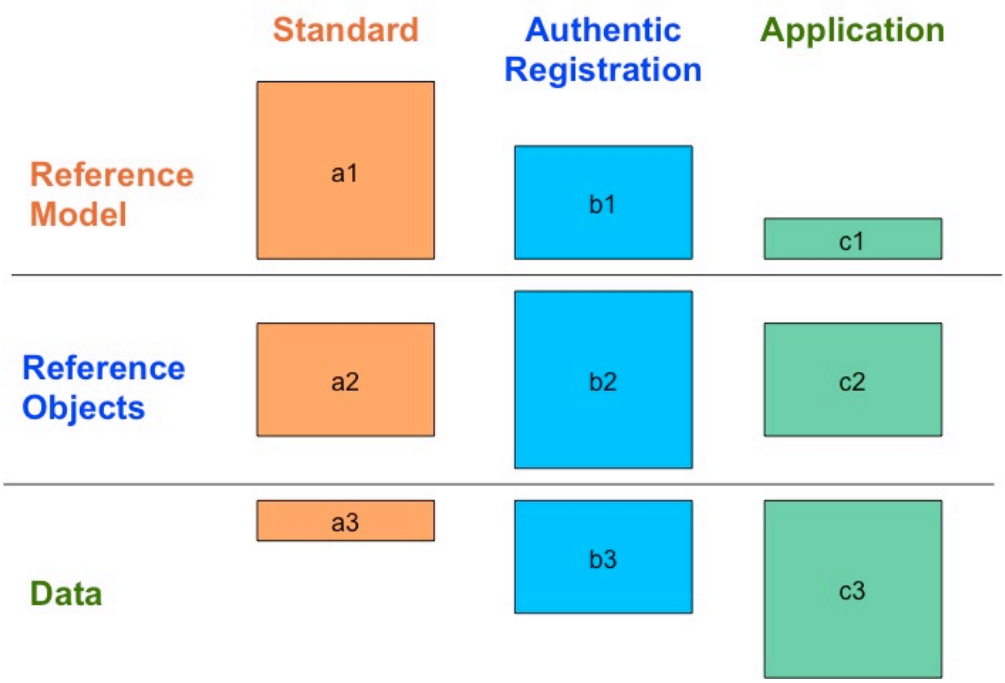
Scope

The URI-strategy is mainly intended for data that is used to define objects or concepts, to which other applications can refer. Data to which no links exist is out of scope. To illustrate this, we distinguish between three categories of sources of information:

1. Standards
2. Authentic registrations
3. 'ordinary' Applications

Each with an emphasis on one of three categories of concepts:

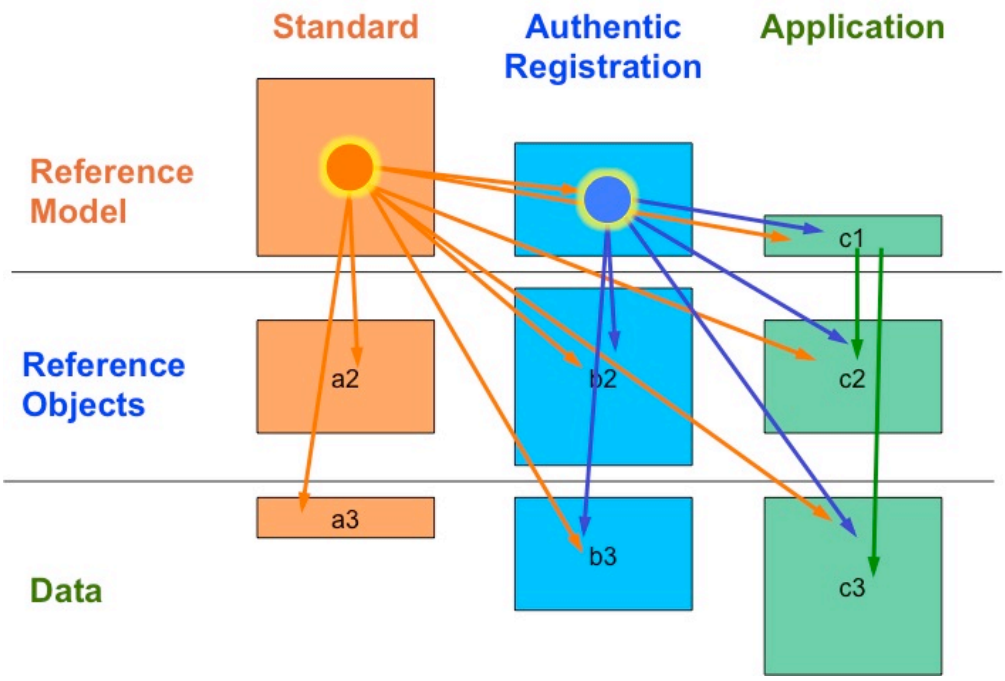
1. Terms in a conceptual Model
2. Reference objects
3. 'ordinary' Data



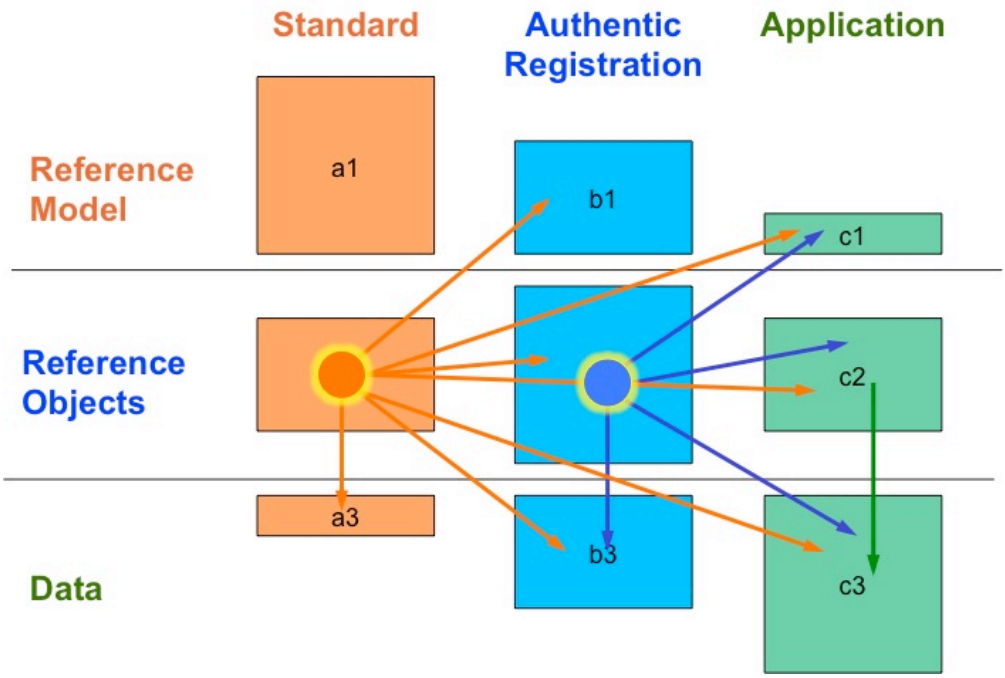
The size of a cell in the diagram is an indication of the importance of the concepts in that category of sources of information. The most important function of a standard is usually to define a conceptual model. (a1) Authentic registrations are generally set up to keep an account of the Reference objects (b2) and the function of an 'ordinary' Application is generally merely to collect Data for a specific goal (c3).

Of course, an Authentic registration and an Application can have a Model of their own (b1 and c1), some Standards provide a list with Reference values (a2) and Applications may have local Reference data (c2). In addition, Standards and Registers may require some 'ordinary' data (a3 and b3), for the purpose of recording changes and the origin of the reference data, for example.

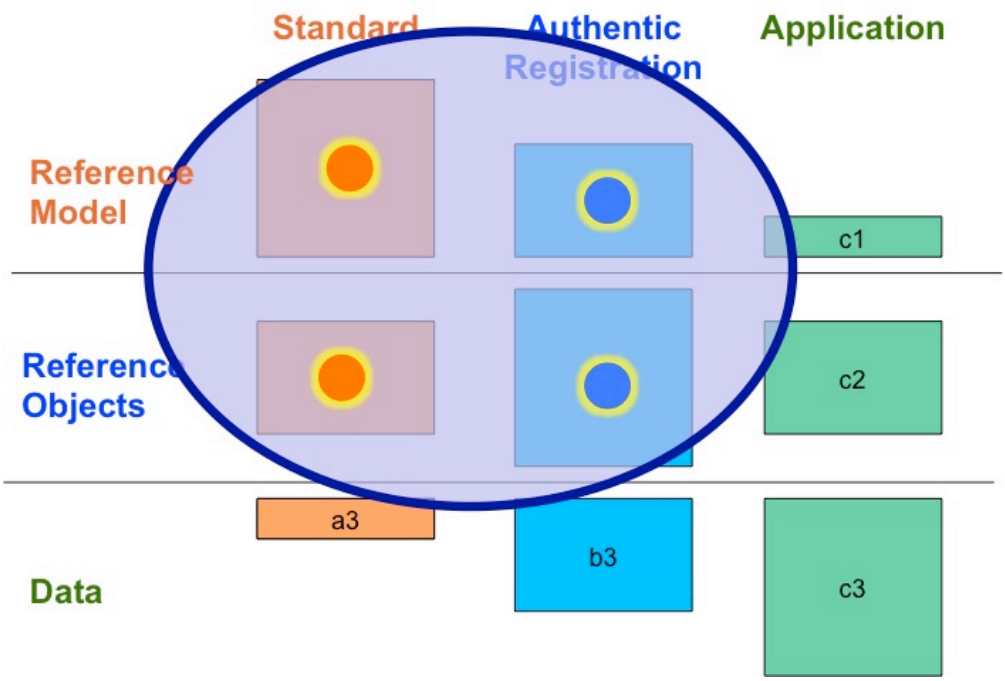
The URI-strategy supports the reuse of Concepts and Reference objects by other data-collections. And so the interesting categories are the terms in the Models and the Reference data.



Terms, such as classes and properties, that are defined in the Models of a Standard or an Authentic registration are used to categorise Reference objects and Data.



Reference objects, defined by Standards (value lists for example), but particularly those that are administered in Authentic registrations, are used in 'ordinary' Applications.



The URI-strategy is intended for Models and Reference objects of both Standards and Authentic registrations. And so not in first instance for the 'ordinary' Data (row 3) or concepts in 'ordinary' Applications (column c), for the simple reason that their links are accessed less frequently.

Insights

The Working group URI-strategy gained a number of insights upon analysing the proposed alternatives during the PiLOD.

'No register, no identifier'

It is claimed fairly often in Linked Data theory that one must define a URI for every concept or object, making it seem that you cannot start until you have devised and minted a new Linked Data URI for each concept or object. But why define everything all over again? Mankind has been defining authentic identification for standard terms and reference objects for centuries and centuries. Consider, for example, encyclopaedias, taxonomies and registrations of inhabitants or real estate. In this context, we refer to a provision for the authentic definition and identification of concepts or reference objects as a register. And so, in this context, we understand a register to mean either a specification of terms/concepts in a standard or an authentic registration of reference objects.

The goal of this far from minor effort to set up registers is to be able to refer to accurate and more extensive definitions of abstract concepts and objects from different administrations in a univocal manner using a term as agreed upon (an identifier) so that everyone knows what is asserted. Information from different administrations can then be linked by – manually – grouping similar terms (usually names or numbers) together.

Now that we wish to automate that, there is no reason not to continue using these existing registers. But what if we want to refer to concepts or objects for which no registers exist? The

only way to mint a URI for concepts and/or objects that are lacking is by recording these in a new register. If we find that no register exists for certain concepts or objects, whereas we do desire URIs for these, then the only solution will be to establish a register. In short: you can only mint URIs for concepts or objects that are recorded in a register. We have summarised this significant insight in the adage: 'No register, No identifier'.

No mandatory shared Internet domain

The W3C recommends the use of http-URIs for identifying Linked Data. An http-URI is a URL that starts with 'http://', followed by a domain and, optionally, a local path. The British URI-strategy assumes that all of the Linked Data URIs of the government reside under a single main domain: 'data.gov.uk'. In order to maintain a degree of scalability, they propose to divide that domain into sectors. The sectors mentioned include, for example, location, education, transport and health. The corresponding domains are then: 'location.data.gov.uk', 'education.data.gov.uk', 'transport.data.gov.uk' and 'health.data.gov.uk'. This, however, leads to two problems:

1. For each of these sectors, it is necessary to find a party that is willing to be owner and administrator of the sub-domain concerned.
2. It is not always clear in which domain information is to be categorised. Do train stations belong to ‘location’ or to 'transport'?

And so it is difficult to determine the sector-owners and they will differ from one another in terms of the services offered and the procedures followed. But what’s more, the national main domain requires an owner as well. And with that, we would create both a single-point-of-failure as well as organisational dependency between the register holders and the holder of the central domain and the sector domains.

This is why we do not want to implement this UK-guideline in the Dutch URI-strategy. These problems have meanwhile been acknowledged in the UK as well. Moreover, they have met with problems further to the use of the main domain 'data.gov.uk'. One aspect of the advice that we were given in London during ODW13 was therefore to lay down by law every domain that is an intended component of permanent URIs.

Starting points

The Working group URI-strategy has formulated a number of starting points that should be observed upon drawing up the strategy:

1. **Link up with international best-practices.** You can go faster on your own, but you will go farther by working together. By linking up with international developments, you benefit from solutions that are devised on a global scale. In addition, European regulations are becoming increasingly important to the Dutch government.
2. **Link up with existing developments.** The strategy concerns many parties and systems and cannot be implemented all at once as something new. And so it is wise to assess what is already taking place in the sphere of standardisation and authentic registrations and to reuse that as much as possible.
3. **Anticipate deviating systems.** Even if systems are developed that, for whatever reason, do not observe the national strategy, it must still be possible to link to these systems.

4. **Keep it as simple as possible, but not simpler.** If the approach is too complex, then the strategy will not be adequately applied, or not applied at all. If the approach is too simple, then the strategy will not yield sufficient results.

Give sufficient thought to the following with respect to standardisation:

Persistence.
: Persistence means that solutions persist even if the organisation around them changes. Even if we have to accept that we do not know everything yet and that advancing insights may lead to different choices. Persistence does not mean for eternity, but a company or authority must feel confident to develop critical operational systems on its basis.

Scalability
: Scalability is important in order to keep the management costs calculable, even if the applications expand. No one can predict how many applications will be developed in the years to come. And so scalability must be taken into account upon setting up each component of the strategy.

Intelligibility.
: Intelligibility is essential in ensuring that agreements are easily understood and adopted.

Trust.
: Trust is a necessity in order to get organisations to strategically opt for the use and publication of Linked Data on their own accord.

Machine-readability.
: Machine-readability ensures that working solutions can actually be generated using Linked Data.

Human readability.
: Human readability is also important in ensuring that users trust and understand the solutions. But if the machine cannot make proper use of the data, then it will not work anyway.

… and preferably in that order.

URI-pattern

Following the three sources referred to above, we assume that http-URIs are the obvious choice. All three of the strategies depart from further agreements regarding the pattern to be used in constructing the http-URI. The pattern for http-URIs that is recommended in these sources – and which we have adopted for that reason - is:

```
http://{domain}/{type}/{concept}/{reference}
```

We will examine each of the four components separately below.

{domain}

The `{domain}` component contains the Internet domain and, optionally, a path within that domain:

`{domain}` = `{internet domain}/{path}.`

The {domain} serves two purposes. It is first of all an important instrument in obtaining unique identifications: two objects that are administered in two different databases can coincidentally be designated with the same identification (for example, a plot in the land register with ID 010101 and a legal person with ID 010101). Should both the Land registry as well as the *Nieuw HandelsRegister (NHR)*[3] decide to publish these objects as linked data, then two unique URIs will still be generated: one will start with http://brk.nl/, for example, and the other with http://nhr.nl/. Secondly, a well-chosen domain will ensure recognisability and trust. Plots in the Land registry with a URI such as http://data.brk.nl/perceel/010101 seem more reliable than http://data.findithere.eu/perceel/010101, for example.

The {path} can be used if various collections of objects exist within a register, in which double IDs may be present. The {path} can then be used to create extra name spaces.

Recommendations for the {domain}

1. A single task: the register

   The {domain} is preferably exclusively reserved for the publication of the register and for resolving the URIs of the register. The fact is, if the domain is part of a more extensive domain where other publications take place as well, then it may become necessary sooner or later to re-organise the publications, with all its consequences for the persistence of the URIs in the register.

2. No organisation name in the {domain}

   It is highly recommended not to include the name of an organisation in the {domain}, however tempting that may be from a marketing perspective. Again, persistence is the most important argument. The fact is, organisations can be split up, merged or renamed and they are then generally given a new name and will opt for a new Internet domain. Renaming the URIs disrupts the persistence. Continuing to use the old domain – which would be an acceptable option in a pure technical sense – may, however, give the impression that the data is obsolete as well. In general, registers will continue to exist for as long as they serve a certain purpose. If the register is indeed discontinued or is converted into a new register, then the models and reference objects in the old register are usually genuinely out of date.

3. Observe restraint with {path}

   Try to avoid the use of {path} as much as possible. The shorter the URI, the easier it is to use. The less information in the URI, the smaller the chance that it must be reconsidered later on.

{type}

The {type} indicates which kind of URI is involved. This may be:

'id'

[3] The NHR is the Dutch register of the Chambers of Commerce, containing all legal persons such as corporations and government bodies

identifier of an object (individual/instance) in a register.

`'doc'`

documentation (metadata) on the object in the register.

`'def'`

definition of a term in an ontology.

Recommendations for `{type}`

1. Use 303 redirect of the 'id'-URI to the 'doc'-URI.

   What is meant by this is explained in 'Cool URIs for the Semantic Web' [17] in section 4.2 [18].

2. Use Hash-URIs for terms included in the model

   It is sometimes difficult to distinguish between the model and the content in a Linked Data application. This distinction is generally clearer in a relational database: the tables and columns hold the model and the data in the tables make up the content. In Linked Data, however, a class can also be considered an instance (namely of the class rdfs:Class). In order to provide the user of a register with more clarity regarding which terms genuinely make up part of the model and which terms can be considered content of the register, it is recommended to define the URIs of the first as hash-URI (#-URI): `http://{domain}/def#{term}`. An additional advantage of this is that the URI `http://{domain}/def` yields all of the terms from the model.
   What is meant by this is explained in section 4.1 of 'Cool URIs for the Semantic Web'. [19]
   If the ontology is very extensive, then one may also opt not to make use of hash-URIs.

`{concept}`

The `{concept}` gives the human reader an indication of the type of concept that is identified by the URI. The `{concept}` is important for two reasons. First of all, it may offer a solution if there are objects within the registration that have no unique identifiers, but that are unique per type of object. The municipality of Utrecht and the province of Utrecht, for example. Secondly, and this is more important, it will yield a more comprehensible URI. A human reader may assume that `http://bagregister.nl/id/pand/01010101` is the URI of a building in the BAG.

A possible disadvantage of including a `{concept}` in the URI is that this gives meaning to the URI, whereas it is generally easier to make meaningless IDs persistent.

Recommendations for `{concept}`

1. `{concept}` means nothing.

   It is extremely ill-advised to ascribe any meaning to `{concept}` for the machine. URIs are opaque in a technical sense. [20] And so the `{concept}` is not necessarily the class to which an object belongs. This only helps the human reader, the manager of a semantic model for example, to recognise the URIs. [21] and [22]

2. One should also consider persistence when choosing the `{concept}`.

   If it is conceivable in a registration that object types (classes) can change names, while continuing to represent the same class, then it is not wise to include this component in the URI. You should include a higher class in such cases. For some, changing the type of an instance means per definition that the same instance no longer applies, but rather some other instance, of some other type. Example: take the case in which the Central Organ Asylum seekers Centre (COA) is transformed from an independent administrative body (zbo) to an agency. [23] And that we were to opt for: `{domein}/id/zbo/coa` as the URI of the COA. Following the transformation, this then becomes `{domein}/id/agentschap/coa`. If we were to opt for `{domein}/id/organisatie/coa` then we need not adjust the URI, but we can also no longer distinguish between the COA as a ZBO and the COA as an agency.

{reference}

The `{reference}` is the identifying name or code of the individual object. The URI strategy offers a lot of freedom in terms of `{reference}`, seeing that the requirements in various applications may differ widely. A `{reference)` may be: an identifying number, an alphanumerical code, a word or name, et cetera. Each register has a way to uniquely designate the individual objects in the collection. This unique designation can be included in the `{reference}`.

Recommendations for {reference}

1. Names or numbers?

   There is much discussion regarding the use of 'meaningless' identifiers versus 'meaningful' identifiers. So long as computers have no awareness, every URI is a meaningless string to the machine. There are cases in which a meaningless string can take on a meaning to people (020 is used by many who do not want to use the label 'Amsterdam' or 'Ajax', 013 (Pop podium in Tilburg), 9292 (Public transport information), number 14 (Johan Cruijff).
   Names or numbers, both have advantages. The advantage of numbering is that it appears to be more accurate and there can be no homonyms. But you sacrifice recognisability and manageability, without having the labels at hand.

   - In practice, the URIs for the concepts are meaningful in virtually all semantic standards and they generally include the complete label (name) used to indicate the term for humans (usually written as CamelCase [24] so that there are no spaces).
   - In dealing with large numbers of objects, it becomes impossible to devise a recognisable, unique name for each object. We then opt – almost automatically – for numbering.
   - There is a grey area between these two extremes. It is advantageous in the case of small, stable sets of objects (provinces, for example) to include the entire name in the URI, whereas in the case of larger sets, which involve more mutations, there are often long names that render the URI unmanageable. The use of abbreviations in the URI may be a solution in such cases. [25] and [26]
2. Avoid the use of strange symbols in a URI.

It is best to limit oneself to lowercase letters, digits and, if necessary, hyphens as punctuation.

Open issues

We have observed a number of issues, but we have yet to come to conclusions or to reach consensus. The question is whether this is possible or even necessary. There may always be a certain amount of controversy and Linked Data must then be able to function as well.

Issue1: URNs vs. URIs

The W3C shows the following as best-practice [27]:

1. Use URIs as names for things
2. Use http-URIs, so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
4. Include links to other URIs. So that they can discover more things.

This is, in itself, an extremely powerful paradigm: http-URIs make use of the existing infrastructure of the web and every browser is capable of requesting an http-URI. By means of content negotiation – also a standard http functionality – the server transcribes the request into the requested format on the basis of ID and supplies a document – the 'useful information' – to the client with reference to the http-returncode '303-redirect'. This redirect is required because the documents, in varying formats, each have their own URL, which differs from the ID that is being requested. And precisely that is found to be poorly understood and often incorrectly implemented in practice.

In addition, we can already observe a standardisation taking place on numerous domains in the identification of information that is available online. This is necessary – even if it does not concern Linked Data – because the information on the Internet can be copied and combined with other information in numerous ways. A major difference compared to the old, paper world in which copies were made sparingly and the documents to which one referred generally made up part of the physical file. That standardisation is seldom (for now?) in the form of an http-URI, but is usually a number, or in other words a URN (Unified Resource Name). The best known example of this is the ISBN-number for books. A less known, but more recent and more relevant example concerns the new standard ECLI-code of the EU, which is used to identify court judgments. And so, taking into account the starting point of reusing whatever is already available, we could consider generating http-URIs using the URNs that are already being made. The W3C best-practice will remain intact with just a small alteration:

1. Use URNs as names for things
2. Use REST-services as resolvers for those URNs, so that people can look up those names.
3. When someone looks up a URN, provide useful information, using the standards (RDF*, SPARQL)
4. Include links to other URNs. So that they can discover more things.

It is now clear that the URN does not produce the desired result and that the resolver returns a description of the object. This does not require a 303-redirect. This approach also renders it

possible to generate various resolvers for the same URN. Each resolver can decide which 'useful information' it provides regarding a concept. This is an advantage compared to the http-URI, which always leads to a single location.

Issue 2: Recognisable Internet domain

It is assumed in the British strategy that all of the http-URIs of Linked Data of the British government are ranged under a single principal domain: 'data.gov.uk'. Meanwhile, however, there are ideas to revise the organisation of the entire ‘gov.uk’ domain, making the future of data.gov.uk uncertain as well. An important piece of advice on the part of one of the devisers of the English strategy is therefore to lay down by law any domain that is a permanent component of persistent URIs, because otherwise the persistency cannot be guaranteed.

The notion of a principal domain is, however, quite appealing. It ensures the recognisability of the URIs, which evokes trust on the part of the buyers of the Linked Data. In the Netherlands, we could opt to use the domain data.gov.nl to that end. An additional advantage of this domain is that it is not yet in use and so it does not call up any other associations as yet.

On the other hand, a single principal domain also leads to problems. The domain has to be subdivided in order to keep this solution scalable. This can be achieved by defining sub-domains. The division into sectors was already rejected above. But 'No register, no identifier' provides the solution: each register could have its own sub-domain: `{register}.data.gov.nl`. This guarantees scalability, but this will of course require an administration of all of the registers: A register of registers.

Of course, in itself, it is quite useful to have such a register, but it does involve management costs and it is also a potential single point of failure.

The conclusion in the final workshop was that it is not realistic to make a solution entirely dependent upon a central provision that, moreover, is to be managed by a third party. That would make the keepers of the registers dependent upon this party when minting their URIs in accordance with the URI-strategy. And so a central register of registers should never be an indispensable part of the system. The register keepers must be in full control over the domain of their register completely independently.

And so each register must reside under an Internet domain of its own: `{register}.nl`. The register domain is used for the namespace and for the authentic resolver of the register itself.

It will then still be useful to set up a Register of Registers. Not as an essential component, without which the system cannot function, but rather as a convenient catalogue (/signpost/gazetteer?) for developers who seek registers, resolvers and register keepers.

It is conceivable (but not essential) that a resolver is set up by the Register of Registers (for example: `{register}.data.gov.nl`) that redirects to the resolver of the register (to for example `{register}.nl`). By that way `{register}.data.gov.nl` becomes an alternative resolver for the authentic resolver of the register. But it is not wise to see this as purl-server.

Issue 3: Degree to which the strategy must be formalised

The ultimate goal is a situation in which each object has only 1 URI. This URI is minted by the authentic registration of that object. This has been laid down by law for some of the data. The question is whether or not it is desirable and feasible to implement a formal basis for authenticity of this kind in the strategy, or, perhaps even better, if we can have faith in an organic process in which the most frequently used sources become de facto registers? We could formulate the formal criteria with which a register must comply as follows:

A register

- is a Standard or an Authentic Registration
- has a formal basis for authenticity:
    - the Standard is included on the 'comply or explain'-list with open standards [28] of the Forum Standaardisatie
    - the Authentic Registration is designated as such in Dutch or EU-regulations
- has a register keeper
- has its own Internet domain (namespace)
- has its own URI-pattern that complies with the national URI-strategy

This also applies to the URI-strategy itself: Can the URI-strategy be defined as a standard? And if so, will it suffice to include this standard on the 'apply or explain'-list of the Forum Standaardisatie or are more extensive measures desired, perhaps even legislation?

Taken from "http://www.pilod.nl/index.php?title=Boek/BrinkEtAl-URI&oldid=3098"