



Spatial Data on the Web using the current SDI

Report of the research results in the Geonovum testbed “Spatial Data on the Web” (topic 4)

16/03/2016

Authors

Clemens Portele, interactive instruments GmbH
Paul van Genuchten, GeoCat BV
Lieke Verhelst, Linked Data Factory
Andreas Zahnen, interactive instruments GmbH

This report is published under a CC/by license (<http://creativecommons.org/licenses/by/4.0/>)

Table of Contents

[Table of Contents](#)
[Abstract](#)
[The Goal](#)
[Our Motivation](#)
[Our Approach - an Overview](#)
[Schema.org](#)
 [Relevance of schema.org](#)
 [Mapping of features in datasets to schema.org](#)
 [BAG data](#)
 [Mapping of BAG dataset](#)
 [Mapping of BAG layer](#)
 [Mapping of BAG feature](#)
 [Farmland data](#)
 [Bekendmakingen data](#)
 [Mapping of metadata records to schema.org](#)
 [Mapping of other WFS resources to schema.org](#)
 [WFS landing page](#)
 [WFS feature type](#)
 [Subsets of a WFS feature type](#)
 [WFS feature](#)
[GeoDCAT-AP](#)
 [Mapping of metadata records to GeoDCAT-AP](#)
[URI strategy](#)
 [General considerations](#)
 [WFS proxy](#)
 [CSW proxy](#)
[Representations](#)
 [WFS proxy](#)
 [Main page](#)
 [HTML](#)
 [WFS landing page](#)
 [HTML](#)
 [XML](#)
 [WFS feature type](#)
 [HTML](#)
 [JSON-LD](#)
 [GML](#)
 [GeoJSON](#)
 [Subsets of a WFS feature type](#)
 [WFS feature](#)
[CSW proxy](#)
 [HTML](#)

[RDF/XML](#)

[JSON-LD](#)

[Linking](#)

[Semantic workspace](#)

[Link storage location](#)

[Establishing links](#)

[Link maintenance](#)

[Links from BAG to Bekendmakingen features](#)

[Links between the spatial data and metadata](#)

[Implementations](#)

[Idproxy](#)

[GeoNetwork](#)

[Validation and testing](#)

[Indexing by search engines](#)

[Indexing results](#)

[Ranking in search results](#)

[Proxy performance](#)

[Understandability by non-experts](#)

[Interoperability with the Linked Data Cloud](#)

[Test method](#)

[Observations related to schema.org and its use by search engines](#)

[Schema.org geo extension](#)

[Schema.org validation and Linked Data principles](#)

[Dereferenceability](#)

[Multiple Types](#)

[Linking principles](#)

[Experiences with existing WFS deployments](#)

[Experiences with existing metadata](#)

[Open challenges](#)

[Challenges related to the existing infrastructure](#)

[Support for establishing links dynamically](#)

[Metadata today is often incomplete and inconsistent](#)

[Performance & data compactness](#)

[Challenges related to the Web](#)

[Search engines are a black box](#)

[Content negotiation based on media types insufficient](#)

[Links](#)

[Deployments](#)

[Source code repositories](#)

[Discussions and other resources](#)

[Appendix A](#)

[XML Schema: Farmland data](#)

[XML Schema: Simple BAG data](#)

[Appendix B - Crosswalk: Spatial Data on the Web best practices in the WFS proxy](#)

[Appendix C: Example of an ISO 19139 document transformed to GeoDCAT-AP](#)

Abstract

This document is the report of interactive instruments, GeoCat and the Linked Data Factory on the findings from the work on research topic 4 in Geonovum's "Spatial data on the Web" testbed.

Finding, accessing and using data disseminated through spatial data infrastructures (SDI) based on OGC web services is difficult for non-expert users. Our research has investigated how to improve this while keeping the current spatial data infrastructures intact. I.e., we have been exploring ideas how to realize synergies between the current spatial data infrastructures and the developments on the Web of data.

The approach has been to design and implement an intermediate layer using proxies that will make data and metadata from the OGC web services available using the following principles that we consider important from a Web perspective:

- All resources are identified using persistent HTTP URIs
- All interaction is using the HTTP protocol and consistent with its design
- All resources are discoverable via search engines
- Resources can be accessed and understood by citizens and developers
- Resources are either explicitly linked using HTTP URIs or data is structured so that links can be established dynamically

To a large extent we were successful. This report in particular documents

- the resources that are made available by the proxies,
- the mapping of the data and metadata resources to the schema.org vocabulary (plus the GeoDCAT-AP vocabulary for metadata),
- our strategy for assigning URIs to the resources,
- the representations (formats), in which each resource is available,
- our experiences with establishing links across datasets and between data and metadata, and
- our experiences with search engines crawling and indexing the resources.

There are a number of open challenges and takeaways that we have learned during our work on the research questions.

- Existing infrastructure
 - Data integration across datasets: Existing datasets are sometimes structured in ways that make it difficult to dynamically establish links between resources in the datasets. Users of data on the Web would benefit from properly structured data using common vocabularies as this decreases the cost of data integration (in this testbed, this is linking) significantly.
 - Metadata today is often incomplete or inconsistent: Metadata provided in the WFS capabilities and the dataset metadata in the catalogs is either incomplete or both have inconsistent information.

- Performance & data compactness: Response times and data size may have an impact on the indexing and ranking by search engines as well as on the usability in general. More work is needed to understand this aspect better and to be able to assess, if existing SDI components meet expectations.
- The Web of data
 - Search engines are largely a black box when we look at spatial data: It is not clear to us, if and how structured spatial data is used by search engines and how users can use it to improve finding data they are interested in.
 - schema.org is a divergent kind of vocabulary - from the Linked Data perspective: Its use is in particular intended for RDFa/JSON-LD markup of HTML pages to facilitate the search engines and mobile digital personal assistants. The capabilities for RDFa/JSON-LD to mark up geospatial data are not optimal. Therefor a geo extension to schema.org is proposed.
 - Different vocabularies needed for different use cases / communities: We distinguish two general use cases for making spatial data available as Linked Data. The first is the same as the testbed: make spatial data part of the indexed web by providing a common structured format. The other use case is: make spatial data available and spatially queryable in the Linked Data cloud. For the first use case schema.org is necessary. For the second use case we recommend to use other vocabularies such as GeoSPARQL. To support both use cases, we advise to use both schema.org and GeoSPARQL.
 - Content negotiation based on media types insufficient: The HTTP content negotiation mechanism has its limits as it supports only selection based on the “data format”, but is unable to support selection of different vocabularies. The approach that we have taken, and which mostly solves the issue for us, is to simply provide separate URIs for the different vocabularies / user communities.

The Goal

Finding, accessing and using data disseminated through OGC services is difficult for non-expert users. This has several reasons, including:

- In spatial data infrastructures, catalog services are intended to be used for discovering spatial assets, not the general purpose search engines of the Web. OGC web services do not address indexing of their content by those search engines.
- By design, the catalog services only provide access to metadata - and in general metadata that is focussed on the needs of expert users - not the data itself.
- Users cannot just “follow links” to access data, it is typically necessary to construct some kind of query to access data.
- In addition, it is often difficult for non-expert users to understand and use the data. Part of this are domain-specific complexities that are difficult for non-experts (e.g., handling of coordinates in different coordinate reference systems), but hard to avoid entirely. But the datasets often address requirements of expert communities with diverse needs, resulting in comprehensive, but complex specifications that cover many edge cases, too. And the data is typically available in formats that are not easy to process for non-expert users.

The research documented in this report has investigated how to improve this while keeping the current spatial data infrastructure (SDI) intact. I.e., how can synergy be realized between the current spatial data infrastructures and the developments on the Web?

Our Motivation

This research topic combines two aspects that are both important to us. First of all, it allows to think ahead and investigate how to publish feature-based spatial data sets available on the Web. At the same time, it does not simply start from a clean sheet, but takes the existing infrastructure for spatial datasets and the related workflows for data management and dissemination into account.

We believe that it is important to avoid being constrained by the technical details of WFS, GML, CSW and ISO 19139 when exploring how the data should be made available on the Web, i.e. when focusing on the first aspect. In the end, it is essential that the data is useful for those that want to use it, e.g. for implementing web or mobile applications for the new environmental act.

At the same time, there is a lot of utility in leveraging the current infrastructure. This approach evaluates the extent in which the current spatial data infrastructures can support the provision of spatial data on the Web, and what are its conceptual limitations. If the approach proves to be feasible, then it provides a migration path with a potentially low effort for a data provider - or even a central operator of a spatial data infrastructure - to make his spatial dataset(s) also available “on the Web” as all the software that is used is made available under an open source license.

Since the time and resources available for this research are limited, it is also helpful that the scope of the research topic does not cover everything related to spatial data on the Web, but is focused on a few aspects that are particularly important:

- crawlability and linkability, i.e. making each resource available via a persistent URI and ensure that all resources can be reached via links from a “landing page” for a data set
- classification of the resources using vocabularies supported by the main search engines on the Web
- discovery of spatial data by the main search engine
- representations of data for consumption by humans (HTML), web-developers (JSON) and search engine crawlers (HTML with annotations)
- establishing and maintaining links between data

Our Approach - an Overview

There are ongoing discussions what good practices are for publishing spatial data on the Web.

For example, a joint [Spatial Data on the Web working group](#) of W3C and OGC is currently in the process of documenting good practices. In an [appendix of this report](#), we summarize an assessment in how far our work in this testbed follows the current draft of best practices identified by the working group.

A relevant ongoing discussion in that working group is in how far a [Linked Data approach with its 5-star ranking and preference for RDF](#) should be the basis for the best practices or whether the approach should be more open with respect to technologies (see, for example, [here](#)) as long as they are consistent with the [Web architecture](#).

In our work we are taking more the latter approach, although we keep in mind that given the strong support for Linked Open Data in the Netherlands there is value in following also the Linked Open Data principles. I.e. our design aims at being consistent with the Web as it is today as well as with the Linked Open Data principles.

As organisations already have an infrastructure set up to publish OGC web services, this research investigates, if we can design an intermediate layer that will make data from those OGC web services available in a ‘webby’ way. With ‘webby’ we mean in the context of our research:

- All resources are identified using persistent HTTP¹ URIs
- All interaction is using the HTTP protocol and consistent with its design (support for the HTTP verbs, content negotiation, etc.)
- All resources are discoverable via search engines
- Resources can be accessed and understood by citizens and developers
- APIs to access data should be self-describing

¹ When we say “HTTP”, this includes the use of HTTPS throughout this document.

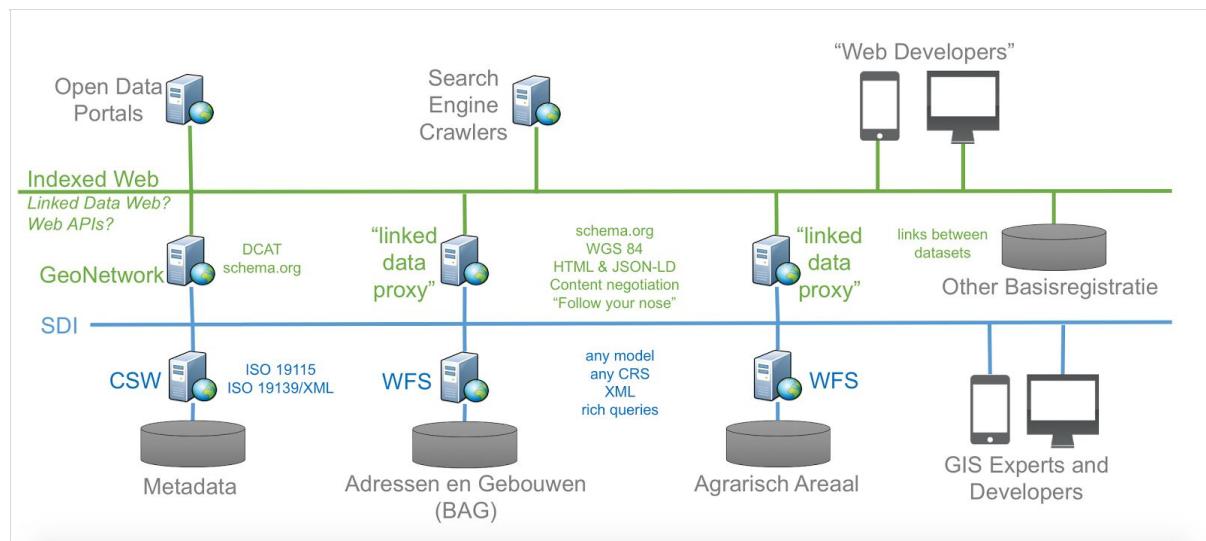
- Resources are either explicitly linked using HTTP URIs or data is structured so that links can be established dynamically

The research focusses on tools daily used by citizens, such as Google Search or Cortana. It is conducted in the scope of the upcoming wide environmental law, ‘wet van de leefgomgeving’, as the basis for use cases. Questions like ‘can i build a fish pond in my garden’ , or, more general: ‘which recent announcements are available for this area’ are envisioned.

Proxy approach

To establish that intermediate layer, proxies are introduced on top of the WFS (data service) and CSW (metadata service) so the contained resources are made available for other communities and supporting the practices they follow.

The “blue” layer in the figure below represents the spatial data infrastructure that addresses needs of the community of geospatial experts.



The “green” proxy layer is intended to support the following communities and their practices (not pretending to be complete):

- Search engines
Search engines generally use the schema.org ontology, use HTTP URLs as identifiers and limit their encodings to RDFa/Microdata and JSON-LD.
- Linked Open Data (LOD) community
Use a minimal set of common ontologies. Largely influenced by the DBpedia project.
- e-government community
Require the use of authoritative ontologies (DCAT, “basisregistraties”, INSPIRE)
- Web API developer community
Don’t use ontologies, but have additional best practices, such as GeoJSON, Swagger, CKAN, etc.

An essential part of the proxies is to transform between the resources, their representations as published by the proxies and their hypermedia controls on one hand and the OGC web service requests and response messages and formats. Transformations are usually combined efforts of

- selection using queries
- schema transformation and
- format-serialisation.

Content negotiation plays an important role in this area.

This report discusses the following aspects of the implementation:

- the mapping of the data and metadata resources to the identified vocabularies, [schema.org](#) and [GeoDCAT-AP](#)
- the strategy for [assigning URLs](#) to the resources
- the [representations](#) (formats), in which each resource is available
- [linking](#) across datasets and between data and metadata

Based on this discussion, we document our [findings](#) on the research questions.

Schema.org

Relevance of schema.org

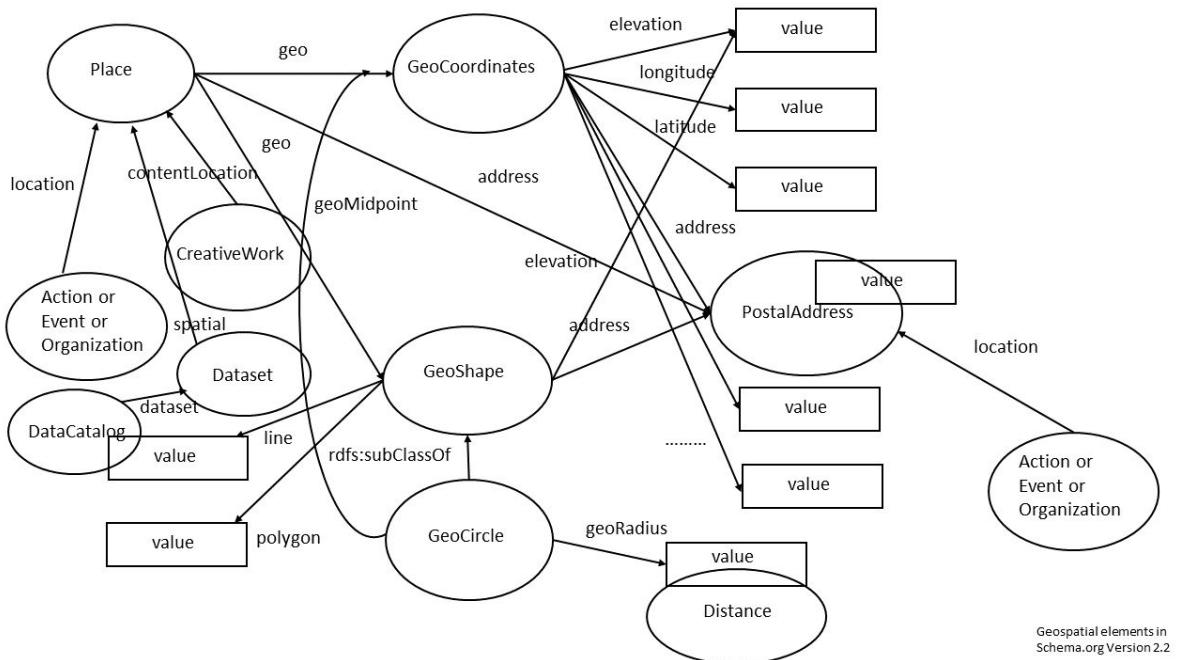
This research focusses on common tools like Cortana and search engines. For that reason transformation to the schema.org ontology is relevant since it is *the* vocabulary understood, and as a result mandated, by the search engines. The scope of Schema.org is to provide a extensive data model of objects commonly advertised on the web. Not all OGC feature types and properties can be transformed without loss of information to schema.org. For that reason [schema.org supports extensions](#). However, it is unclear, if using extensions may have a negative impact when data is indexed or ranked in searches². In our research we make use of extensions to facilitate applications that do support the extensions.

Mapping of features in datasets to schema.org

Schema.org dictates a certain hierarchy and order for the way data sources must be modeled. Details are given in the documentation of the schemas³. For an easy overview the most important geo relations are depicted by us in a drawing of classes (the ovals) and properties (the arrows):

² The testing tools for structured data that the search engine providers offer sometimes report errors when extensions are included, but it is not clear, if that would impact indexing. We do not have any evidence from our research that this is case.

³ <http://schema.org/docs/schemas.html>



This is a partial representation of schema.org that depicts the most important classes (circles) and properties (arrows) in which geo information is presented.

This image makes clear that it is not possible to model every class with every property, but that sometimes intermediate classes must be used to attach properties to a class. For example: a Place does not have a latitude and longitude, but a GeoCoordinates does. So in order to add latitude and longitude to a Place, one needs to model the GeoCoordinates class into the mapping to make the mapping compliant with schema.org.

BAG data

We used the simple BAG WFS⁴ to demonstrate the working of schema.org for point datasets. The source XML Schema is listed in Appendix A. This BAG data model contains a postal address and coordinates, so if we use Place as the main class, we can use GeoCoordinates and PostalAddress as secondary classes to model every property necessary.

Mapping of BAG dataset

The BAG WFS proxy landing page can be represented by a schema:Dataset that refers to another Dataset (the layer) via the schema:hasPart property.

Mapping of BAG layer

The layer dataset contains a collection of Places. The property that links the Dataset to the Place is schema:spatial.

⁴

<http://geodata.nationaalgeoregister.nl/inspireadressen/wfs?service=WFS&request=DescribeFeatureType&version=2.0.0>

Mapping of BAG feature

source	type	destination	type	remark
inspireadressen:inspireadressen	complex type	instance of schema:Place	schema:Place	has URI
		instance of schema:PostalAddress	schema:PostalAddress	Blank Node
all address info		schema:address	rdf:Property	to schema:PostalAddress
straatnaam	element of type string	schema:streetAddress		concatenate these values into schema:streetAddress
huisnummer	element of type string	schema:streetAddress		concatenate these values into schema:streetAddress
huisletter	element of type string	schema:streetAddress		concatenate these values into schema:streetAddress
toevoeging	element of type string	schema:streetAddress		concatenate these values into schema:streetAddress
woonplaats	element of type string	schema:addressLocality	string	
postcode	element of type string	schema:postalCode	string	
geom	element of type pointPropertyType	schema:geo	rdf:Property	to schema:GeoCoordinates
		instance of type schema:GeoCoordinates	schema:GeoCoordinates	Blank Node
		schema:latitude	number	
		schema:longitude	number	

A JSON-LD example:

```
{
  "@context" : "http://schema.org",
  "@type" : "Place",
  "@id" : "http://www.ldproxy.net/bag/inspireadressen/inspireadressen.23079",
  "url" : "http://www.ldproxy.net/bag/inspireadressen/inspireadressen.23079",
```

```

"geo" : {
    "@type" : "GeoCoordinates",
    "longitude" : "5.456203403470931",
    "latitude" : "51.34206514984284"
},
"address" : {
    "@type" : "PostalAddress",
    "streetAddress" : "de Molensteen 6",
    "addressLocality" : "Valkenswaard",
    "postalCode" : "5554VC"
}
}

```

Farmland data

The farmland data XML Schema is listed in Appendix A. It maps to schema.org by using Place as the main class, and GeoShape as the class that contains the polygon coordinates.

source	type	destination	type	remark
aan:aan	complexType	schema:Place		has URI
objectid	element of type integer	ignore		this attribute is a system internal id
aanid	element of type integer	qname of the resource	the QN of the identifier (URI) of the instance of schema:Place, must have the URL of the WFS proxy	see aan:aan
versiebron	element of type string	must be added in an extension to schema.org		there is no property for Place or above Place that can represent “source”
type	element of type string	schema:description	string	the information in the data contains info about the type land
geom	element of type gml:MultiSurfacePropertyType	schema:geo	rdf:Property	to schema:GeoShape
gml:MultiSurfacePropertyType	complexType	instance of schema:GeoShape	Blank Node	
		schema:polygon	number list	the information in the data contains the polygon coordinates

A JSON-LD example, which also includes additional properties following [the approach described in the JSON-LD standard using null values in the context](#) for now:

```
{
  "@context" : {
    "@vocab" : "http://schema.org/",
    "versiebron" : null
  },
  "@type" : "Place",
  "@id" : "http://www.ldproxy.net/aan/aan/aan.1",
  "url" : "http://www.ldproxy.net/aan/aan/aan.1",
  "geo" : {
    "@type" : "GeoShape",
    "polygon" : "4.73669815895657,53.04210214780725
4.737386778482686,53.04225965287612 4.738189184603757,53.041560320183144
4.738334449604591,53.04144258121632 4.7383578153531065,53.04142028437658
4.737935049254453,53.04125675245224 4.737676098195687,53.04115710925473
4.737337237145229,53.041069844303244 4.737220698890215,53.04104358436285
4.736999734927018,53.0410103498384 4.736914748307701,53.04100668450981
4.736679349996972,53.04146456389503 4.736380359462239,53.04202780911646
4.73669815895657,53.04210214780725"
  },
  "name" : "1356489",
  "versiebron" : "luchtfoto 2013",
  "description" : "BTR-landbouw"
}
}
```

Bekendmakingen data

The testbed facilitates to research how search engines respond to a mapping between an existing already published Linked Data set that is linked by the WFS proxy. We have used a website that publishes government notification. This dataset supports the following use case: a person that is located at a particular “Place” wants to know “what official notifications have been published by governmental organisations at this location?”.

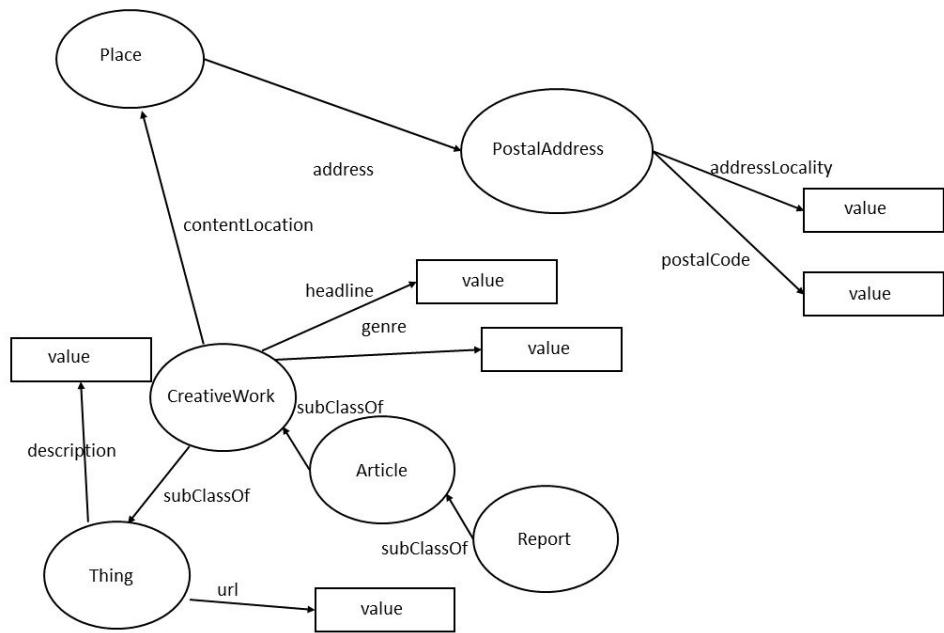
The data that is behind the website <https://www.officielebekendmakingen.nl/> is served by more than one type of service, amongst them a WFS service⁵ and a SRU service⁶. Output from the SRU service filtered on ‘Valkenswaard’ was used to set up two versions of a Linked Data version of “Bekendmakingen”, a spatial version with coordinates and no postal code, and an administrative version with postal codes and no coordinates. The data was not cleansed to illustrate issues with data quality.

We have used a selection of properties to represent the bekendmakingen in both services. The schema.org class used is Report. The non spatial version is represented as follows:

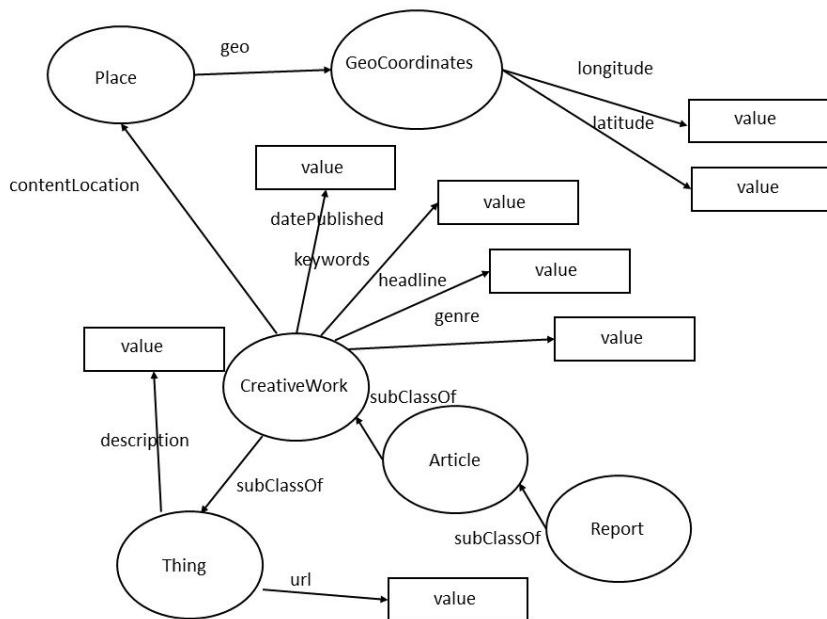
⁵ <https://geozet.koop.overheid.nl/wfs?request=getcapabilities>

⁶

<http://zoekdienst.overheid.nl/sru/Search?version=1.2&operation=searchRetrieve&x-connection=BM&startRecord=1&maximumRecords=1000&query=keyword=valkenswaard>



The spatial version like this:



It is important to mention that the address information in the original Bekendmakingen dataset that was retrieved is quite incomplete. Many address fields are empty, and the address information is a scatter of text snippets in the other fields. At best the postal code field is used with a 4 digit postal code. Needless to say this does not help with automatically trying to integrate this data with other data sets.

Both datasets are published in a SPARQL endpoint and published dereferenceable via their URLs. For details see: <https://github.com/geo4web-testbed/topic4-task2>.

Mapping of metadata records to schema.org

Catalog records are typically structured as ISO 19139 XML documents. As part of this research we've looked at optimal scenario's to transform this metadata to schema.org and make it crawlable by search engines.

Schema.org does not have a concept of a catalog record, so we map each ISO 19139 catalog record describing a resource directly to a schema.org class.

As part of ISO 19115 the hierarchyLevel element determines the type of object described in the metadata. The available types are listed in the ScopeCode⁷ codelist of ISO 19115. For each of these concepts a different schema.org class can be used.

dataset	http://schema.org/Dataset
series	http://schema.org/DataCatalog
service	http://schema.org/DataCatalog
application	http://schema.org/SoftwareApplication
collectionHardware	?
nonGeographicDataset	http://schema.org/Dataset
dimensionGroup	?
featureType	http://schema.org/Dataset
model	http://schema.org/APIReference
tile	http://schema.org/Dataset
fieldSession	?
collectionSession	?
other	http://schema.org/Thing

Schema.org does not have a concept for series and service metadata. A viable approach is to model the series and service as a list of Datasets. GeoDCAT-AP suggests to model a service as a dcat:catalogue. We suggest to model the series and service in schema.org in a similar way. For quite some of the hierarchylevels no suitable classes are available in schema.org, creating a schema.org extension or using one of the available extensions, such as the web-observatory extension may make sense. In the scope of this research only metadata describing datasets, services and series is relevant. Other hierarchy levels are not commonly used in the current national spatial data infrastructure.

Dataset descriptions structured as ISO 19139 map quite well on Schema.org/Dataset. Most obvious missing properties are typical spatial properties such as spatial resolution and projection info. Spatial extent of a dataset is poorly modelled in schema.org. Most of the ISO 19139 metadata uses a bounding box element to indicate the spatial extent of a dataset. Schema.org expects (in line with DCAT) an identifier of a geographic location. Off course a

⁷ <http://www.isotc211.org/2005/resources/Codelist/gmxCodelists.xml>

transformation may create a location object specifically for the dataset having a geometry property similar to the dataset bounds, but from a linking perspective less powerful. Instead we recommend dataproviders creating ISO 19139 documents, to include both a geographic identifier as a geographic extent element as the location of a dataset.

For each ISO 19139 having a link to a WFS, the original WFS endpoint is provided as a schema.org/DataDownload.contentUrl. The dutch metadata profile requires the featuretypename be made available in the name element of the gmd:online section. This property can be used to create a getfeature operation as part of the contentUrl that will return the full dataset in any of the formats supported by the WFS.

For those catalog records describing a dataset that is available via the Idproxy software an additional schema.org/DataDownload.contentUrl is added with a link to Idproxy. It would be interesting to experiment with automatic configuration of the Idproxy software for any WFS endpoint available in the catalog.

To support such a workflow from the catalogue to Idproxy automatically configured on any WFS, the catalogue should notify Idproxy of the wfs-url to be used. Idproxy supports this by a query using the WFS URL. For example,

<http://www.Idproxy.net/?wfsUrl=http://geodata.nationaalgeoregister.nl/inspireadressen/wfs>

This URL will return a HTTP 307 redirect to the proxy service <http://www.Idproxy.net/bag/>.

The WFS URLs are normalized by removing known parameters like SERVICE, VERSION, etc.

If the WFS proxy does not have a configuration for that WFS yet, it has to be set up first. An idea for the future would be to add an API to Idproxy that can be used by the catalogue (or others) to retrieve and set up such a proxy endpoint. That administration endpoint would need to be protected, so it can not be misused.

Testing in Structured Data Testing Tool

While testing the schema.org document in google structured dataset testing tool, we identified that google enforces some additional, non documented, properties as part of the contactpoint information. For example a url and contact-type are required and a phone number should have the international syntax. It is to be expected that newer versions of (implementations of) the ontology will introduce more restrictions, as soon as the search engines will start to interact with the structured data about datasets.

Dataset (1)	All good
Dataset:	http://opendatacat.net/geonetwork-geo4web/doc/dataset/gn-metadata-view-5416
name:	Werelderfgoed
datePublished:	2014-07-08
description:	Begrenzingen van het Nederlandse cultureel erfgoed dat vanwege haar uitzonderlijke universele waarde door het Werelderfgoed comite van UNESCO is geplaatst op de Werelderfgoedlijst. De begrenzingen zijn gebaseerd op de kaarten die destijds opgenomen waren in het nominatiedossier dat aan Unesco is aangeboden.
thumbnailUrl:	http://services.rce.geovoortziening.nl/www/PreviewWerelderfgoedKlein200x150.png
thumbnailUrl:	http://services.rce.geovoortziening.nl/www/PreviewWerelderfgoedGroot800x600.png
keywords:	werelderfgoedUNESCO

Mapping of other WFS resources to schema.org

WFS landing page

An OGC Web Feature Service provides access to a dataset. As described above, the proxy landing page for a WFS is therefore represented by a [schema:Dataset](#).

Information from the capabilities document of the WFS is mapped to schema.org properties, where possible.

Service identification section:

- ows:ServicIdentification/ows:Title → [schema:name](#)
- ows:ServicIdentification/ows:Abstract → [schema:description](#)
- ows:ServicIdentification/ows:Keywords/ows:Keyword → [schema:keywords](#)
- ows:ServicIdentification/ows:AccessConstraints → [schema:licence](#)

Service Provider section:

- ows:ServiceProvider → [schema:provider](#) with a [schema:Organization](#) resource as its value
 - ows:ProviderName → [schema:name](#)
 - ows:ServiceContact → [schema:contactPoint](#) with a [schema>ContactPoint](#) resource as its value.
 - ows:ServiceContact/ows:ContactInfo/ows:Address → [schema:address](#) with a [schema:PostalAddress](#) resource as its value.

- Google requires that schema:url is provided in several places of a [schema:Organization](#). If the information is not included in the capabilities, we have used a URL in the domain example.com, but that is of course an issue that would need to be addressed in a proper solution, for example, by working with the WFS provider to add the information or by configuring the missing information in the proxy service.

Feature type section:

- ows:FeatureTypeList/ows:FeatureType → [schema:hasPart](#) with a [schema:Dataset](#) resource as its value
- The combined bounding box of all feature types is also mapped to [schema:spatial](#) with a blank node [schema:Place](#), which has a [schema:geo](#) property with a blank node [schema:GeoShape](#) with a [schema:box](#) property.

The URL of the GetCapabilities request is mapped to [schema:sameAs](#).

In addition, we also want to include the link to the dataset metadata, which is stored in an OGC Catalog Service (CSW). The value is represented as [schema:isPartOf](#) with a [schema:DataCatalog](#) resource. The general approach is discussed in more detail [here](#).

For the dataset metadata, Idproxy extracts the metadata URL information from the WFS capabilities document, if available. In an INSPIRE service the link to the service metadata is contained in the ows:ExtendedCapabilities.

The structured data according to schema.org from the BAG WFS is shown below (using Google's Structure Data Testing Tool):

▼ Dataset (1)	Keine Fehler
Dataset	
name:	INSPIRE Adressen WFS
description:	INSPIRE Adressen themalaag, gevuld met relevante objecten uit de Basisregistratie Adressen en Gebouwen (BAG), beheerd door het Kadaster.
keywords:	Adressen
license:	otherRestrictions; Geen beperkingen
url:	http://www.ldproxy.net/bag/
sameAs:	http://geodata.nationaalgeoregister.nl/inspireadressen/wfs?REQUEST=GetCapabilities&SERVICE=WFS&
hasPart [Dataset]:	
url:	http://www.ldproxy.net/bag/inspireadressen/
spatial [Place]:	
geo [GeoShape]:	
box:	3.053,47.975 7.24,53.504
provider [Organization]:	
name:	PDOK
url:	http://www.example.com/todo
contactPoint [ContactPoint]:	
contactType:	technical support
name:	KlantContactCenter PDOK
url:	http://www.example.com/todo
address [PostalAddress]:	
addressLocality:	Apeldoorn
addressCountry [Country]:	
name:	Nederland
isPartOf [DataCatalog]:	
url:	http://opendatacat.net/geonetwork-geo4web/doc/dataset/3a97fbe4-2b0d-4e9c-9644-276883400dd7

WFS feature type

A feature type is also represented by a [schema:Dataset](#), which is part of the WFS dataset described above.

Information from the capabilities document of the WFS is mapped to schema.org properties, where possible.

Feature type section:

- ows:FeatureTypeList/ows:FeatureType/ows:Title → [schema:name](#)
- ows:FeatureTypeList/ows:FeatureType/ows:Abstract → [schema:description](#)
- ows:FeatureTypeList/ows:FeatureType/ows:Keywords/ows:Keyword → [schema:keywords](#)

- ows:FeatureTypeList/ows:FeatureType/ows:WGS84BoundingBox → [schema:spatial](#) with a blank node [schema:Place](#), which has a [schema:geo](#) property with a blank node [schema:GeoShape](#) with a [schema:box](#) property.
- ows:FeatureTypeList/ows:FeatureType/ows:MetadataURL → [schema:isPartOf](#) with a [schema:DataCatalog](#) resource. The URL of the DataCatalog is constructed as described [above](#).

The WFS dataset is included using [schema:isPartOf](#).

In case any named subsets have been configured, like the municipalities for the BAG dataset, the list of subsets is referenced using [schema:hasPart](#) with a [schema:Dataset](#) resource as its value. Note that the use of hasPart does not seem right as this is an address dataset and the list of municipalities is just derived information - distinct values of all municipality attributes; i.e., this aspect is an [open issue](#).

In addition, (up to) 25 features of the dataset are shown, basically a page. As schema.org does not include any pagination support, links to other pages are not included.

The structured data according to schema.org from the address feature type in the BAG WFS is shown below (using Google's Structure Data Testing Tool):

Dataset (1)		Keine Fehler ✓
Dataset		
name:	Adressen	
description:	INSPIRE Adressen afkomstig uit de basisregistratie Adressen, beschikbaar voor heel Nederland	
url:	http://www.ldproxy.net/bag/inspireadressen/	
keywords:	Adressen	
isPartOf [Dataset]:		
url:	http://www.ldproxy.net/bag/	
spatial [Place]:		
geo [GeoShape]:		
box:	3.053,47.975 7.24,53.504	
isPartOf [DataCatalog]:		
url:	http://opendatacat.net/geonetwork-geo4web/doc/dataset/76091be7-358a-4a44-8182-b4139c96c6a4	
hasPart [Dataset]:		
url:	http://www.ldproxy.net/bag/inspireadressen/?fields=addressLocality&distinctValues=true	
Place (25)		Keine Fehler ✓
Place 1		
url:	http://www.ldproxy.net/bag/inspireadressen/inspireadressen.1/	
name:	Appingedam, Fivelkade W 4	
address [PostalAddress]:		
streetAddress:	Fivelkade W 4	
addressLocality:	Appingedam	
postalCode:	9901GE	
Place 2		
url:	http://www.ldproxy.net/bag/inspireadressen/inspireadressen.2/	
name:	Appingedam, Fivelkade W 5	
address [PostalAddress]:		

Subsets of a WFS feature type

The schema.org representation of a subset of a feature type is basically the same as for the feature type itself.

WFS feature

The mapping to schema.org will differ by feature type. The main elements of the mappings for the address and the farmland features are discussed [above](#).

In addition:

- [schema:hasMap](#) references the HTML representation which includes a map using leaflet.

- [schema:sameAs](#) references the feature in the WFS.

The structured data according to schema.org from an address feature in the BAG WFS is shown below (using Google's Structure Data Testing Tool):

Place (1)		Keine Fehler ✓
Place		
name:	Valkenswaard, Boomvalk 24	
url:	http://www.ldproxy.net/bag/inspireadressen/inspireadressen.23097/	
hasMap:	http://www.ldproxy.net/bag/inspireadressen/inspireadressen.23097/?f=html	
sameAs:	http://geodata.nationaalgeoregister.nl/inspireadressen/wfs?REQUEST=GetFeature&OUTPUTFORMAT=application%2Fgml%2Bxml%3B+version%3D3.2&VERSION=2.0.0&NAMESPACES=xmlns%28inspireadressen%2Chttp%3A%2F%2Finspireadressen.geonovum.nl%29&TYPENAMES=inspireadressen%3Ainspireadressen&SERVICE=WFS&COUNT=1&RESOURCEID=inspireadressen.23097	
address [PostalAddress]:		
streetAddress:	Boomvalk 24	
addressLocality:	Valkenswaard	
postalCode:	5554LZ	
geo [GeoCoordinates]:		
latitude:	51.34249325134195	
longitude:	5.456679823278034	

GeoDCAT-AP

A proxy approach as described in this report can also be applied to facilitate spatial data discovery from the linked data community. The proxies should be configured in such a way that they can also return RDF encodings of objects structured using common ontologies from the linked data community. A widely used ontology to describe datasets in the linked data domain is [DCAT](#)⁸.

[GeoDCAT-AP](#) is an extension on DCAT and [PROV](#)⁹ developed in the scope of joinup¹⁰, to provide an ontology to express the full richness of INSPIRE metadata encoded according to ISO 19139. Initially the goal of the GeoDCAT-AP initiative was to facilitate discovery of INSPIRE data in the open data domain. But recent discussions tend to promote GeoDCAT-AP as an alternative metadata format within an INSPIRE discovery service¹¹.

⁸ <https://www.w3.org/TR/vocab-dcat/>

⁹ <https://www.w3.org/TR/prov-o/>

¹⁰ https://joinup.ec.europa.eu/asset/dcat_application_profile/asset_release/geodcat-ap-v10

¹¹

http://inspire.ec.europa.eu/documents/Network_Services/TechnicalGuidance_DiscoveryServices_v3.0.pdf

Mapping of metadata records to GeoDCAT-AP

Joinup released, together with the GeoDCAT-AP specification, an XSLT which can transform ISO 19139 metadata to GeoDCAT-AP. We've used this XSLT to improve existing RDF export capabilities in GeoNetwork. Until recently RDF/XML could only be exported from GeoNetwork using an RSS type of search. In recent versions RDF/XML can also be exported using CSW and as a full catalog dump. Only RDF/XML is supported, no transformations to turtle or json-ld are currently available.

The Joinup XSLT has been improved on a number of aspects.

- The Joinup XSLT tends to create frequent ‘blank nodes’, which is not forbidden, but a bad practice in linked data. Instead GeoNetwork ‘mints’ uri’s for new objects that are created as part of the mapping (eg organisations, locations, etc).
- The Joinup XSLT is quite ambiguous on the type of data-links available in the metadata. The reason is that INSPIRE metadata does not require to indicate the type of resource behind a link and XSLT has no capabilities to probe a link, to derive the type of resource. The dutch metadata profile however requires data providers to indicate the type of resource behind a link, which facilitates the mapping for the type of data link.
- The dutch profile requires explicit conventions for stating types of open data licences. This facilitates the link to DCAT.

An example of a transformed ISO 19139 document is available in [Appendix C](#).

Besides DCAT also the [VOID¹² ontology](#) is relevant in the scope of our research. DCAT is widely used to describe traditional datasets in a structure other than RDF, VOID is used to describe datasets that are structured as RDF. As part of this research we suggest ways to convert non RDF data structures to RDF. To make those structures discoverable on the semantic web, VOID is a relevant ontology.

To facilitate Semantic Web Bots a SPARQL endpoint can be set up based on a (nightly) full RDF dump of the catalogue. Alternatively semantic web users (bots and people) can follow links to metadata URI’s from external sources (using a content negotiation accept header “application/rdf+xml”).

URI strategy

General considerations

The Platform Linked Data Nederland has formulated a national strategy¹³ for the minting of URIs. The idea behind this is that by formulating a number of recommendations for minting

¹² <https://www.w3.org/TR/void/>

¹³ http://www.pilod.nl/w/images/a/aa/D1-2013-09-19_Towards_a_NL_URI_Strategy.pdf

URIs organisations have a quick start procedure. The URI strategy describes a pattern for URIs: `http://{domain}/{type}/{concept}/{reference}`

Since the publication of this strategy there have been a number of review sessions in which lessons learned have been discussed. One lesson was that “what could work for one use case might not work for another”, so the “URI strategy” is now more of a guideline than a strategy.

As the URI strategy may well have an impact on search engine crawling and ranking, and since it is important for APIs and developers, the work on research topic 3 has compared different URI strategies. The findings are documented in their report. In our work with the proxies, we have implemented a straightforward URI schema that is documented in this section.

Depending on the use cases, some resources which are “child” properties of higher-level resources, may not be referenced directly. Our general approach was to not assign a persistent URI to those resources, i.e. use blank nodes, as they do not need to be linked.

A typical example is geometry, which we consider as “owned” by the feature ([schema:Place](#) in schema.org terminology) as sharing geometry is out-of-scope for our experiments. In other contexts, this might be handled differently.

WFS proxy

The WFS proxy provides URIs for the following resources.

Resource: [WFS landing page](#)

URI template: `{baseuri}/{wfs}[/?f={format}]`

Example: <http://www.ldproxy.net/aan>

Remarks:

- `{baseuri}` is the base URI of the ldproxy deployment
- `{wfs}` is configured in the ldproxy when the proxy service is set up
- The optional query parameter “format” can be used to select a specific media type, if HTTP content negotiation cannot be used. Values are:
 - “html” for HTML (media type: “text/html”), this is the default value
 - “xml” for WFS Capabilities (media type: “application/xml”)

Resource: [WFS feature type](#)

URI template: `{baseuri}/{wfs}/{featuretype}[/?page={n}&f={format}]`

Example: <http://www.ldproxy.net/aan/aan>

Remarks:

- `{featuretype}` is taken from the unqualified feature type name in the WFS capabilities
- The optional query parameter “page” can be used to select a specific page. The default page is the first page ($n=1$).
- The optional query parameter “format” can be used to select a specific media type, if HTTP content negotiation cannot be used. Values are:

- “html” for HTML (media type: “text/html”), this is the default value
- “jsonld” for JSON-LD (media type: “application/ld+json”)
- “json” for GeoJSON (media type: “application/vnd.geo+json”)
- “xml” for GML (media type: “application/gml+xml”)

Resource: [WFS feature](#)

URI template: {baseuri}/{wfs}/{featuretype}/{id}[/?f={format}]

Example: <http://www.ldproxy.net/aan/aan/aan.81>

Remarks:

- {id} is the gml:id of the feature
- The optional query parameter “format” can be used to select a specific media type, if HTTP content negotiation cannot be used. Values are the same as for the feature type.

Resource: [Named subsets of a WFS feature type](#)

URI template: {baseuri}/{wfs}/{featuretype}/?fields={attributes}&distinctValues=true

Example:

<http://www.ldproxy.net/bag/inspireadressen/?fields=addressLocality&distinctValues=true>

Remarks:

- {attributes} is an attribute of the feature type; the attribute must have a simple value
- Idproxy must be configured to support the subsetting for the attribute

Resource: [WFS feature type subset](#)

URI template: {baseuri}/{wfs}/{featuretype}/?{attributes}={value}[&page={n}&f={format}]

Example: <http://www.ldproxy.net/bag/inspireadressen/?addressLocality=Valkenswaard>
<http://www.ldproxy.net/bag/inspireadressen/?postalCode=5551AB>

Remarks:

- {value} is the attribute value to select the relevant features
- Idproxy must be configured to support the subsetting for {attribute}
- The optional query parameter “page” can be used to select a specific page. The default page is the first page (n=1).
- The optional query parameter “format” can be used to select a specific media type, if HTTP content negotiation cannot be used. Values are the same as for the feature type.

CSW proxy

If a catalog needs to be accessed, which does not offer similar transformations and access methods as implemented in GeoNetwork, an approach is envisioned in which a component within GeoNetwork acts as a CSW proxy to that service. The component propagates a request for a catalog record to the CSW service and returns the CSW results in the requested RDFa or RDF/xml format. This situation occurs, if a WFS capabilities contains a link to a CSW service which is in a domain that is not nationaalgeoregister.nl. Two approaches are possible:

- Call the CSW proxy directly with the url-encoded CSW url as a query parameter, the proxy will retrieve the document and forward the request to the proper url to retrieve the result
 - Introduce an api on which the WFS proxy can retrieve the identification of an existing endpoint or register a new endpoint and use that identification to call the CSW proxy.
- A basic implementation of the first approach is added to the catalog used in the project.

Representations

WFS proxy

This section describes how the Idproxy software, our WFS proxy implementation, represents the different types of resources.

All representations are both available via content negotiation and by adding a query parameter. The query parameter option is mainly available to support clickable links and to enable copy & paste of the URL into a browser.

Main page

HTML

The [main page](#) of an Idproxy instance simply lists the Web Feature Services that are available through the proxy using the title and abstract provided by the WFS. At the moment, it does not have any schema.org markup or any other representations.

The screenshot shows a web page with a header 'Idproxy Services'. Below the header, there are two service entries:

- INSPIRE Adressen WFS**
INSPIRE Adressen themalaag, gevuld met relevante objecten uit de Basisregistratie Adressen en Gebouwen (BAG), beheerd door het Kadaster.
- Agrarisch Areaal Nederland WFS**
Geografische afbakening van landbouwgrond in Nederland (grond die wordt gebruikt als bouwland, blijvend grasland of de teelt van blijvende gewassen).

WFS landing page

HTML

The [landing page of a WFS](#) displays the information derived from the metadata provided by the service that can be represented in the schema.org vocabulary as described [here](#). The HTML is annotated with the schema.org markup to improve indexing by search engines.

INSPIRE Adressen WFS

INSPIRE Adressen themalaag, gevuld met relevante objecten uit de Basisregistratie Adressen en Gebouwen (BAG), beheerd door het Kadaster.

Feature Types	Adressen
Keywords	Adressen
License	otherRestrictions; Geen beperkingen
Extent	3.053,47.975 7.24,53.504
WFS	http://geodata.nationaalgeoregister.nl/inspireadressen/wfs?REQUEST=GetCapabilities&SERVICE=WFS&
Provider	PDOK KlantContactCenter PDOK Apeldoorn Nederland
Metadata	http://opendatacat.net/geonetwork-geo4web/doc/dataset/3a97fbe4-2b0d-4e9c-9644-276883400dd7

XML

In addition, the resource is also available as XML. The WFS Capabilities document is returned.

WFS feature type

HTML

The HTML page of a [feature type](#) displays the information derived from the metadata provided by the service that can be represented in the schema.org vocabulary as described above in the schema.org discussion for the BAG and farmland features. The HTML is annotated with the schema.org markup to improve indexing by search engines.

In addition, it also displays the first 25 features¹⁴ of that feature type (or less, if there are less than 25 features in the dataset). Pagination controls allow users (and search engines) to browse through all features of that feature type.

Many datasets will include a large number of features. For example, the BAG dataset has about 8 million features. To support browsing by users Idproxy can also be configured to browse by subsets of a feature type. For example, for the BAG dataset, a user can select only the addresses in a municipality (“Browse by addressLocality”).

Leaflet is used to display the features shown on that page on a map. As leaflet supports GeoJSON, we simply make use of the GeoJSON representation of the resource (discussed below).

We are using [Bootstrap](#) for an “informative and pleasant to consume” layout that works on all kinds of devices.

¹⁴ 25 is the number that we used in the testbed, but in general also any other number can be used as it is applicable for the data.

Adressen

INSPIRE Adressen afkomstig uit de basisregistratie Adressen, beschikbaar voor heel Nederland

Keywords	Adressen
Extent	3.053,47,975 7.24,53,504
Metadata	http://opendatacat.net/geonetwork-geo4web/doc/dataset/76091be7-358a-4a44-8182-b4139c96c6a4
Browse By	addressLocality

« < 1 2 3 4 5 > »

Appingedam, Fivelkade W 4

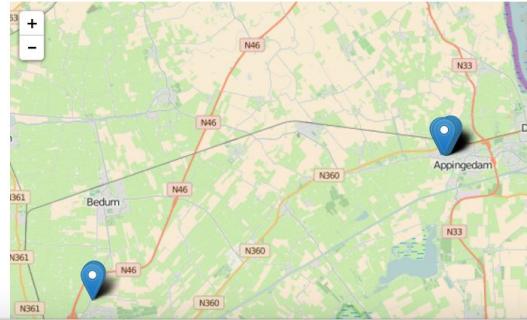
streetAddress Fivelkade W 4
addressLocality Appingedam
postalCode 9901GE

Appingedam, Fivelkade W 5

streetAddress Fivelkade W 5
addressLocality Appingedam
postalCode 9901GE

Appingedam, Fivelkade W 6

streetAddress Fivelkade W 6



JSON-LD

The JSON-LD representation is the same information as included in the HTML as annotations.

```
{
  "@context": {
    "@vocab": "https://schema.org/",
    "features": {
      "@container": "@set"
    }
  },
  "@type": "Dataset",
  "@id": "http://www.ldproxy.net/bag/inspireadressen/",
  "name": "INSPIRE Adressen WFS",
  "description": "INSPIRE Adressen themalaag, gevuld met relevante objecten uit de Basisregistratie Adressen en Gebouwen (BAG), beheerd door het Kadaster.",
  "url": "http://www.ldproxy.net/bag/inspireadressen/",
  "isPartOf": {
    "@type": "Dataset",
    "url": "http://www.ldproxy.net/bag/"
  },
  "keywords": "Adressen",
  "spatial": {
    "@type": "Place",
    "geo": {
      "@type": "GeoShape",
      "box": "3.053,47,975 7.24,53,504"
    }
  },
  "isPartOf": {
```

```

    "@type" : "DataCatalog",
    "url" :
"http://opendatacat.net/geonetwork-geo4web/doc/dataset/3a97fbe4-2b0d-4e9c-964
4-276883400dd7"
},
"features" : [ {
    "@type" : "Place",
    "@id" : "http://www.ldproxy.net/bag/inspireadressen/inspireadressen.1"
}, ...
]
}

```

Some JSON objects have been omitted for better readability.

The representation does not contain any pagination relations yet - which it should, if the JSON-LD should be used by applications, but this has not been the focus of this research. In addition, there is no clear consensus on a vocabulary to support pagination in JSON-LD, but there are examples that could be used. One example is [Hydra](#).

GML

The GML representation is the WFS feature collection of the first 25 features as returned by the WFS (and used to derive the other representations). It also includes links to previous and next pages, if applicable.

```

<wfs:FeatureCollection
next="http://geodata.nationaalgeoregister.nl/inspireadressen/wfs?NAMEPACES=x
mlns%28inspireadressen%2Chttp%3A%2F%2Finspireadressen.geonovum.nl%29&STARTIND
EX=25&COUNT=25&VERSION=2.0.0&TYPENAMES=inspireadressen%3Ainspireadressen&OUTP
UTFORMAT=application%2Fgml%2Bxml%3B+version%3D3.2&SERVICE=WFS&REQUEST=GetFeat
ure" numberMatched="8802046" numberReturned="25">
<wfs:member>
    <inspireadressen:inspireadressen gml:id="inspireadressen.1">
        <inspireadressen:straatnaam>Fivelkade W</inspireadressen:straatnaam>
        <inspireadressen:huisnummer>4</inspireadressen:huisnummer>
        <inspireadressen:woonplaats>Appingedam</inspireadressen:woonplaats>
        <inspireadressen:postcode>9901GE</inspireadressen:postcode>
        <inspireadressen:geom>
            <gml:Point srsDimension="2" srsName="urn:ogc:def:crs:EPSG::28992">
                <gml:pos>252466.7467750016 593751.8139905097</gml:pos>
            </gml:Point>
        </inspireadressen:geom>
    </inspireadressen:inspireadressen>
</wfs:member>
...
</wfs:FeatureCollection>

```

Some XML elements and attributes have been omitted for better readability.

GeoJSON

The GeoJSON representation is a feature collection of the first 25 features. Coordinates in other coordinate reference systems, e.g. RD82, have to be converted to WGS84 long/lat.

```
{
  "type" : "FeatureCollection",
  "features" : [ {
    "type" : "Feature",
    "id" : "inspireadressen.1",
    "geometry" : {
      "type" : "Point",
      "coordinates" : [6.849895040224361 ,53.32130214986961 ]
    },
    "properties" : {
      "straatnaam" : "Fivelkade W",
      "huisnummer" : 4,
      "woonplaats" : "Appingedam",
      "postcode" : "9901GE"
    }
  }, ...
  ]
}
```

Some JSON objects have been omitted for better readability.

The GeoJSON representation does not include any pagination information (out of scope for GeoJSON).

Subsets of a WFS feature type

Subsets of a feature type are established by queries that select the subset. Otherwise, their representation is identical to the representations of a feature type.

In the HTML representation a subset can be selected by name. For example, in the BAG WFS a municipality can be selected.

Adressen

INSPIRE Adressen afkomstig uit de basisregistratie Adressen, beschikbaar voor heel Nederland

Browse By addressLocality

- 's Gravenmoer
- 's-Graveland
- 's-Gravendeel
- 's-Gravenhage
- 's-Gravenpolder
- 's-Gravenzande
- 's-Heer Abtskerke
- 's-Heer Arendskerke
- 's-Heer Hendrikskinderen
- 's-Heerenberg
- 's-Heerenbroek
- 's-Heerenhoek

[In Hoofdregister](#)

Once a subset has been selected, the subset resource is made available in all representations.

Adressen in 's Gravenmoer

INSPIRE Adressen afkomstig uit de basisregistratie Adressen, beschikbaar voor heel Nederland

« < 1 2 3 4 5 > »

's Gravenmoer, Capelseweg 1

streetAddress	Capelseweg 1
addressLocality	's Gravenmoer
postalCode	5109RG

's Gravenmoer, Capelseweg 10

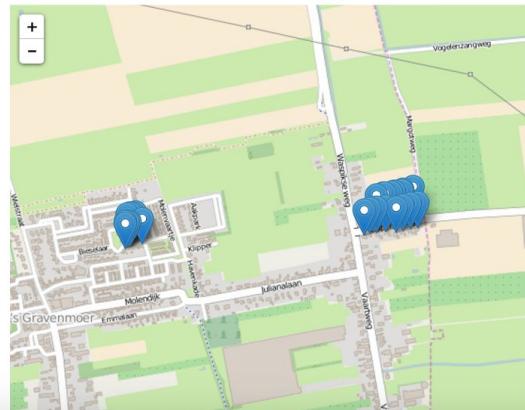
streetAddress	Capelseweg 10
addressLocality	's Gravenmoer
postalCode	5109RG

's Gravenmoer, Capelseweg 12

streetAddress	Capelseweg 12
addressLocality	's Gravenmoer
postalCode	5109RG

's Gravenmoer, Capelseweg 14

streetAddress	Capelseweg 14
---------------	---------------

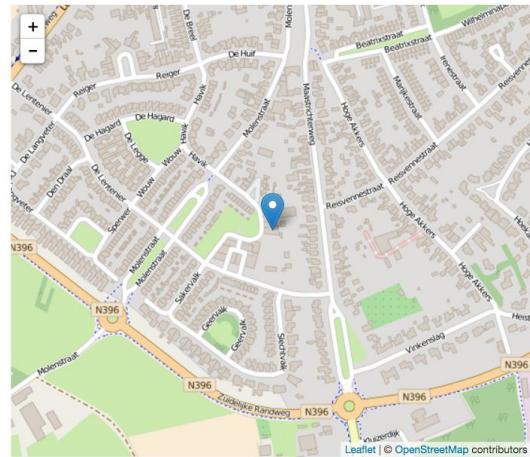


WFS feature

The HTML page of a [feature](#) displays all information associated with the feature. The schema.org vocabulary is used as the basis. I.e., the mapping to schema.org will differ by feature type. The mappings for the address and the farmland features are discussed [here](#) (which also include the JSON-LD representation). The HTML is annotated with the schema.org markup to improve indexing by search engines.

Valkenswaard, Boomvark 24

id	inspireadressen:23097
streetAddress	Boomvark 24
addressLocality	Valkenswaard
postalCode	5554LZ
latitude	51.34249325134195
longitude	5.456679823278034

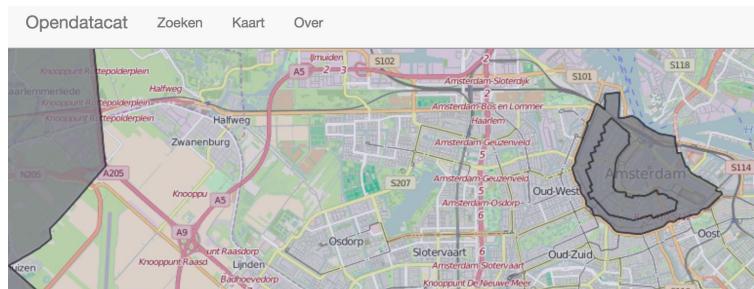


The GML and GeoJSON representations of a feature are identical to the feature in the respective feature collections on the feature type level.

CSW proxy

HTML

The html display of a CSW record is based on a very plain bootstrap template. The html is annotated with schema.org microdata. A leaflet map view is added which displays the bounds of the dataset and, if available, a WMS GetMap request is added as a map overlay. Links are provided to alternative representations of the metadata in either xml (ISO 19139) or rdf/xml (GeoDCAT-AP). If available, a link to a WFS and a link to the same WFS via Idproxy are provided.



Werelderfgoed

Standaard

Datum(publication)
2014-07-08

Identifier
f7d22d90-7651-11e0-a1f0-0800200c9a63

Samenvatting
Begrenzingen van het Nederlandse cultureel erfgoed dat vanwege haar uitzonderlijke universele waarde door het Wereld Werelderfgoedlijst. De begrenzingen zijn gebaseerd op de kaarten die destijds opgenomen waren in het nominatiedossier

RDF/XML

When using an accept-header application/xml+rdf or an explicit url /dcat/{uuid} a document encoded as rdf/xml will be returned by the service. The rdf/xml is structured as GeoDCAT-AP, using a transformation from ISO 19139 using a modified XSLT provided by the joinup GeoDCAT-AP initiative.

JSON-LD

This encoding is currently not available, it can be implemented using RDFconvert (from the sesame library) to transform from rdf/xml to JSON-ld.

Linking

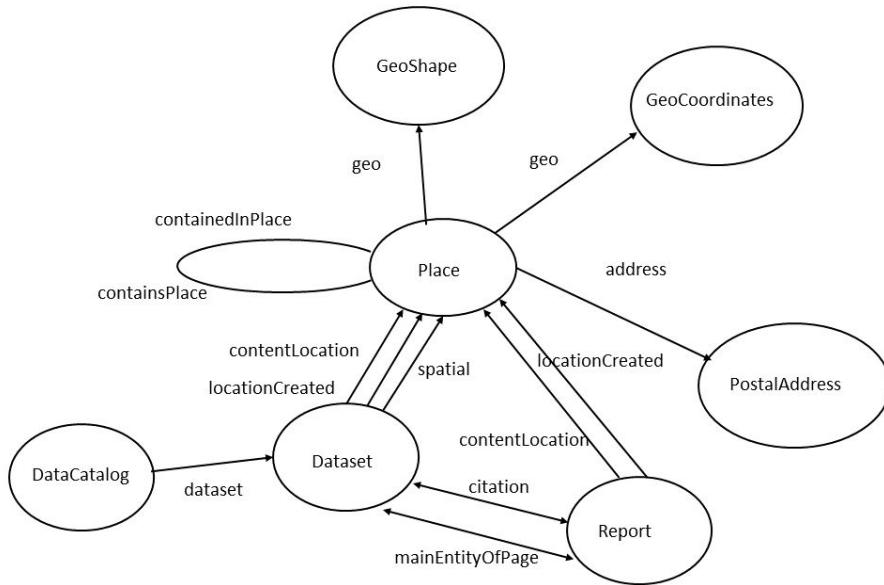
Separate data sets can be linked together to one data set if they both rely on the same principles. They have the same data framework, in this case RDF, and they use the same data modeling fundamentals, in our case Semantic Web modeling principles. The RDF framework dictates that the data is recorded according to the triple sequence of subject, predicate, object. Semantic Web data modeling principles are based on semantic statements, implemented by logical rules that can be validated by dedicated tools such as the structured data validators of the respective schema.org sponsors.

Links are established via HTTP resolvable identifiers in RDF triple statements (example: "<http://www.ldproxy.net/bag/inspireadressen/inspireadressen.1/>"). When these addresses are resolved the RDF data must be retrieved.

To improve discoverability, backward and forward linking is advised.

Semantic workspace

The image below shows which semantic constructions are valid in the context area that we are working. The direction of the arrows is mandatory. Links can only be made to or from the circles.



This learns us that for instance we could in principle create the following relations (links):

A schema:Dataset can be linked to a schema:Report ,used in Bekendmakingen, with the schema:citation property (implemented).

A schema:Report can be linked to a Place in the BAG or Farmland dataset using the property [schema:contentLocation](#). In triples:

```
Subject = <http://the_Bekendmakingen_base_uri/the_Report_identifier>
Predicate =schema:contentLocation
Object =<http://the_BAG_base_uri/the_Place_identifier>
```

As JSON-LD:

```
{"@context": "http://schema.org",
 "@id": "<http://the_Bekendmakingen_base_uri/the_Report_identifier>",

 "contentLocation": {
 "@id": "<http://the_BAG_base_uri/the_Place_identifier>"
 }
}
```

A Place recorded in the Farmland dataset can be linked to a PostalAddress in the BAG dataset using the schema:address property. In JSON-LD:

```
{
  "@context": "http://schema.org",
  "@id": "<http://the_AAN_base_uri/the_Place_identifier>",
  "address": {
    "@id": "<http://the_BAG_base_uri/the_PostalAddress_identifier>"
  }
}
```

The linking possibilities are numerous, and as long as they follow the semantic workspace anything is possible. Meaning: it is in principle also possible to link the Farmland Places with Bekendmakingen PostalAddresses, provided that all these resources have URIs and are not Blank Nodes.

The Report and Dataset resources are a special case, since they are both types of CreativeWork. That means that the properties citation and mainEntityOfPage work in both directions.

Link storage location

Linking datasets means creating triple statements that connect the source and the destination via a property. Where to store these statements? The answer to this question is a matter of data maintenance and link discovery. Technically it does not matter where these statements are located. In the Linked Data cloud there are examples of outgoing links, incoming links and so called Linksets. Linksets are collections of triple statements that represent the connecting links. A linkset can be located anywhere in the Linked Data cloud. As long as the Linkset can be discovered, the connection between the source and destination datasets can be established.

Establishing links

Where the planned use cases, in particular for the new environmental law, expect links between datasets, it will be valuable, if one can *automatically establish* the links between resources. Otherwise, creating and maintaining the expected links might become complicated.

In the case of the links from BAG resources to Bekendmakingen features, [links are dynamically identified in the WFS proxy using SPARQL queries on the Bekendmakingen dataset](#). This supports that the links included with the BAG data are always up-to-date.

Link maintenance

There are tools developed (e.g. semantic pingback) to overcome problems with links. Link issues are not Linked Data issues, they are Web issues. With the Web of data we are at a new level due to the fine-grained nature of the data and the number of new resources compared to the web of documents, but the fundamentals are unchanged. We have solved

broken links in the document Web for many years now, and we are used to it. It is the responsibility of the data provider to maintain, as much as possible, redirects to deprecated links. The user, or software, that downloaded the data must however understand that his version is detached from the data web ecosystem and might become orphaned over time. After all, a broken link in the Web of data is nothing different than a broken link in the Web of documents.

Links from BAG to Bekendmakingen features

The Bekendmakingen features contain information that allows us to dynamically discover links from the municipality (woonplaats) and post code resources that are provided by the BAG WFS proxy.

When a BAG WFS proxy resource with a configuration for dynamically establishing links is being loaded, it will send a SPARQL query to the Bekendmakingen endpoint in the background. The configured query can contain substitution points for values from the proxy query or from the WFS response.

The query to establish the links from the municipality would look like this:

```
prefix schema: <http://schema.org/>
SELECT ?id, ?title
WHERE { GRAPH <http://data.linkeddatafactory.nl/bekendmakingen/> {
  ?id schema:contentLocation [ schema:address [ schema:addressLocality
    """municipality$$" ] ]
  ?id schema:headline ?title
} }
```

The substitution point “\$\$\$municipality\$\$\$” would then be replaced with the attribute value, e.g. “Valkenswaard”, and the response from the Bekendmakingen endpoint would look like this:

```
{ "head": { "link": [], "vars": ["id", "title"] },
"results": { "distinct": false, "ordered": true, "bindings": [
  { "id": { "type": "uri", "value":
    "http://data.linkeddatafactory.nl/bekendmakingen/1000" }, "title": {
      "type": "literal", "xml:lang": "nl", "value": "Strodekkerwei thv 2" }},
  { "id": { "type": "uri", "value":
    "http://data.linkeddatafactory.nl/bekendmakingen/1010" }, "title": {
      "type": "literal", "xml:lang": "nl", "value": "Tinnegieterwei thv 19" }},
  ...
] } }
```

The ids and titles are then parsed from the response and inserted into the BAG WFS proxy resource.

As a result, when we have a municipality or post code with associated public announcements, these are displayed. The [schema:headline](#) and the local id of the Bekendmakingen feature are used.

Adressen in Valkenswaard

INSPIRE Adressen afkomstig uit de basisregistratie Adressen, beschikbaar voor heel Nederland

Announcements

- Strodekkerwei thv 2 (1000)
- Tinnegieterwei thv 19 (1010)
- Larksbeemd plantsoen naast nr. 11 (1020)
- Koolzaadbeemd, plantsoen thv nr. 10 (1030)
- Victordal, plantsoen thv nr. 35 (1040)

[... show more](#)

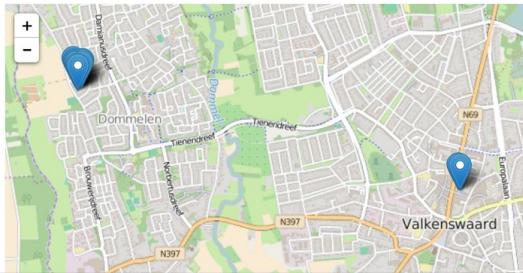
« < < 1 > > »

Valkenswaard, de Molensteen 6

streetAddress de Molensteen 6
addressLocality Valkenswaard
postalCode 5554VC

Valkenswaard, de Molensteen 18A

streetAddress de Molensteen 18A
addressLocality Valkenswaard
postalCode 5554VC



A click on a link leads to the announcement.

1000 at [Linked Data for Testbed Spatial Data on the Web](#)
<http://data.linkeddatafactory.nl/bekendmakingen/1000>



schema:contentLocation	< http://data.linkeddatafactory.nl/bekendmakingen/1000Valkenswaard >
schema:description	▪ Omgevingsvergunning aangevraagd - Strodekkerwei thv 2 (nl)
schema:genre	▪ omgevingsvergunning (nl)
schema:headline	▪ Strodekkerwei thv 2 (nl)
rdf:type	▪ schema:Report
schema:url	http://www.valkenswaard.nl/internet/bekendmakingen_3361/item/strodekkerwei-thv-2_50953.html

[As Turtle](#) | [As RDF/XML](#)

Links between the spatial data and metadata

First we will introduce how catalog records are typically linked to datasets served through services in an OGC / ISO/TC 211 / INSPIRE-based Spatial Data Infrastructure. From a single OGC service endpoint (WMS, WFS, ...) typically one or more datasets are exposed. Each service contains a GetCapabilities operation, which does amongst other things return the available list of datasets in that service. For each dataset a link to an ISO 19115 (profile on ISO 19139) metadata document describing that dataset is provided (usually via a CSW GetRecord operation). Also other metadata elements describing the service are available in the getcapabilities document (contactpoint, license, available operations, formats and projections). INSPIRE requires some additional metadata elements, but allows part of the metadata to be moved into an external ISO 19119 (profile on ISO 19139) document, to which then a link should be established from the capabilities document. Such an ISO 19119 document duplicates links to all the datasets that are available in the service, with their corresponding metadata links. Besides the links based on metadata url's, the INSPIRE technical guidelines suggests a supplementary linking mechanism; the dataset identifier and authority. The capabilities document as well as the metadata documents should contain this information.

Different INSPIRE member states have implemented different conventions around the authority/identifier linkage, hopefully a soon-to-be-released new version of the metadata technical guidelines will provide central guidance to align these conventions. From the perspective of this research the proposal to concatenate the authority, a '#' or '/' and the dataset identifier to be used as a dataset identification (uri), makes most sense.

In the services that were used as input for this research we identified a number of issues with the authority/identifier linkage (Agrarisch Areaal is not an INSPIRE dataset, the inspireadressen wfs capabilities SpatialDataSetIdentifier does not match with the metadata dataset identifier). So we've only used metadata identifiers and service endpoints to discover links between datasets, services and metadata.

We need to lift the existing links between the spatial dataset (e.g., the BAG WFS) and the metadata (e.g., the nationaalgeoregister.nl Catalog Service) from the "SDI layer" to our more 'webby' proxy layer that includes HTML representations with schema.org annotations.

I.e., links from the spatial data made available via the WFS proxy have to go to the Catalog Service proxy, not the Catalog Service itself - and the other way round. The XML provided by the Catalog Service would not be helpful to search engine crawlers or users. I.e., we want to link to the URL of the dataset metadata as it is made available by [the proxy service for the Catalog Service](#), as that representation of the dataset metadata provides HTML with schema.org annotations and/or JSON-LD.

Let us take the link from a WFS capabilities document to a metadata record in the Catalog Service as an example. For the BAG WFS we find the URL

<http://www.nationaalgeoregister.nl/geonetwork/srv/dut/csw?service=CSW&version=2.0.2&request=GetRecordById&outputschema=http://www.isotc211.org/2005/gmd&elementsetname=full&id=3a97fbe4-2b0d-4e9c-9644-276883400dd7>

We use this URL to construct the URL to the corresponding dataset metadata resource in the Catalog Service proxy. This requires knowledge how the URLs of the Catalog Service proxy are constructed. In this case, the UUID value of the id query parameter is extracted and used in the Catalog Service proxy URL.

In case we detect the metadata link in WFS capabilities references a resource outside nationaalgeoregister.nl, a notification has to be send to the CSW proxy to retrieve that resource, before we can ask it to transform it to the required schema/format.

Implementations

Idproxy

Idproxy is our implementation of the WFS proxy. It is reusing two libraries that interactive instruments has previously developed for its [XtraProxy for WFS product](#). These libraries

have been published under an Apache license as part of our contributions to the testbed and are available on GitHub:

- [XtraPlatform](#): implements basic proxy capabilities
- [XtraPlatform-SDI-Tools](#): provides the WFS client, coordinate transformations, etc.

[Idproxy](#) extends these libraries with support for the WFS proxy capabilities described in this report. The software is also available on GitHub and published under the Apache license. Interactive instruments will add sufficient documentation on GitHub so that interested parties can set up own deployments of Idproxy.

The implementation supports

- a configurable mapping from the GML application schema to JSON-LD / schema.org
- the configuration of additional, intermediate resources
- the configuration of rule-based additional links to external resources
- JSON-LD and HTML generation
- a read-only REST interface

Several ideas for enhancements have been identified during the testbed. These are recorded as [issues on GitHub](#).

In principle, such a proxy could be operated by a central SDI operator or individual data providers. An advantage of doing it “close” to the dataset owner would be the domain knowledge necessary for the appropriate schema.org mapping and for configuring other aspects like subsetting resources or how to establish links dynamically for known use cases and related datasets. In addition, the URLs could be in a domain owned by the dataset owner.

GeoNetwork

The technology used in this project to query, store and transform metadata is GeoNetwork opensource. GeoNetwork is an opensource catalogue, common in spatial data infrastructures on local to global scale.

GeoNetwork uses a generalised lucene index to query metadata content structured as a number of pluggable schema's (ontologies). The metadata itself is stored as xml-blobs in a database. GeoNetwork currently has three technologies to transform documents between schema's and encodings; groovy, xslt and jsonix¹⁵. For this project we decided to use XSLT to convert XML according to ISO 19139 to microdata annotated html, crawlable by the search engines.

The latest version of GeoNetwork has a new feature, called formatters¹⁶, which provides an optimal framework for the research. Formatters are renderers based on groovy or xslt that can be plugged into schema-plugins to provide specific rendering for any metadata profile. Output of formatters is cached on the server, so subsequent requests for the same resource

¹⁵ <https://github.com/highsource/jsonix>

¹⁶ <https://github.com/geonetwork/core-geonetwork/wiki/Schema-Formatters-and-Metadata-Views>

are very fast. This is optimal for search engine optimisation, because search engines usually have high penalties in search ranking for slow responding websites.

The dutch metadata profile implementation in nationaalgeoregister.nl (an instance of GeoNetwork) is split into 2 schema-plugins; ISO 19115 and ISO 19119. For each of these profiles a separate schema.org formatter can be developed.

To validate the optimised crawlability of GeoNetwork we have created an experimental setup at <http://opendatacat.net/geonetwork-geo4web>. On that url a modified instance of GeoNetwork has been installed. The sitemap service has been modified to facilitate search engine crawling and the html display has been optimised to include schema.org annotations. The instance has harvested the full content of nationaalgeoregister.nl. The sitemap has been registered at Google and Bing and a Google analytics script has been added. From that point in time the crawling and indexation status has been monitored on Google and Bing Webmaster tools and the Google Structured data testing tool¹⁷.

Validation and testing

Indexing by search engines

Search engines have similar approaches to register content, which is then crawled by the search engine. Google has the widest set of tools to evaluate the crawlability of the Web pages. A small description of the used tooling is below.

Submit a url

Search engines offer an option to submit a url. The search engine will then start the crawl process by following links through that website.

Sitemap

By providing an xml sitemap containing links to each of the resources the initial crawling process can be speeded. Sitemap allows for additional options, such as indicating the last modification date of a resource, so the crawler knows if it has the latest version. CSW Metadata generally has such a date, but that date is not available for WFS features.

A sitemap has a max size of 2500 web pages, the sitemap specification however supports the concept of pagination. In the previous version of the specification, a sitemap allowed to contain a spatial extent for a resource, but this functionality is not supported operational anymore.

A sitemap will also help if a website uses dynamic content (ajax) to display search results. Google is not able to crawl dynamic content, so would not be able to navigate through the search results.

¹⁷ <https://developers.google.com/structured-data/testing-tool/>

Robots.txt

This file is placed at the root of the website and can be used to limit the crawler to not index certain parts of the website. First crawl results indicated that the crawlers also crawled the wfs xml content. this behaviour can be prevented in robots.txt.

Crawl errors

Google offers a dashboard on which webmaster can monitor the crawling process. Get an impression how many pages are indexed, but also see on which pages broken links occur. Especially the catalogue pages showed a huge number of broken links. Most of them related to bad links in the metadata.

Structured data

One of the options in google webmaster tools is the option to display structured data identified in the crawled pages. It seems this functionality is quite critical, google will indicate it hasn't found any structured data if the structured data has a schema-error (todo: check).

Google analytics

By adding google analytics the idea was to allow for user monitoring, which route do users (including the search engines themselves) take through the website.

Content loaded with ajax

Search engines are not able to crawl content that is dynamically added to the website (ajax). Modern websites that use a lot of ajax usually duplicate the content into static web pages to facilitate the search engines. For example most of the Dutch national georegistry is not crawled by google because ajax is used to display the search results and there is no static alternative. The same applies for maps which display vector features. These features are usually loaded with ajax and are not harvested by search engines.

Tests with Yandex and Bing gave similar results as Google, although they usually are less capable to identify the full schema-org annotations in the web pages.

Below we describe our experiences with search engines and the proxies.

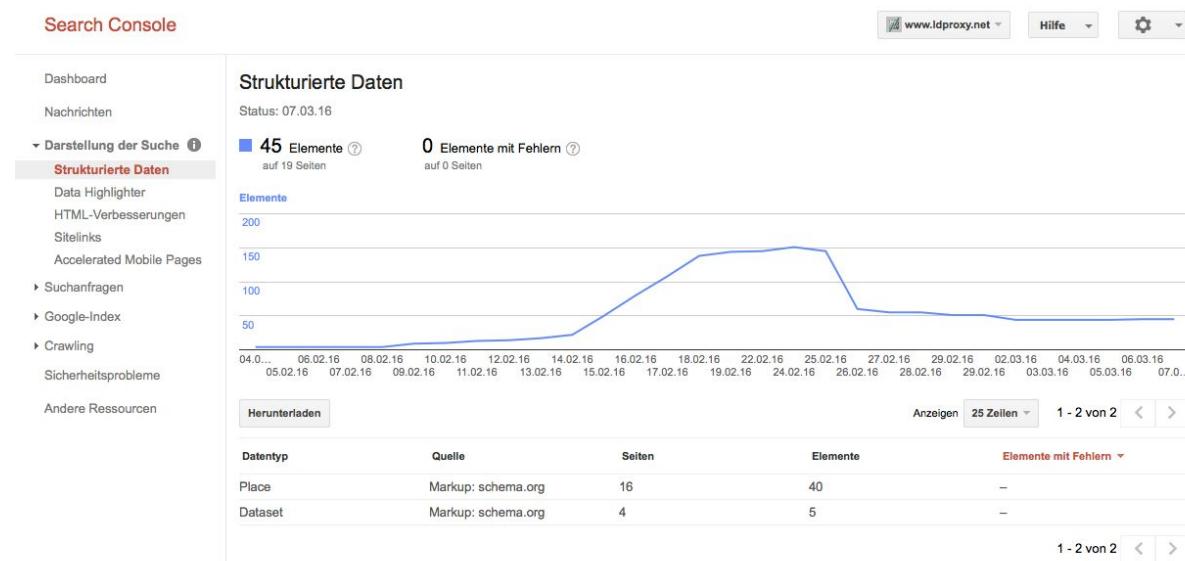
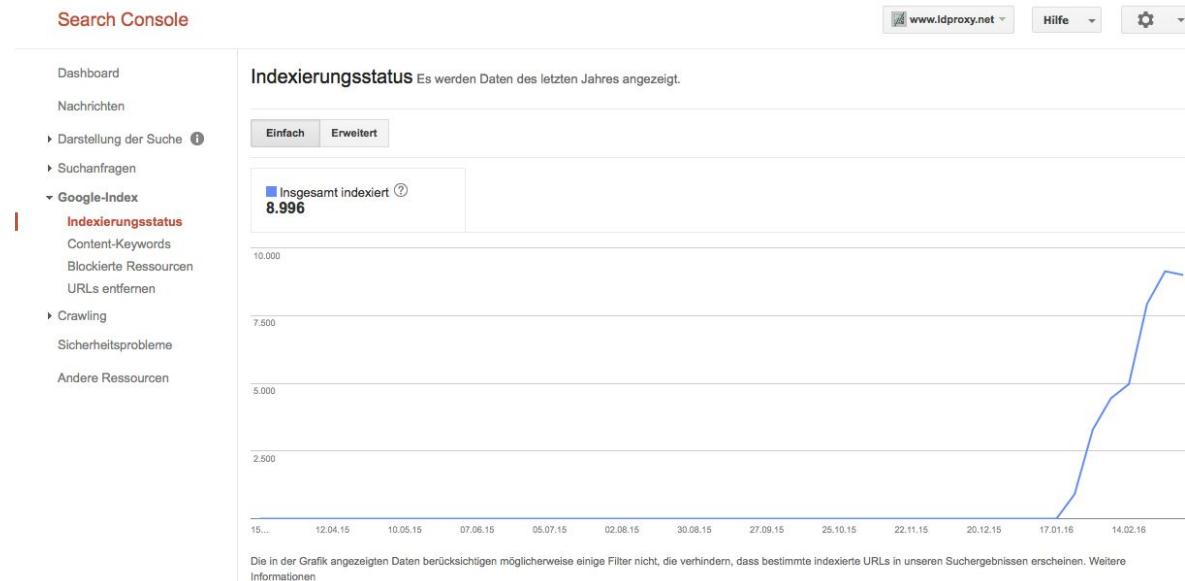
Indexing results

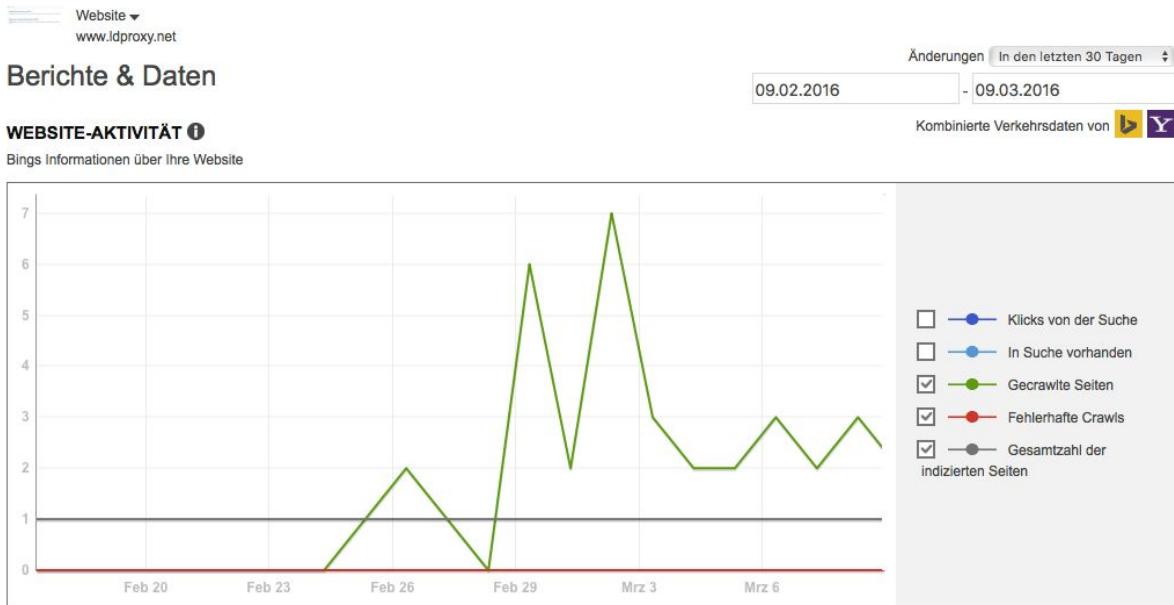
We have registered the WFS proxies in the webmaster consoles of Google and Bing to monitor the indexing progress with two major search engines. The WFS proxy resources have been made available in mid January 2016. At the time of writing (early March 2016), only very few pages have been indexed. The BAG dataset has around 80 million features, the farmland datasets around 50,000 features, each provided in multiple representations.

Bing has crawled a total of 7 pages and indexed one. Google has indexed around 9000 pages, but only 45 structured data resources.

From the indexing we can see that Google indexes HTML and XML representations, JSON and JSON-LD is not indexed.

Interestingly, Google also “forgets” pages and structured data that has been indexed. All crawl results were reported as error-free. The reasons for dropping indexed results again, or how to avoid such drops, are not clear to us.





On the crawling results of the catalog we noticed similar results. Google claims to have crawled and indexed all 5500 catalogue records, but has only found 12 pages with structured data content. Bing crawled 185 pages (of 5500 submitted) and indexed 5. The diagram below shows a steep flat line, instead of the curvy line in the diagram above for Idproxy. This is related to the fact that all catalogue url's were submitted at once, using a sitemap, where Idproxy let's the crawler locate new links to crawl.



For Idproxy no sitemaps are created for the spatial datasets as the sheer number of features (pages) of many spatial datasets seems to indicate that sitemaps, which are limited to 2500 entries per sitemap, are probably not the right mechanism. In addition, dynamically creating sitemaps for spatial datasets may also be a challenge and would involve a crawling through the whole datasets via the WFS to create/update the sitemap.

Ranking in search results

We have tried to investigate, if and in how far the annotations improve the ranking in searches in search engines. This turned out to be really difficult, at least with the (calendar) time available in the testbed as most of the pages have not been indexed yet by search engines as documented above. As a result, we do not have solid evidence, just some indications.

We have executed several searches to test, if data available via the proxies can be discovered through search engines. In all searches we use a different order of the words used in the search than the order in the page name.

In the first search we search using a street name, house number and woonplaats of a BAG feature where the structured data has been indexed. In this search the BAG feature page is listed as the first hit.

The screenshot shows a Google search results page for the query "Wilhelminastraat 15a Didam". The search bar at the top contains the query. Below it, the "Alle" tab is selected, followed by other options like Maps, Shopping, Bilder, News, Mehr, and Suchoptionen. A search button and user profile are also visible. The main content area shows a map of Didam with a red marker indicating the location of Wilhelminastraat 15a. Below the map, the address "Wilhelminastraat 15, 6942 BH Didam, Niederlande" is listed, along with a "Street View" thumbnail and a "Wegbeschreibung abrufen" link. A "Mit dieser Adresse" section follows, containing the name "Uitvaartsprekster Witta Veerink". Further down, a "Didam, Wilhelminastraat 15a - Services" section is shown, listing links to "www.idproxy.net/bag/.../inspireadressen.8795606/" and "https://www.huislink.nl/alle.../didam/wilhelministraa...". Other search results include a link to a Funda page for a house on Wilhelminastraat 96 in Didam.

Wilhelminastraat 15a Didam

Ungefähr 1.480.000 Ergebnisse (0,60 Sekunden)

Wilhelminastraat 15, 6942 BH Didam, Niederlande

Mit dieser Adresse

Uitvaartsprekster Witta Veerink

Didam, Wilhelminastraat 15a - Services

Wilhelminastraat, Didam - Alle adressen | Huislink - Het ...

Huis te koop: Wilhelminastraat 96 6942 ET Didam [funda]

From the response by Google we can also see that the address is identified as such and that the search engine manages information associated with the address. However, unlike with other types of data like events or reviews, information about places ([schema:Place](#)) is not supported by Google with [rich snippets](#).

Using the recently published [Google Knowledge Graph Search API](#) we can also determine that our structured data is not added to the Google Knowledge Graph. The searches using the API only return curated content.

In the next search we use an address where the HTML page has been indexed by Google, but not the structured data. In this search the BAG feature page is listed as the second entry.

Google Aduarderdiep 1014 Den Ham

Alle Maps Shopping News Bilder Mehr ▾ Suchoptionen

Ungefähr 13.200 Ergebnisse (0,54 Sekunden)

Aduarderdiep 1014, 9833 TG Den Ham, Niederlande

Map data ©2016 Google

Aduarderdiep 1014, Den Ham - Postcode.nl

https://www.postcode.nl/9833TG/1014 ▾ Diese Seite übersetzen

Aduarderdiep 1014. id: inspireadressen.944. streetAddress: Aduarderdiep 1014. addressLocality: Den Ham. postalCode: 9833TG

Kaart Satelliet | Kaartgegevens: © CBS, Kadaster, ...

Den Ham, Aduarderdiep 1014 - Services

www.idproxy.net/bag/inspireadressen/inspireadressen.944/ ▾

Den Ham, Aduarderdiep 1014. id: inspireadressen.944. streetAddress: Aduarderdiep 1014. addressLocality: Den Ham. postalCode: 9833TG

Verkocht: Aduarderdiep 5 9833 TG Den Ham Gn [funda]

www.funda.nl > ... > Verspreide huizen Aduard ▾ Diese Seite übersetzen

11.11.2015 - Verkocht: Aduarderdiep 5 9833 TG Den Ham Gn. Zoek een koopwoning, huurwoning, nieuwbouwwoning of recreatiewoning op funda, ...

Postcode 9833TG Aduarderdiep Den Ham

www.postcodeboek.info/.../Aduarderdiep-Den+Ham ▾ Diese Seite übersetzen

Postcode 9833TG. Aduarderdiep in Den Ham. Gemeente: Zuidhorn Provincie: Groningen. Nummers: 1, 3, 5, 7, 9, 1002, 1004, 1006, 1008, 1010, 1012, 1014, ...

When we search for an address where the HTML page has not yet been indexed by Google, the BAG feature page is (unsurprisingly) not listed.

Google Aambeeld 2 Valkenswaard

Alle Maps Shopping Bilder Videos Mehr ▾ Suchoptionen

Ungefähr 33.300 Ergebnisse (0,79 Sekunden)



Aambeeld 2, 5554 CV Valkenswaard, Niederlande Wegbeschreibung abrufen

Aambeeld 2, Valkenswaard (5554CV) | Huislink - Het ...
<https://www.huislink.nl/valkenswaard/aambeeld/2/> ▾ Diese Seite übersetzen
Is Aambeeld 2, Valkenswaard (5554CV) jouw woning? Meld hem aan bij Huislink. Flirt met woningen die *jij* leuk vindt!

Goedkoop tanken bij Aambeeld 2 - 10 Valkenswaard, Noord ...
www.brandstof-zoeker.nl ▾ ... ▾ Aambeeld ▾ Diese Seite übersetzen
Vind goedkope benzine en goedkope diesel in Nederland - Noord-Brabant - Valkenswaard - Aambeeld. Laagste benzine- en dieselprijs 2 - 10 (5554CV) van ...

5554 CV | Aambeeld (Valkenswaard) - Drimble
drimble.nl ▾ ... ▾ Wijk ▾ Buurt ▾ Diese Seite übersetzen
Deze postcode hoort bij de straat Aambeeld in Valkenswaard, gelegen in de buurt ...
Huisnummers: 2, 4, 1, 3, 5, 7, 9, 11, 13, 17, 10, 21, 23, 25, 27, 29. Plaats ...

Aambeeld 2 in Valkenswaard | 5554CV | Kadasterdata
valkenwaard.kadasterdata.nl ▾ ... ▾ Aambeeld ▾ Diese Seite übersetzen
Alles weten over Aambeeld 2 in Valkenswaard? Dit gebouw is in 1958 gebouwd en heeft een inhoud van 270m3. De waarde van dit pand is...

The same is the case in Bing:

The screenshot shows a Bing search results page with the following details:

- Search Query:** Wilhelminastraat 15a Didam
- Results Count:** 3.110 ERGEBNISSE
- Filter Options:** Datum ▾ Sprache ▾ Region ▾
- Results:**
 - Postcode 6942BH De Schutterije Didam - ...** [Diese Seite übersetzen](https://www.zoekplaats.nl/postcode/6942BH/Wilhelminastraat/Didam)
www.zoekplaats.nl/postcode/6942BH/Wilhelminastraat/Didam ▾
Postcode 6942BH De Schutterije Didam ... Postcode 6942BH De Schutterije / Wilhelminastraat in Didam ... 11, 12, 13, 14, 15, 15 a, 16, 17, 18, 19 ...
 - Protestantse gemeente te Didam - Pastoraat** [Diese Seite übersetzen](https://www.pkn-didam.nl/over-ons/organisatie/taakgroepen/pastoraat)
https://www.pkn-didam.nl/over-ons/organisatie/taakgroepen/pastoraat ▾
Protestantse Gemeente Didam PKN. Informatie over pastoraat, ... Didamseweg 15A, ... 4 Wilhelminastraat nr. 2-74 - 8 Wilhelminastraat nr. 3-71 - 8 ...
 - Hendriks Didam | De Telefoonongids** [Diese Seite übersetzen](https://www.detelefoonongids.nl/hendriks/didam/7-1)
www.detelefoonongids.nl/hendriks/didam/7-1 ▾
32 resultaten voor Hendriks in Didam vindt u in De Telefoonongids. Telefoonnummer, adres, postcode, routeplanner en plattegrond - detelefoonongids.nl »
 - Ontwerpen stickers Didam - ...** [Diese Seite übersetzen](https://www.telefoonboek.nl/zoeken/ontwerpen-stickers/didam)
www.telefoonboek.nl/zoeken/ontwerpen-stickers/didam ▾
Ontwerpen stickers Didam op Telefoonboek.nl - Naast een ontwerpen stickers vind je ook ingenieur, reclamebureau, software, interieurbouw, ontwerpstudio
 - Jansen Didam | De Telefoonongids** [Diese Seite übersetzen](https://www.detelefoonongids.nl/jansen/didam/7-1/?page=3)
www.detelefoonongids.nl/jansen/didam/7-1/?page=3 ▾
75 resultaten voor Jansen in Didam vindt u in De Telefoonongids. Telefoonnummer, adres, postcode, routeplanner en plattegrond - detelefoonongids.nl »
 - Radio brabant Didam - Telefoonboek.nl - ...** [Diese Seite übersetzen](https://www.telefoonboek.nl/zoeken/radio-brabant/didam)
www.telefoonboek.nl/zoeken/radio-brabant/didam ▾
Radio brabant Didam op Telefoonboek.nl ... Wilhelminastraat 34 ... Schoolstraat 15-A . 6901HD ...
 - PostcodeZoeken.net - Petuniastreet** [Diese Seite übersetzen](https://www.postcodezoeken.net/Straat/Gelderland/Didam/Petuniastreet)
www.postcodezoeken.net/Straat/Gelderland/Didam/Petuniastreet ▾
Hier vind je informatie over de straat Petuniastreet in Didam, zoals de provincie, gemeente, plaats, locatie en alle postcodes in Petuniastreet.
 - Favoriet FM | Didam: spoorwegovergang ...** [Diese Seite übersetzen](https://rttfavoriet.nl/didam-spoorwegovergang-bievankweg-binnenkort-afgesloten)
rttfavoriet.nl/didam-spoorwegovergang-bievankweg-binnenkort-afgesloten ▾
Didam: spoorwegovergang ... de Hengelderweg en Wilhelminastraat. Fietsers worden

The catalogue has fierce competition in the search rankings from other catalogues that harvest the same content, but have a longer history. With a search term “wederopbouwgebieden” for example, our page is not on the first pages of google. Adding the term “dataset” shows the page at rank 6 (behind data.overheid.nl and europeandataportal).

dataset wederopbouwgebieden

Alle Afbeeldingen Video's Maps Nieuws Meer ▾ Zoekhulpmiddelen

Ongeveer 144 resultaten (0,49 seconden)

Wederopbouwgebieden - Datasets - Data.overheid.nl
<https://data.overheid.nl/data/dataset/wederopbouwgebieden> ▾
Wederopbouw is één van de vijf onderdelen van de Visie Erfgoed en Ruimte. Als onderdeel van het speerpunt wederopbouw zijn in door de Rijksdienst voor ...

Verwant - Wederopbouwgebieden - Datasets - Data ...
<https://data.overheid.nl/data/dataset/wederopbouwgebieden/related>
Ministerie van Onderwijs, Cultuur en Wetenschap | Rijksdienst voor het Cultureel Erfgoed. Licentie: CC-BY 3.0. Catalogus: Nationaal Georegister. Permalink:...

Activiteiten stroom - Wederopbouwgebieden - Datasets ...
<https://data.overheid.nl/data/dataset/activity/wederopbouwgebieden> ▾
Nationaal Georegis... heeft de dataset **Wederopbouwgebieden** bijgewerkt 4 uren geleden. Nationaal Georegis... heeft de dataset **Wederopbouwgebieden** ...

Amsterdam Westelijke Tuinsteden: een ...
<cultureelerfgoed.nl/.../amsterdam-westelijke-tuinsteden-een-wederopbouwgebieden> ▾
Deze brochure laat zien waarom de Westelijke Tuinsteden in Amsterdam een van de 30 **wederopbouwgebieden** van nationaal belang is. De Westelijke ...

Contact - European Data Portal
<www.europeandataportal.eu/.../form?...Dataset.../data...> ▾ Vertaal deze pagina
Dataset URL: <http://www.europeandataportal.eu/data/en/dataset/wederopbouwgebieden>. Name *. Please enter your name. E-mail *. Please enter your email.

Wederopbouwgebieden
<opendatacat.net/.../doc/dataset/5e2f489b-0a94-4375-a6f2-e3c6fd9eaefa> ▾
Wederopbouwgebieden ... Aanvullende informatie: [Indien aanwezig: URL van het document dat de datastructuur van de **dataset beschrijft**]. Referentiesysteem ...

Another interesting aspect is observed in image search. Search engines offer options to query only images. The search engines will not find any imagery from our test setup, since our metadata is linking to imagery that is hosted at the data provider or at [nationaalgeoregister.nl](#). Our setup however does help in making that imagery available in the search engine index.

One of the original goals of the research was to explore, if the search engine can also be used to help answering more structured questions, e.g. “buildings in Valkenswaard built before 1960”. As the structured data is not available via the Google Knowledge Graph Search API and searching using spatial and temporal aspects does not seem to be supported in general, our conclusion is that currently the search engines can only be used to identify resources by searching via keywords.

Proxy performance

The performance of the WFS proxy is basically the same as the performance of the underlying WFS. The overhead for processing the WFS responses is minimal. The proxy of the farmland WFS (<http://www.ldproxy.net/aan/>) has significantly less latency in its responses than the BAG WFS proxy (<http://www.ldproxy.net/bag/>), which is a result of the higher latency of the BAG WFS.

The load times of pages will not only impact the user experience, but it may have an impact on how search engines crawl sites and how they rank results, too, see for example this [blog post](#). The usability of proxies will therefore directly depend on the performance of the underlying SDI components.

Data from the WFSs is not cached - with one exception. When defining browsable subsets, like with the municipalities and the post code in the BAG dataset, it is necessary to periodically determine and cache the list of named subsets in the proxy. It is not possible to determine the named subsets in a dynamic way that would support an interactive use of the proxy by a user.

The CSW proxy uses a html cache provided by the Formatter framework. The benefit is that the proxy is very fast on subsequent requests. The initial request will have similar performance challenges as the WFS proxy.

Understandability by non-experts

The HTML pages have been designed to use layouts that users are familiar with. [Screenshots of the layouts](#) are included in this report.

An aspect that we know could be improved in the HTML representation, would be the use of human readable labels where currently type and attribute names are used. For example, we could write “woonplaats” / “municipality” instead of the schema.org property name “addressLocality”.

Also we have to be aware that users will consider each proxied WFS or CSW record as an average website. It’s important to add a statement on each page indicating the purpose and origin of the data represented on that page. A related aspect is that for some datasets, like a network topology or heatmap, presenting individual records as webpages doesn’t make much sense, the added value of the set is in the combined representation of multiple records.

Interoperability with the Linked Data Cloud

The Linked Data Cloud is established on the principles originally formulated by Tim Berners-Lee¹⁸. One can test if this all works as expected by executing a series of tests.

The requirements from Tim Berners-Lee are:

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.

1 and #2 are self explanatory and do not need any tests. # 3 is somewhat cryptic and #4 is mostly understood, but also sometimes not correctly implemented.

How #3 can be implemented is described, amongst others, on the [Data on the Web Best Practices](#) . There are some tests that you can do, to check if everything works ok, these are described in the next section.

#4 might sound simple, however there are important requirements:

- a) the link URLs must dereference to RDF, otherwise the data can not be queried
- b) the semantics of the link must validate correctly. Example: The expected value of a property type must comply with the specification. See image below for schema:addressLocality. The object in the triple must be of type text and can not be an instance of a Class. This automatically implies that you cannot link to the object of

¹⁸ <https://www.w3.org/DesignIssues/LinkedData.html>

schema:addressLocality :



addressLocality

[Thing](#) > [Property](#) > [addressLocality](#)

The locality. For example, Mountain View.

Usage: Between 500,000 and 1,000,000 domains

Values expected to be one of these types
Text

Used on these types
PostalAddress

Test method

To test #3 it you can use a tool that can make HTTP requests, such as cURL¹⁹. With cURL you can test if a resource responds correctly to content negotiation for RDF. The application type should be either ld+json or rdf+xml (or any other RDF serialisation, but they are not used in this testbed).

A screenshot of a Windows Command Prompt window titled "Command Prompt". The window shows two separate curl commands being run. The first command is "curl -I -H "Accept: application/ld+json" http://www.ldproxy.net/bag/inspireadressen/inspireadressen.22410" and the second is "curl -I -H "Accept: application/rdf+xml" http://opendatacat.net/geonetwork-geo4web/dcat/88a3ciae-902e-4efa-8663-d". Both commands return a 200 OK status code and standard HTTP headers like Date, Server, Pragma, Cache-Control, Expires, Content-Type, Transfer-Encoding, Set-Cookie, and Via.

To test #4 a) properly, one must install a SPARQL client that supports querying a web resource via its persistent URI address. [Jena/ark](#) is such a tool. It runs from the command line with the following syntax:

¹⁹ <https://en.wikipedia.org/wiki/CURL>

```
> arq --data <the resource URL> --query <a local file with a query statement>
```

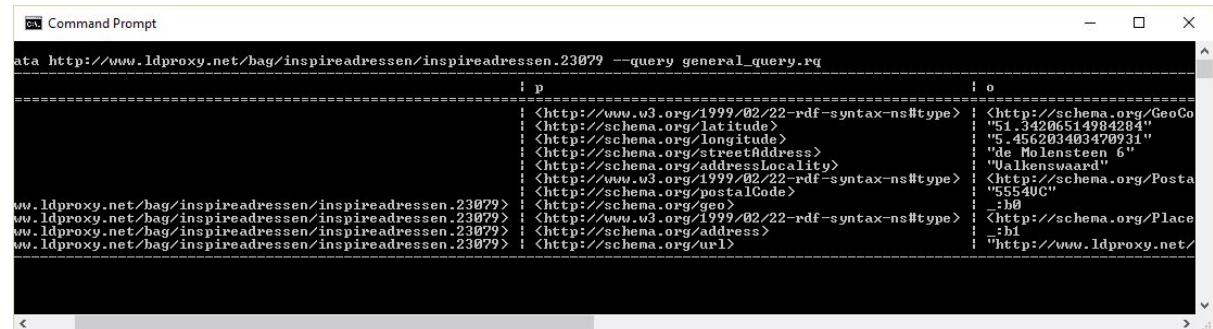
The contents of a simple query file is this:

```
SELECT * WHERE
{ ?s ?p ?o }
LIMIT 20
```

The response should be a table of 20 rows with the triples belonging to that resource: .

```
> arq --data
http://www.ldproxy.net/bag/inspireadressen/inspireadressen.23079
--query general_query.rq
```

Requirement #4b) has been met by human interference only, simply by looking at the schema definitions. There are tools²⁰ that can do this automatically but since this is not a complex schema we haven't done so.



p	o
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://schema.org/GeoCo
<http://schema.org/latitude>	"51.34286514984284"
<http://schema.org/longitude>	"5.456203403428931"
<http://schema.org/streetAddress>	"de Molensteen 6"
<http://schema.org/addressLocality>	"Walkenswaard"
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://schema.org/Posta
<http://schema.org/postalCode>	"55540C"
<http://schema.org/geo>	_b0
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://schema.org/Place
<http://schema.org/address>	_b1
<http://schema.org/url>	"http://www.ldproxy.net/

The same for the catalog proxy URLs:

```
> arq --data
http://opendatacat.net/geonetwork-geo4web/dcat/88a3c1ae-902e-4efa-
8663-ddd836101929 --query general_query.rq
```



p	o
<http://geonetwork-geo4web/dcat/88a3c1ae-902e-4efa-8663-ddd836101929#ge4web/organization/Rijksdienst20voor20het20Cultureel20Erfgoed>	<http://xmlns.com/foaf/0.1/name> "Rijksdiens
<http://geonetwork-geo4web/dcat/88a3c1ae-902e-4efa-8663-ddd836101929#ge4web/organization/Rijksdienst20voor20het20Cultureel20Erfgoed>	<http://xmlns.com/foaf/0.1/member> <http://ope
<http://geonetwork-geo4web/dcat/88a3c1ae-902e-4efa-8663-ddd836101929#ge4web/resource/f7d22d90-7651-11e0-a1f0-0800200c9a63>	<http://xmlns.com/foaf/0.1/thumbnail> <http://ser
<http://geonetwork-geo4web/resource/f7d22d90-7651-11e0-a1f0-0800200c9a63>	<http://xmlns.com/foaf/0.1/thumbnail> <http://ser
<http://geonetwork-geo4web/resource/f7d22d90-7651-11e0-a1f0-0800200c9a63>	<http://www.w3.org/ns/dcat#dataQuality> "De begrenz
<http://geonetwork-geo4web/resource/f7d22d90-7651-11e0-a1f0-0800200c9a63>	<http://www.w3.org/ns/dcat#distribution>

This query can easily be extended by including (for instance) the DBpedia endpoint.

```
PREFIX schema: <http://schema.org/>
```

²⁰ <https://github.com/AKSW/RDFUnit>

```

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dbpo: <http://dbpedia.org/ontology/>
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT ?postaladdress ?dbp ?dbppredicate ?dbpobject WHERE
{ ?postaladdress schema:addressLocality ?locality .
  BIND (strlang(?locality , "nl") as ?localitylang ) .
  SERVICE <http://nl.dbpedia.org/sparql>
    {      SELECT * WHERE { ?dbp rdfs:label ?localitylang .
      ?dbp ?dbppredicate ?dbpobject .
    }
}
}

```

The result is a combination of the BAG adressen data and DBpedia data.

			dbpObject
?postaladdress	: dbp	: dbppredicate	
-:b0	<http://nl.dbpedia.org/resource/Categorie:Valkenswaard>	: rdf:type	<http://www.w3.org/2000/01/rdf-schema#Class>
-:b0	<http://nl.dbpedia.org/resource/Valkenswaard>	: rdfs:label	'Valkenswaard'@nl
-:b0	<http://nl.dbpedia.org/resource/Categorie:Valkenswaard>	: http://www.w3.org/2002/07/owl#sameAs	<http://nl.dbpedia.org/resource/Valkenswaard>
-:b0	<http://nl.dbpedia.org/resource/Valkenswaard>	: http://www.w3.org/2002/07/owl#sameAs	<http://nl.dbpedia.org/resource/Valkenswaard>
-:b0	<http://nl.dbpedia.org/resource/Valkenswaard>	: http://www.w3.org/2002/07/owl#sameAs	<http://nl.dbpedia.org/resource/Valkenswaard>
-:b0	<http://nl.dbpedia.org/resource/Valkenswaard>	: http://www.w3.org/2002/07/owl#sameAs	<http://nl.dbpedia.org/resource/Valkenswaard>
-:b0	<http://nl.dbpedia.org/resource/Valkenswaard>	: http://www.w3.org/2002/07/owl#broader	<http://nl.dbpedia.org/resource/Brabant>
-:b0	<http://nl.dbpedia.org/resource/Valkenswaard>	: http://www.w3.org/2002/07/skos/core#prefLabel	'Valkenswaard'@nl
-:b0	<http://nl.dbpedia.org/resource/Valkenswaard>	: dbpo:wikiPageDisambiguates	<http://nl.dbpedia.org/resource/Valkenswaard>
-:b0	<http://nl.dbpedia.org/resource/Categorie:Valkenswaard>	: dbpo:wikiPageDisambiguates	<http://nl.dbpedia.org/resource/Valkenswaard>
-:b0	<http://nl.dbpedia.org/resource/Categorie:Valkenswaard>	: http://nl.dbpedia.org/property/wikiPageUsesTemplate	<http://nl.dbpedia.org/property/wikiPageUsesTemplate>
-:b0	<http://nl.dbpedia.org/resource/Valkenswaard>	: http://nl.dbpedia.org/property/wikiPageUsesTemplate	<http://nl.dbpedia.org/property/wikiPageUsesTemplate>

Observations related to schema.org and its use by search engines

This section summarizes observations from our use of the schema.org vocabulary and how search engines support the vocabulary.

Schema.org geo extension

During the mapping exercise of the datasets to schema.org we noticed that the schema.org geo classes and properties were not optimally modeled. Details about this can be found in the text below. During the testbed we have developed a straw man for an alternative geo extension for schema.org²¹. The main idea behind this extension is to allow for a simple kind of geo model at the highest level of the schema vocabulary (i.e. schema.Thing). Discussions about how to proceed with this extension are still happening.

²¹ <https://github.com/geo4web-testbed/geo-extension-to-schemaorg>

Schema.org validation and Linked Data principles

The correct use of schema.org can be checked in (amongst others) Google's structured data testing tool²². Making use of this tool learns us that schema.org does not comply to some of the Linked Data principles.

What we noticed in our research about how schema.org differs from Linked Data principles is summarised in this chapter.

Dereferenceability

The schema.org resource URLs are not dereferenceable. This can be tested by submitting an arq request to the schema.org URLs (see [above](#) for detailed instructions):



```
D:\>arq --data http://schema.org/Thing --query general_query.rq
Failed to load data
D:\>
```

Multiple Types

Having more than one type for a resource is a common mechanism in Linked Data. The Google validator does not allow a resource to have more than one type (although insiders tell us the crawler will ignore the non recognised annotations). The following JSON-LD does not recognise two types:

```
[
  {"@context" : "http://schema.org",
   "@type" : "Organization",
   "@id" : "http://example.com/1234"},

  { "@context" : "http://schema.org",
    "@type" : "Website",
    "@id" : "http://example.com/1234"}
]
```

If you remove the @id tag, and change it into a “url” tag, the following happens:

²² <https://developers.google.com/structured-data/testing-tool/>

The screenshot shows the Google JSON-LD validator interface. On the left, there is a code editor containing the following JSON-LD code:

```

1 [ 
2   { "@context" : "http://schema.org",
3     "@type" : "Organization",
4     "url" : "http://example.com/1234"
5   ,
6 
7   { "@context" : "http://schema.org",
8     "@type" : "WebSite",
9     "url" : "http://example.com/1234"
10  ]

```

On the right, under the "VALIDATE" button, the results are displayed:

- Organization (1)**
 - Organization**
 - url: <http://example.com/1234>
- WebSite (1)**
 - WebSite**
 - url: <http://example.com/1234>

Note: The “url” tag does not represent the identifier of the resource. This is done by the @id tag. In the Google validator an identifier is represented by an italic face representation, every other value is normal text, labels are bold face.

Usually the JSON-LD markup in websites do not have any @id tags. According to Linked Data principles this would mean that a blank node is created. Although this is allowed, the result is that the resource cannot be used for external linking since the identifier is not known to the outside world.

Linking principles

In the Google validator it is not allowed to use owl:sameAs to specify similarity between resources, as can be seen in the screenshot below.

The screenshot shows the Google JSON-LD validator interface with two errors indicated by a red exclamation mark icon in the top right corner.

Place (1)

Place:	The property http://www.w3.org/2002/07/owl#sameAs is not recognised by Google for an object of type Place.
description:	
Errors ! <ul style="list-style-type: none"> ! http://www.w3.org/2002/07/owl#sameAs [Place]: http://anotherexample.org/xyz/place/12345 ! http://www.w3.org/2002/07/owl#sameAs [http://dbpedia.org/ontology/Place]: http://dbpedia.org/resource/Valkenswaard 	
geo [GeoShape]:	http://example.org/aan/shape/356490
polygon:	111370.8638000004 561886.1823000014 111417.20600000024 56 1903.2932999991 111470.32140000165 561824.9866999984 1114 79.94629999995 561811.7969999984 111481.49109999835 56180 9.3016999997 111452.9728000015 561791.3588999994 111435.50 530000031 561780.4268999994 111412.69060000032 561770.920 6999987 111404.84809999913 561768.0689000003 111389.99480 000138 561764.5040999986 111384.2910999991 561764.1475999 020 111368 06910000851 561815 2130000007 111310 176000001

So linking between schema.org classes must happen by means of other properties, that are member of the schema.org vocabulary. Moreover, the classes that we link to must also be member of the schema.org vocabulary:

▼ Place (1) 1 Error !

Place:	<code>http://dbpedia.org/ontology/Place</code> is not a known valid target type for the <code>containedInPlace</code> property.
description:	
containedInP	
! containedInPlace [<code>http://dbpedia.org/ontology/Place</code>]:	<code>http://dbpedia.org/resource/Valkenswaard</code>
geo [GeoShape]:	<code>http://example.org/aan/shape/356490</code>
polygon:	<code>111370.8638000004 561886.1823000014 111417.2060000024 56190 3.2932999991 111470.32140000165 561824.9866999984 111479.946 29999995 561811.7969999984 111481.49109999835 561809.3016999 997 111452.9728000015 561791.3588999994 111435.5053000031 5 61780.4268999994 111412.6906000032 561770.9206999987 111404 .84809999913 561768.0689000003 111389.99480000138 561764.504 0999986 111384.2910999991 561764.1475999989 111368.962499998 51 561815.2430000007 111349.47500000149 561878.1020999998 11 1370.8638000004 561886.1823000014</code>

We have asked the schema.org developers for feedback on these findings ²³ and after having received this we can conclude that schema.org allows the use of other vocabularies but these are not used by the search engines.

Experiences with existing WFS deployments

This section documents common issues that we have seen in the past using OGC Web Feature Services. We focus on WFS version 2.0 as it is the version referenced from the INSPIRE technical guidelines for download services.

First of all, before publishing a WFS 2.0 service, it should be validated using the [OGC Conformance Test Suite for WFS 2.0](#) against all conformance classes that the WFS deployment is implementing.

For a WFS that also implements the INSPIRE technical guidance for pre-defined datasets or direct access (or the Dutch profiles), Geonovum provides a [validator](#).

A key issue that often blocks the use of a WFS in a client or library are **XML Schema problems**. Validation of XML is often a pre-condition and failure to conform makes the WFS unusable for many clients. This is also the case with Idproxy, which uses Java libraries that will only accept valid XML.

Some typical problems related to XML Schemas that we have seen with WFSs:

- References are to local XML Schema documents (using file-URIs) not available to clients. XML Schema documents need to be referenced using absolute HTTP URIs.

²³ <https://github.com/geo4web-testbed/general/issues/2>

- References are protected XML Schema documents, but the XML Schema documents need to be publically available.
- Incorrect copies of standardized schemas are used, e.g. for WFS 2.0, Filter 2.0 or GML 3.2.
- The XML Schema documents published by the WFS are incomplete. For example, they sometimes import namespaces without actually providing information where the imported namespace can be found. Or a broken URL is provided for the XML Schema document of the namespace.
- The XML Schema documents published by the WFS are invalid themselves.
- XML in the WFS responses is invalid as a namespace is not declared or imported.

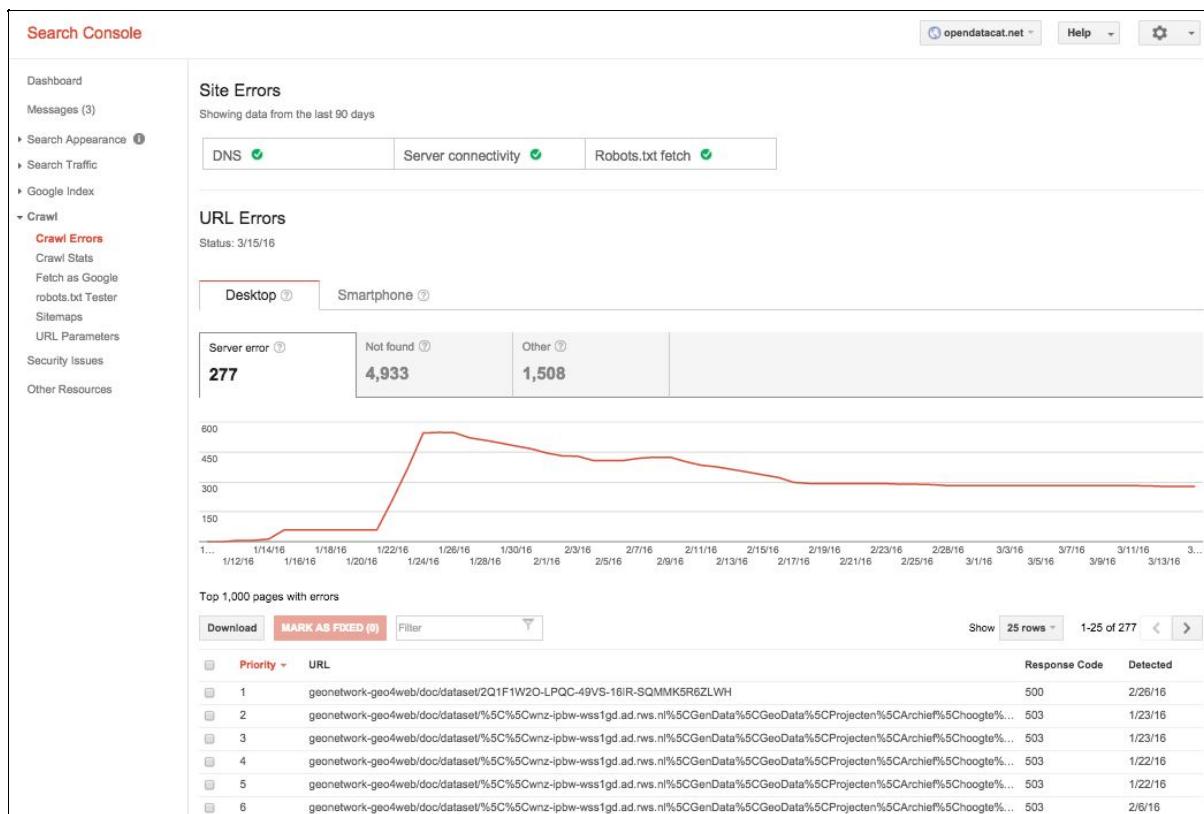
Other typical problems are:

- Inconsistent information between the WFS capabilities and information used in the feature encodings. For example, the URI of the coordinate reference system(s) for a feature type is sometimes not consistent with the URIs used in the GML encoding.
- Links to other resources, encoded in xlink:href attributes, do not resolve due to errors how the URIs are compiled.

In addition to those issues, some WFS seem to have only been tested to a limited extent and requests that are not just a basic request may lead to errors and exceptions. Using the validators listed above should help to identify many of such issues.

Experiences with existing metadata

As stated at various places in this document we have identified multiple broken links between metadata and services and asynchronous metadata statements between capabilities en catalog content. But there were other observations related to metadata content that surfaced once the search engine crawlers started crawling the content and reported multiple errors in the crawled web pages. These errors are typically related to bad input in the metadata. Having this type of errors in crawled content is not optimal for the search ranking. It would be good to point metadata providers to these issues and or improve the software so it can fix common mistakes in metadata while transforming to schema.org.



A listing of common mistakes:

Links in metadata not containing a http:// prefix.

The crawler will then try to resolve the link as local

<http://.../geonetwork/www.rijkswaterstaat.nl>.

Illegal modification date

For example a modification date of '1969-01-01' will cause crawling of the sitemap to fail, search engines use this field to determine if the resource should be revisited.

Missing thumbnails

A link to a thumbnail which does not exist, causes a broken link error.

Use of illegal characters in a url

Some data providers use a uuid enclosed in {...}. These characters are used in a url in a non-encoded way by GeoNetwork. This causes issues in some http clients.

Open challenges

This section summarizes challenges that we have been facing in the testbed and that cannot be addressed or resolved as part of our work.

Challenges related to the existing infrastructure

Support for establishing links dynamically

Existing datasets are sometimes structured in ways that make it difficult to dynamically establish links between resources (features) in the datasets.

For example, the Bekendmakingen dataset contains an [entry that a human will understand](#) [links to the addresses Dragonder 48 and Dragonder 50 in Valkenswaard](#). However, the information is part of text attributes and there is no reliable query that would allow to dynamically establish the links, for example, in the WFS proxy. The result is that we can link to the public announcement only from the Valkenswaard or the postcode resource, but not from the more specific resources of the two addresses.

In fact, the schema used by the Bekendmakingen dataset supports structured address information, but these fields are always left empty and street name and house number information is only provided as part of textual descriptions.

For setting up the data infrastructure supporting the new environmental law, it seems important that data that needs to be linked supports dynamic link establishment, where feasible. This may require changes to schemas of the data sets and / or the data capturing processes. In the example of the Bekendmakingen dataset, it would be helpful, if the announcements would be captured with structured address information in case an announcement is related to an address.

Metadata today is often incomplete and inconsistent

We found that the metadata provided in the WFS capabilities and the dataset metadata in the catalog is either incomplete or both have inconsistent information.

Just a few examples:

Metadata element	Catalog	WFS
Bounding box (West)	3.37087°	3.053°
Bounding box (East)	7.21097°	7.24°
Bounding box (South)	50.7539°	47.975°
Bounding box (North)	53.4658°	53.504°
Contact, individual name	Beheer PDOK	KlantContactCenter PDOK
Contact, organisation name	Beheer PDOK	-

Contact, phone	-	-
Contact, email	beheerPDOK@kadaster.nl	-
Contact, address	-	Apeldoorn Nederland
Contact, role	pointOfContact	pointOfContact

As spatial data infrastructures separate between metadata and data, often the metadata will be collected and maintained through separate processes. This has to be taken into account when planning how to support the new environmental law use cases.

Performance & data compactness

The response time and size of resources on the Web may have an impact on the indexing and ranking by search engines as well as on the usability in general. This has to be taken into account when publishing spatial data on the Web. Where proxies on top of the existing infrastructure are used, the capabilities of the existing SDI components have to be analysed, if they meet expectations.

The report of research topic 3 (“crawlable spatial data”) has analysed this issue in more detail. See their report for additional information.

Challenges related to the Web

Search engines are a black box

As described [in more detail earlier in this report](#), search engines are largely a “black box”. It is not clear - at least to us, if and how structured spatial data will be used and how users can use it to improve finding data they are interested in. Therefore, it is not entirely clear, if there are better ways (URI strategies, data formats, use of vocabularies, use of RDFa vs microformats vs JSON-LD, etc.) to make spatial data ‘crawlable’.

Again, the report of research topic 3 (“crawlable spatial data”) has additional information.

Content negotiation based on media types insufficient

The HTTP content negotiation mechanism has its limits as it supports only selection based on the “data format”, but is unable to support selection of different schemas (e.g. schema.org for the search engines vs DCAT for governmental data portals).

The approach that we have taken, and which mostly solves the issue for us, is to simply provide separate URIs for the different schemas / user communities.

Links

In this section we summarize links to all resources that are results of our work on research topic 4, “Spatial Data on the Web using the current SDI”. These resources will exist up to a half year after the ending of the research, but may be subject to change afterwards.

Deployments

ldproxy for BAG WFS	http://www.ldproxy.net/bag/
ldproxy for Farmland WFS	http://www.ldproxy.net/aan/
GeoNetwork for NSDI CSW	http://opendatacat.net/
BaseURI for Bekendmakingen dataset	http://data.linkeddatafactory.nl/bekendmakingen/
BaseURI for Bekendmakingen Spatial dataset	http://data.linkeddatafactory.nl/bekendmakingenspatial/
SPARQL endpoint	http://data.linkeddatafactory.nl:8890/sparql use either “ http://data.linkeddatafactory.nl/bekendmakingenspatial/ ” or “ http://data.linkeddatafactory.nl/bekendmakingen ” as the “Default Data Set Name (Graph IRI)”

Source code repositories

ldproxy	https://github.com/interactive-instruments/ldproxy https://github.com/interactive-instruments/XtraPlatform-SDI-Tools https://github.com/interactive-instruments/XtraPlatform
GeoNetwork	https://github.com/geo4web-testbed/core-geonetwork https://github.com/geonetwork/core-geonetwork https://github.com/metadata101/iso19139.nl.geografie.1.3.1

Discussions and other resources

Geonovum website	http://www.geonovum.nl/onderwerp-artikel/testbed-locatie-data-het-web
Our proposal	https://github.com/geo4web-testbed/topic4-general/blob/master/Proposal_topic_4_ii_GeoCat_LDF.pdf

General discussions and presentations	https://github.com/geo4web-testbed/topic4-general
schema.org mapping (BAG features)	https://github.com/geo4web-testbed/simpleBAG-to-schemaorg
schema.org mapping (farmland features)	https://github.com/geo4web-testbed/aan_to_schema_org
Identification and processing of the Bekendmakingen dataset	https://github.com/geo4web-testbed/topic4-task2
schema.org geo-extension	https://github.com/geo4web-testbed/geo-extension-to-schemaorg
ISO 19139 to GeoDCAT-AP mapping	https://github.com/geo4web-testbed/iso-19139-to-dcat-ap
ldproxy design	https://github.com/geo4web-testbed/ldproxy-design
ldproxy development	https://github.com/interactive-instruments/ldproxy/issues
Research topic 2 report: Spatial Data Platform	To be added when available
Research topic 3 report: Crawlable Spatial Data	To be added when available

Appendix A

XML Schema: Farmland data

```
<?xml version="1.0" encoding="utf-8"?><xsd:schema
xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:aan="http://aan.geonovum.nl"
xmlns:gml="http://www.opengis.net/gml/3.2" xmlns:wfs="http://www.opengis.net/wfs/2.0"
elementFormDefault="qualified" targetNamespace="http://aan.geonovum.nl">
  <xsd:import namespace="http://www.opengis.net/gml/3.2"
schemaLocation="http://geodata.nationaalgeoregister.nl/schemas/gml/3.2.1/gml.xsd"/>
  <xsd:complexType name="aanType">
    <xsd:complexContent>
      <xsd:extension base="gml:AbstractFeatureType">
        <xsd:sequence>
          <xsd:element maxOccurs="1" minOccurs="0" name="objectid" nillable="true"
type="xsd:int"/>
          <xsd:element maxOccurs="1" minOccurs="0" name="aanid" nillable="true"
type="xsd:int"/>
          <xsd:element maxOccurs="1" minOccurs="0" name="versiebron" nillable="true"
type="xsd:string"/>
          <xsd:element maxOccurs="1" minOccurs="0" name="type" nillable="true"
type="xsd:string"/>
          <xsd:element maxOccurs="1" minOccurs="0" name="geom" nillable="true"
type="gml:MultiSurfacePropertyType"/>
        </xsd:sequence>
      </xsd:extension>
    </xsd:complexContent>
  </xsd:complexType>
  <xsd:element name="aan" substitutionGroup="gml:AbstractFeature" type="aan:aanType"/>
</xsd:schema>
```

XML Schema: Simple BAG data

```
<xsd:complexType name="inspireadresstenType">
  <xsd:complexContent>
    <xsd:extension base="gml:AbstractFeatureType">
      <xsd:sequence>
        <xsd:element maxOccurs="1" minOccurs="0" name="straatnaam" nillable="true"
type="xsd:string"/>
        <xsd:element maxOccurs="1" minOccurs="0" name="huisnummer" nillable="true" type="xsd:int"/>
        <xsd:element maxOccurs="1" minOccurs="0" name="huisletter" nillable="true"
type="xsd:string"/>
        <xsd:element maxOccurs="1" minOccurs="0" name="toevoeging" nillable="true"
type="xsd:string"/>
        <xsd:element maxOccurs="1" minOccurs="0" name="woonplaats" nillable="true"
type="xsd:string"/>
        <xsd:element maxOccurs="1" minOccurs="0" name="postcode" nillable="true"
type="xsd:string"/>
        <xsd:element maxOccurs="1" minOccurs="0" name="geom" nillable="true"
type="gml:PointPropertyType"/>
      </xsd:sequence>
    </xsd:extension>
  </xsd:complexContent>
</xsd:complexType>
<xsd:element name="inspireadressen" substitutionGroup="gml:AbstractFeature"
type="inspireadressen:inspireadresstenType"/>
```

Appendix B - Crosswalk: Spatial Data on the Web best practices in the WFS proxy

A joint working group of W3C and OGC on “Spatial Data on the Web” is working on a best practices document. The first stable version is a First Public Working Draft (<https://www.w3.org/TR/2016/WD-sdw-bp-20160119/>), the latest draft version is the Editor’s Draft (<http://w3c.github.io/sdw/bp/>).

This annexes uses the Editor’s Draft dated 02/03/2016 and lists for each best practice, if and how the WFS proxy implementation follows the best practice.

Best Practice 1: Use globally unique HTTP identifiers for entity-level resources - *entities within datasets should have unique, persistent HTTP or HTTPS URLs as identifiers*

Best Practice 3: Convert or map dataset-scoped identifiers to URIs - *find (or create) URIs for resources in spatial datasets that have locally-scoped identifiers*

Implemented. For example,

<http://www.ldproxy.net/bag/inspireadressen/inspireadressen.3456789>. This URI is stable and persistent. It will only change, if the address is no longer part of the dataset (assuming that the WFS implementation conforms to the rules and does change the local `gml:id` identifiers of features).

Best Practice 2: Reuse existing (authoritative) identifiers when available - *avoiding the creation of lots of identifiers for the same resource*

Not applicable. There are no (authoritative) URIs that we are aware of for the features in the datasets.

Best Practice 4: Provide stable identifiers for Things (resources) that change over time - *even though resources change, it helps when they have a stable, unchanging identifier*

Not implementable. We do not know how the data providers handle changes. It is, for example, entirely possible that the data provider of the Farmland dataset would consider a 2017 release a new dataset, with new identifiers.

Best Practice 5: Provide identifiers for parts of larger information resources - *identify subsets of large information resources that are a convenient size for applications to work with*

Implemented. For the BAG dataset, there are [subsets](#) for each woonplaats (municipality) and post code.

Best Practice 6: Provide a minimum set of information for your intended application - *when someone looks up a URI for a SpatialThing, provide useful information, using the common representation formats*

Implemented. At least to the extent possible. All information in the dataset is provided and it is possible to configure the proxy to map information to well-known properties, e.g. in the schema.org vocabulary.

Best Practice 7: How to describe geometry - *geometry data should be expressed in a way that allows its publication and use on the Web*

Implemented. We use schema.org GeoCoordinates and GeoShape in [HTML annotations](#) and [JSON-LD](#), since we are targetting search engine crawlers. In addition, we provide GeoJSON and GML representations of the features for clients that prefer these representations. Content negotiation is supported.

Best Practice 8: Specify Coordinate Reference System for high-precision applications - *coordinate referencing system should be specified for high-precision applications to locate geospatial entities*

HTML/schema.org: <http://www.ldproxy.net/aan/aan/28>

JSON-LD/schema.org: <http://www.ldproxy.net/aan/aan/28/?f=jsonld>

GeoJSON: <http://www.ldproxy.net/aan/aan/28/?f=json>

GML: <http://www.ldproxy.net/aan/aan/28/?f=xml>

Implemented. As schema.org and GeoJSON requires WGS84 this is what is used for these representations. The GML representation however uses the Dutch CRS used in the source data.

JSON-LD/schema.org:

<http://www.ldproxy.net/bag/inspireadressen/inspireadressen.3456789/?f=jsonld>

GeoJSON: <http://www.ldproxy.net/bag/inspireadressen/inspireadressen.3456789/?f=json>

GML: <http://www.ldproxy.net/bag/inspireadressen/inspireadressen.3456789/?f=xml>

Best Practice 9: How to describe relative positions - *provide a relative positioning capability in which the entities can be linked to a specific position*

Not applicable. This might be relevant for other data linking to the BAG data, but it is not relevant for the two datasets accessible via the proxy.

Best Practice 10: How to describe positional (in)accuracy - *accuracy and precision of spatial data should be specified in machine-interpretable and human-readable form*

Not implementable. We have no information about the accuracy. In principle, if the dataset metadata linked from the WFS capabilities would provide this information in a consistent way, this could be implemented, too. In this case, the dataset metadata did not provide this information.

Best Practice 11: How to describe properties that change over time - *entities and their data should have versioning with time/location references*

Not applicable. The owners of the dataset did not include any such information in the data. If it would be included, this information would also be exposed via the proxy - in the same way it is included in the source dataset.

Best Practice 12: Use spatial semantics for Spatial Things - *the best vocabulary should be chosen to describe the available SpatialThings*

Implemented. We use schema.org as the key use case is indexing by search engine crawlers.

Best Practice 13: Assert known relationships - *spatial relationships between Things should be specified in forms of geographical, topological and hierarchical links*

Partly implemented. Relationships that could be identified dynamically [between address data and public announcements](#) are established and included in the representations, where possible. However, if the data would include [street address information in a structured, reliable way](#), links on the address level could be included, too.

Best Practice 14: Provide context required to interpret observation data values

Best Practice 15: Describe sensor data processing workflows

Best Practice 16: Relate observation data to the real world

Best Practice 18: How to publish (and consume) sensor data streams

Not applicable. There is no observation data involved.

That said, the Farmland data seems to have information about the year in which the data was captured, but there is no information where more information about that data capturing process could be found.

Best Practice 17: How to work with crowd-sourced observations

Not applicable. There is no crowd-sourced data involved.

Best Practice 19: Make your entity-level links visible on the web - *the data should be published with explicit links to other resources*

Best Practice 21: Link to spatial Things - *create durable links by connecting Spatial Things*

Best Practice 23: Link to related resources - *link your spatial resources to other related resources*

Implemented. Through [content negotiation and multiple formats](#) (HTML, JSON-LD, GeoJSON, GML) both human and machines are supported. Links have been added [between datasets](#) and [to the catalogue](#).

See also Best Practice 13.

Best Practice 20: Provide meaningful links - *when providing a link, a data publisher should opt for a level of formal and meaningful semantics that helps data consumers to decide if the target resource is relevant to them*

Implemented. We use schema.org (or other vocabularies) to express the semantics.

Best Practice 22: Link to resources with well-known or authoritative identifiers - *link your spatial resources to others that are commonly used*

Not implemented. Due to time/resource constraints we have not addressed this, as this was not the focus of the research.

Best Practice 24: Use links to find related data - *related data to a spatial dataset and its individual data items should be discoverable by browsing the links*

Partly implemented. This has been implemented only to a limited extent with the links to the catalogue information and Bekendmakingen features.

Best Practice 25: Make your entity-level data indexable by search engines - *search engines should receive a metadata response to a HTTP GET when dereferencing the link target URI*

Implemented. See <https://www.google.de/search?q=site:ldproxy.net>.

Note that from the [experience so far](#), it is unclear in how far the structured data included in the annotations has any effect.

Best Practice 26: Include spatial information in dataset metadata - *the description of datasets that have spatial features should include explicit metadata about the spatial information*

Implemented. While this is mainly addressed via links to the [CSW proxy](#), the WFS Capabilities include bounding box information and this is included in the schema.org Datasets, i.e. is provided on the WFS landing page (<http://www.ldproxy.net/aan/>) and for each feature type (<http://www.ldproxy.net/aan/aan/>).

Best Practice 27: Publish data at the granularity you can support - *granularity of mechanisms provided to access access a dataset should be decided based on available resources*

Implemented. Since WFS provides feature data and a proxy can support that granularity, the proxy provides feature-level granularity. In addition, [named subsets](#) have been introduced that are not directly available via the WFS:

Best Practice 28: Expose entity-level data through 'convenience APIs' - *if you have a specific application in mind for publishing your data, tailor the spatial data API to meet that goal*

Best Practice 30: Include search capability in your data access API - *if you publish an API to access your data, make sure it allows users to search for specific data*

Implemented. Since the main use case is supporting search engine crawlers, the API is quite restricted. However, it provides [basic selection capabilities](#) in addition to links to each feature.

Best Practice 29: APIs should be self-describing - *APIs should provide a discoverable description of their contents and how to interact with it; ideally this should be a machine-readable description*

Considered in the design. The idea is to automatically generate Swagger API descriptions and provide those. We will not address this in the testbed due to time and resource constraints. There are also open questions (Swagger uses something similar to JSON Schema, but we want to support JSON-LD). See <https://github.com/interactive-instruments/ldproxy/issues/9>.

Appendix C: Example of an ISO 19139 document transformed to GeoDCAT-AP

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <dcat:CatalogRecord xmlns:dcat="http://www.w3.org/ns/dcat#">
    rdf:about="http://opendatacat.net/geonetwork-geo4web/metadata/88a3c1ae-902e-4
    efa-8663-ddd836101929">
      <foaf:primaryTopic xmlns:foaf="http://xmlns.com/foaf/0.1/">
        rdf:resource="http://opendatacat.net/geonetwork-geo4web/resource/f7d22d90-765
        1-11e0-a1f0-0800200c9a63"/>
        <dct:issued xmlns:dct="http://purl.org/dc/terms/" />
        <dct:modified xmlns:dct="http://purl.org/dc/terms/" />
        <dct:references xmlns:dct="http://purl.org/dc/terms/">
          <rdf:Description
            rdf:about="http://opendatacat.net/geonetwork-geo4web/srv/eng/xml.metadata.get
            ?uuid=88a3c1ae-902e-4efa-8663-ddd836101929">
            <dct:format>
              <dct:IMT>
                <rdf:value>application/xml</rdf:value>
                <rdfs:label
                  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">XML</rdfs:label>
              </dct:IMT>
            </dct:format>
          </rdf:Description>
        </dct:references>
        <dct:references xmlns:dct="http://purl.org/dc/terms/">
          <rdf:Description
            rdf:about="http://opendatacat.net/geonetwork-geo4web?uuid=88a3c1ae-902e-4efa-
            8663-ddd836101929">
            <dct:format>
              <dct:IMT>
                <rdf:value>text/html</rdf:value>
                <rdfs:label
                  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">HTML</rdfs:label>
              </dct:IMT>
            </dct:format>
          </rdf:Description>
        </dct:references>
      </dcat:CatalogRecord>
      <dcat:Dataset xmlns:dcat="http://www.w3.org/ns/dcat#">
        rdf:about="http://opendatacat.net/geonetwork-geo4web/resource/f7d22d90-7651-1
        1e0-a1f0-0800200c9a63">
          <dct:identifier
            xmlns:dct="http://purl.org/dc/terms/">f7d22d90-7651-11e0-a1f0-0800200c9a63</d
            ct:identifier>
          <dct:title
            xmlns:dct="http://purl.org/dc/terms/">Werelderfgoed</dct:title>
```

<dct:abstract xmlns:dct="http://purl.org/dc/terms/">Begrenzingen van het Nederlandse cultureel erfgoed dat vanwege haar uitzonderlijke universele waarde door het Werelderfgoed comite van UNESCO is geplaatst op de Werelderfgoedlijst. De begrenzingen zijn gebaseerd op de kaarten die destijds opgenomen waren in het nominatiedossier dat aan Unesco is aangeboden.</dct:abstract>

<dcat:keyword>werelderfgoed</dcat:keyword>

<dcat:keyword>UNESCO</dcat:keyword>

<dcat:theme

rdf:resource="http://opendatacat.net/geonetwork-geo4web/thesaurus/GEMET%20-%20INSPIRE%20themes%20version%201.0/Beschermde gebieden"/>

<dcat:theme

rdf:resource="http://opendatacat.net/geonetwork-geo4web/thesaurus/iso/topicCategory/structure"/>

<foaf:thumbnail xmlns:foaf="http://xmlns.com/foaf/0.1/"

rdf:resource="http://services.rce.geovoortziening.nl/www/PreviewWerelderfgoedKlein200x150.png"/>

<foaf:thumbnail xmlns:foaf="http://xmlns.com/foaf/0.1/"

rdf:resource="http://services.rce.geovoortziening.nl/www/PreviewWerelderfgoedGroot800x600.png"/>

<dct:spatial xmlns:dct="http://purl.org/dc/terms/">

<ogc:Polygon xmlns:ogc="http://www.opengis.net/rdf#">

<ogc:asWKT rdf:datatype="http://www.opengis.net/rdf#WKTLiteral">

<http://www.opengis.net/def/crs/OGC/1.3/CRS84>

Polygon((3.4 50.8, 3.4 53.6, 7.2 53.6, 7.2 50.8, 3.4 50.8))

</ogc:asWKT>

</ogc:Polygon>

</dct:spatial>

<dct:temporal xmlns:dct="http://purl.org/dc/terms/">

<dct:publisher xmlns:dct="http://purl.org/dc/terms/">

rdf:resource="http://opendatacat.net/geonetwork-geo4web/organization/Rijksdienst%20voor%20het%20Cultureel%20Erfgoed"/>

<dct:accrualPeriodicity

xmlns:dct="http://purl.org/dc/terms/">notPlanned</dct:accrualPeriodicity>

<dcat:granularity>10000</dcat:granularity>

<dct:language xmlns:dct="http://purl.org/dc/terms/">dut</dct:language>

<dct:license

xmlns:dct="http://purl.org/dc/terms/">otherRestrictions</dct:license>

<dct:license xmlns:dct="http://purl.org/dc/terms/">Geo Gedeeld licentie</dct:license>

<dct:license

xmlns:dct="http://purl.org/dc/terms/">http://services.rce.geovoortziening.nl/www/GeoGedeeld_WereldErfgoed_2011101.pdf</dct:license>

<dcat:distribution

rdf:resource="http://services.rce.geovoortziening.nl/rce/wms"/>

<dcat:distribution

rdf:resource="http://services.rce.geovoortziening.nl/rce/wfs"/>

<dcat:distribution

rdf:resource="http://services.rce.geovoortziening.nl/www/download/data/Unesco_Werelderfgoed_nl.xml"/>

```

<dcat:dataQuality>De begrenzingen zijn door de Rijksdienst voor het
Cultureel Erfgoed gedigitaliseerd op basis van TOP10vector, GBKN en
topografische kaarten 1:25.000.</dcat:dataQuality>
</dcat:Dataset>
<skos:Concept xmlns:skos="http://www.w3.org/2004/02/skos/core#"
rdf:about="http://opendatacat.net/geonetwork-geo4web/thesaurus/GEMET%20-%20IN
SPIRE%20themes%2C%20version%201.0/Beschermde%20gebieden">
    <skos:inScheme
rdf:resource="http://opendatacat.net/geonetwork-geo4web/thesaurus/GEMET%20-%2
0INSPIRE%20themes%2C%20version%201.0"/>
    <skos:prefLabel>Beschermde gebieden</skos:prefLabel>
</skos:Concept>
<dcat:Distribution xmlns:dcat="http://www.w3.org/ns/dcat#"

rdf:about="http://services.rce.geovoorziening.nl/rce/wms">

<dcat:accessURL>http://services.rce.geovoorziening.nl/rce/wms</dcat:accessURL
>
    <dct:title
xmlns:dct="http://purl.org/dc/terms/">WorldHeritage</dct:title>
    <dct:format xmlns:dct="http://purl.org/dc/terms/">
        <dct:IMT>
            <rdf:value>OGC:WMS</rdf:value>
            <rdfs:label
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">OGC:WMS</rdfs:label>
        </dct:IMT>
    </dct:format>
</dcat:Distribution>
<dcat:Distribution xmlns:dcat="http://www.w3.org/ns/dcat#"

rdf:about="http://services.rce.geovoorziening.nl/rce/wfs">

<dcat:accessURL>http://services.rce.geovoorziening.nl/rce/wfs</dcat:accessURL
>
    <dct:title
xmlns:dct="http://purl.org/dc/terms/">WorldHeritage</dct:title>
    <dct:format xmlns:dct="http://purl.org/dc/terms/">
        <dct:IMT>
            <rdf:value>OGC:WFS</rdf:value>
            <rdfs:label
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">OGC:WFS</rdfs:label>
        </dct:IMT>
    </dct:format>
</dcat:Distribution>
<dcat:Distribution xmlns:dcat="http://www.w3.org/ns/dcat#"
rdf:about="http://services.rce.geovoorziening.nl/www/download/data/Unesco_Wer
elderfgoed_nl.xml">
<dcat:accessURL>http://services.rce.geovoorziening.nl/www/download/data/Unesc
o_Werelderfgoed_nl.xml</dcat:accessURL>
    <dct:title xmlns:dct="http://purl.org/dc/terms/">Atom feed
WorldHeritage</dct:title>

```

```
<dct:format xmlns:dct="http://purl.org/dc/terms/">
  <dct:IMT>
    <rdf:value>INSPIRE Atom</rdf:value>
    <rdfs:label
      xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">INSPIRE Atom</rdfs:label>
  </dct:IMT>
</dct:format>
</dcat:Distribution>
<foaf:Organization xmlns:foaf="http://xmlns.com/foaf/0.1/"
  rdf:about="http://opendatacat.net/geonetwork-geo4web/organization/Rijksdienst
  %20voor%20het%20Cultureel%20Erfgoed">
  <foaf:name>Rijksdienst voor het Cultureel Erfgoed</foaf:name>
  <foaf:member
    rdf:resource="http://opendatacat.net/geonetwork-geo4web/organization/info%40c
    ultureelerfgoed.nl"/>
  </foaf:Organization>
  <foaf:Agent xmlns:foaf="http://xmlns.com/foaf/0.1/"
    rdf:about="http://opendatacat.net/geonetwork-geo4web/person/info%40cultureele
    rfgood.nl">
    <foaf:name>InfoDesk</foaf:name>
    <foaf:phone>033 4217456</foaf:phone>
    <foaf:mbox rdf:resource="mailto:info@cultureelerfgoed.nl"/>
  </foaf:Agent>
</rdf:RDF>
```