
CKANEXT-DCATDONL Documentation

Kennis- en Exploitatiecentrum Officiële Overheidspublicaties

Oct 29, 2018

Table of Contents

1	Summary	3
1.1	Contact	3
1.2	References	4
2	Announcements	5
2.1	Scheduled update of public test application on https://dcat-ap-donl.nl	5
3	Changelog	7
3.1	Changes applied on 29/10/2018	7
4	Installation	9
4.1	Requirements	9
4.2	Installing the ckanext-dcatdonl plugin	10
4.3	Installation of Solr	10
4.4	Setting up the background process	11
5	Usage	13
5.1	Creating or updating datasets and resources	13
5.2	Viewing a dataset or resource	17
5.3	Searching for datasets or resources	17
6	CKAN Schema	19
6.1	DCAT to CKAN mapping	19
6.2	Dataset	21
6.3	Resource	24
6.4	Schema validation	24
6.5	Valuelists	27

The CKAN extension that implements the DCAT-AP-DONL metadata standard into CKAN.

This documentation describes the `ckanext-dcatdonl` plugin developed by Textinfo B.V. on behalf of KOOP. This plugin implements the DCAT-AP-DONL 1.1 metadata standard into CKAN. DCAT-AP-DONL 1.1 is a DCAT application profile based on the profile of DCAT-AP-NL 1.1, which itself is based on the application profile of DCAT-AP-EU 1.1. It aims to reduce data duplications and to standardize values wherever possible.

Note: It is important to note that while this extension updates the schema of CKAN, it does not provide front-end templates for these extended schemas. Users of this extension with the desire for such templates need to implement those templates themselves in a separate plugin.

A modified version of this extension is currently used in the CKAN environment of the [Data.Overheid.nl](https://data.overheid.nl)¹ webapplication. The most recent version of the public extension can be found on <https://gitlab.textinfo.nl/opensource/ckanext-dcatdonl>.

The following subjects are described

1. how to install the plugin into a existing CKAN installation
2. how to use the functionality the plugin introduces once it is installed
3. the modifications to the CKAN schema that the plugin is responsible for
4. the validation of this schema as defined by DCAT-AP-DONL 1.1
5. description of all the valuelists and their locations
6. the structure of the plugin code itself

1.1 Contact

For questions and/or comments regarding this CKAN extension please contact KOOP(Kennis- en Exploitatiecentrum Officiële Overheidspublicaties) via:

¹ <https://data.overheid.nl>

Online [koopoverheid.nl](https://www.koopoverheid.nl)²

Email opendata@overheid.nl

Telephone (070) 7000 526

1.2 References

- [Data.Overheid.nl](https://data.overheid.nl)³
- [DCAT-AP-NL 1.1](http://dcat-ap-nl.nl)⁴
- [DCAT-AP-DONL 1.1](https://dcat-ap-donl.readthedocs.io)⁵
- [CKANEXT-DCATDONL documentation](https://ckanext-dcatdonl.readthedocs.io)⁶
- [CKANEXT-DCATDONL repository](https://gitlab.textinfo.nl/opensource/ckanext-dcatdonl/)⁷

² <https://www.koopoverheid.nl/>

³ <https://data.overheid.nl>

⁴ <http://dcat-ap-nl.nl>

⁵ <https://dcat-ap-donl.readthedocs.io>

⁶ <https://ckanext-dcatdonl.readthedocs.io>

⁷ <https://gitlab.textinfo.nl/opensource/ckanext-dcatdonl/>

CHAPTER 2

Announcements

Below you'll find important announcements regarding this CKAN extension.

2.1 Scheduled update of public test application on <https://dcat-ap-donl.nl>

On 05/11/2018 the dcat-ap-donl.nl test environment will be unavailable while we roll out the latest changes of the CKAN extension.

Contains the functional changelog of this CKAN extension.

3.1 Changes applied on 29/10/2018

A new version has been released, the changes are list below!

- **updated installation instructions.**
 - Now includes the installation of additional requirements
 - Added instructions for targeting specific Apache Solr versions
- **Introduced instructions to setup Apache Solr in various versions**
 - Currently contains support for Apache Solr 5.4 and 7.4
- Introduced several backwards compatible fixes for CKAN versions below 2.6
- **Included Solr optimizations, searches against Solr now include several facets by default, namely**
 - facet_referentie_data
 - facet_access_rights
 - facet_publisher
 - facet_authority
 - facet_high_value
 - facet_basis_register
 - facet_dataset_status
 - facet_metadata_language
 - facet_frequency
 - facet_license_id

- facet_source_catalog
- facet_theme
- **Changes to the schemas:**
 - To declare a license for a package and/or resource you must now provide it in the `license_id` key rather than the *license* key.
 - The fields `highvalue`, `basisregister` and `referentiedata` are no longer considered data.overheid system properties and are now part of the base DCAT-AP-DONL scheme
 - The field `highvalue` has been renamed to `high_value`
 - The field `referentiedata` has been renamed to `referentie_data`
 - The field `basisregister` has been renamed to `basis_register`
 - The field `dataset_status` will now default to the URI for `beschikbaar`
 - The field `high_value` will now default to `false`
 - The field `referentie_data` will now default to `false`
 - The field `basis_register` will now default to `false`
 - The list validation is now less strict, when a single value is provided it will silently convert this to a list of size 1 rather than returning a validation error message
- Updated the Usage chapter to incorporate the changes to the schemas
- The documented error messages have been updated
- Updated the logging format of the `controlled_vocabulary_updater.py`
- Small fixes to various chapters of this documentation containing inaccuracies

This chapter covers all the information required to install the `ckanext-dcatdonl` plugin into a CKAN installation.

4.1 Requirements

The plugin was developed with the following versions in mind.

4.1.1 CKAN

The plugin functions correctly with these CKAN versions:

Version	Reference
2.7.3	http://docs.ckan.org/en/ckan-2.7.3/
2.7.4	http://docs.ckan.org/en/2.7/
2.8.0	http://docs.ckan.org/en/2.8/

It is likely that the plugin functions correctly in earlier and later versions, however only the above mentioned CKAN versions have been tested and confirmed to work.

4.1.2 PostgreSQL

CKAN uses PostgreSQL with version 9.2 or higher.

4.1.3 Python

CKAN itself, and the `ckanext-dcatdonl` plugin are written in Python 2.7.x. As such, the CKAN host must have this version of Python installed.

4.2 Installing the ckanext-dcatdonl plugin

Follow the steps listed below to install and activate the ckanext-dcatdonl extension into CKAN.

1. With your CKAN virtual environment activated:

```
. /usr/lib/ckan/default/bin/activate
pip install -e git+https://gitlab.textinfo.nl/opensource/ckanext-dcatdonl.git
↳#egg=ckanext-dcatdonl
cd ckanext-dcatdonl
pip install -r requirements.txt
```

2. Edit your CKAN .ini configuration file and add the following

```
ckan.plugins = ... dcatdonl
```

3. In the same file, add (or change) the following properties to:

```
licenses_group_url = file:///usr/lib/ckan/default/src/ckanext-dcatdonl/ckanext/
↳dcatdonl/resources/overheid_license.json
solr_url = http://{host}:8983/solr/ckan
ckan.mimetype_guess = None
```

4. Restart apache2

You have now successfully installed the *ckanext-dcatdonl* plugin

4.3 Installation of Solr

To install Solr version 7.5.0 (assuming no previous Solr installation):

```
sudo apt-get install openjdk-9-jre-headless
cd /opt
sudo wget http://www-eu.apache.org/dist/lucene/solr/7.5.0/solr-7.5.0.tgz
sudo tar xzf solr-7.5.0.tgz solr-7.5.0/bin/install_solr_service.sh --strip-
↳components=2
sudo bash ./install_solr_service.sh solr-7.5.0.tgz
```

To create the CKAN core into Solr:

```
sudo -u solr /opt/solr/bin/solr create -c ckan
sudo rm /var/solr/data/ckan/conf/protowords.txt
sudo rm /var/solr/data/ckan/conf/solrconfig.xml
sudo rm /var/solr/data/ckan/conf/managed-schema
sudo rm /var/solr/data/ckan/conf/stopwords.txt
sudo rm /var/solr/data/ckan/conf/synonyms.txt
sudo mkdir /var/lib/solr
sudo chown solr /var/lib/solr -R
cd ~
sudo ln -s /usr/lib/ckan/default/src/ckanext-dcatdonl/ckanext/dcatdonl/resources/solr/
↳7.4/currency.xml /var/solr/data/ckan/conf
sudo ln -s /usr/lib/ckan/default/src/ckanext-dcatdonl/ckanext/dcatdonl/resources/solr/
↳7.4/elevate.xml /var/solr/data/ckan/conf
sudo ln -s /usr/lib/ckan/default/src/ckanext-dcatdonl/ckanext/dcatdonl/resources/solr/
↳7.4/protowords.txt /var/solr/data/ckan/conf
sudo ln -s /usr/lib/ckan/default/src/ckanext-dcatdonl/ckanext/dcatdonl/resources/solr/
↳7.4/schema.xml /var/solr/data/ckan/conf
```

(continues on next page)

(continued from previous page)

```
sudo ln -s /usr/lib/ckan/default/src/ckanext-dcatdonl/ckanext/dcatdonl/resources/solr/  
↳7.4/solrconfig.xml /var/solr/data/ckan/conf  
sudo ln -s /usr/lib/ckan/default/src/ckanext-dcatdonl/ckanext/dcatdonl/resources/solr/  
↳7.4/spellings.txt /var/solr/data/ckan/conf  
sudo ln -s /usr/lib/ckan/default/src/ckanext-dcatdonl/ckanext/dcatdonl/resources/solr/  
↳7.4/stopwords.txt /var/solr/data/ckan/conf  
sudo ln -s /usr/lib/ckan/default/src/ckanext-dcatdonl/ckanext/dcatdonl/resources/solr/  
↳7.4/synonyms.txt /var/solr/data/ckan/conf  
sudo ln -s /usr/lib/ckan/default/src/ckanext-dcatdonl/ckanext/dcatdonl/resources/solr/  
↳7.4/synonyms_themes.txt /var/solr/data/ckan/conf  
sudo ln -s /usr/lib/ckan/default/src/ckanext-dcatdonl/ckanext/dcatdonl/resources/solr/  
↳7.4/synonyms_themes_hierarchy.txt /var/solr/data/ckan/conf  
sudo service solr restart
```

If you want to use the `ckanext-dcatdonl solr` optimizations for earlier CKAN versions it is advised to use the files present in the `ckanext/dcatdonl/resources/solr/5.5` directory instead.

4.4 Setting up the background process

In order for the `ckanext-dcatdonl` plugin to function properly, a background process must run at least once a day. This background process retrieves the latest versions of the valuelists and saves these locally. This process is run by executing the following command once a day via a CRON job for example.

```
python /usr/lib/ckan/default/src/ckanext-dcatdonl/ckanext/dcatdonl/task/controlled_  
↳vocabulary_updater.py
```

Ensure that the python script has READ and WRITE access to the following directory and its contents

The extension provides a `.sh` file which executes the above command, this file can easily be added to your servers crontab. This file is located in `shell/valuelist_updater.sh`.

```
/usr/lib/ckan/default/src/ckanext-dcatdonl/ckanext/dcatdonl/resources/controlled_  
↳vocabularies
```

The extension can function without the background process running, however this means that the valuelists that are used as part of the DCAT-AP-DONL metadata standard will never be updated.

This plugin does not introduce new functionality into CKAN. Rather, it augments certain actions of the CKAN installation by extending its schema. The usage of these augmented actions is detailed in this chapter.

5.1 Creating or updating datasets and resources

This extension adds several new mandatory and optional properties. A [Postman collection](#)⁸ is available that includes examples on how to communicate with a CKAN installation that uses ckanext-dcatdonl.

Below several examples are given on how to interact with CKAN.

5.1.1 Minimum dataset creation request

(POST) `{{CKAN_HOST}}/api/3/action/package_create`

```
{
  "owner_org":           "{{ORG_ID}}",
  "identifier":          "https://www.mijn.organisatie.nl/datasets/mijndataset1",
  ↪,
  "name":                "mijndataset1",
  "title":               "mijndataset1",
  "notes":               "De omschrijving van mijndataset1!",
  "metadata_language":  "http://publications.europa.eu/resource/authority/
  ↪language/NLD",
  "authority":           "http://standaarden.overheid.nl/owms/terms/'s-
  ↪Hertogenbosch",
  "publisher":           "http://standaarden.overheid.nl/owms/terms/Centraal_
  ↪Bureau_voor_de_Statistiek",
  "contact_point_name": "John Doe",
  "license_id":          "http://creativecommons.org/licenses/by/4.0/deed.nl",
}
```

(continues on next page)

⁸ <https://www.getpostman.com/collections/c54a66d658d1dec274bb>

(continued from previous page)

```

    "language": [
      ↪ "http://publications.europa.eu/resource/authority/
↪ language/NLD"
    ],
    "theme": [
      ↪ "http://standaarden.overheid.nl/owms/terms/
↪ Arbeidsomstandigheden_(thema)"
    ],
    "modified": "2018-04-11T09:00:00"
  }

```

5.1.2 Full dataset creation request

(POST) `{{CKAN_HOST}}/api/3/action/package_create`

```

{
  "owner_org": "{{ORG_ID}}",
  "identifier": "https://www.mijn.organisatie.nl/datasets/mijndataset1",
  ↪ ",
  "alternate_identifier": [
    ↪ "https://www.cbs.nl/datasets/denbosch-mijndataset1",
    ↪ "
  ],
  "language": [
    ↪ "http://publications.europa.eu/resource/authority/
↪ language/NLD"
  ],
  "source_catalog": "https://data.overheid.nl",
  "authority": "http://standaarden.overheid.nl/owms/terms/'s-
↪ Hertogenbosch",
  "publisher": "http://standaarden.overheid.nl/owms/terms/Centraal_
↪ Bureau_voor_de_Statistiek",
  "contact_point_email": "john.doe@cbs.nl",
  "contact_point_address": "Straatnaam 12, 1234AB, Amsterdam",
  "contact_point_name": "John Doe",
  "contact_point_phone": "020 - 1234567",
  "contact_point_website": "https://cbs.nl",
  "contact_point_title": "",
  "access_rights": "http://publications.europa.eu/resource/authority/
↪ access-right/PUBLIC",
  "url": "https://www.mijn.organisatie.nl/datasets/mijndataset1",
  ↪ ",
  "conforms_to": [
    ↪ "https://standaarden.nl/mijn_standaard"
  ],
  "related_resource": [
    ↪ "https://www.mijn.organisatie.nl/datasets/
↪ anderedataset12"
  ],
  "source": [
    ↪ "https://www.mijn.organisatie.nl/datasets/
↪ mijndataset0"
  ],
  "version": "1.0",
  "has_version": [

```

(continues on next page)

(continued from previous page)

```

        "https://www.mijn.organisatie.nl/datasets/
↔mijndataset0"
    ],
    "is_version_of":
    [
        "https://www.mijn.organisatie.nl/datasets/
↔mijndataset2",
        "https://www.mijn.organisatie.nl/datasets/
↔mijndataset3"
    ],
    "legal_foundation_ref": "art-1",
    "legal_foundation_uri": "http://wetten.nl/de-wet",
    "legal_foundation_label": "De wet!",
    "frequency": "http://publications.europa.eu/resource/authority/
↔frequency/DAILY",
    "provenance":
    [
        "https://www.mijn.organisatie.nl/datasets/uitleg"
    ],
    "sample":
    [
        "https://www.mijn.organisatie.nl/datasets/
↔mijndataset1/samples"
    ],
    "license_id": "http://creativecommons.org/licenses/by/4.0/deed.nl",
    "name": "mijndataset1",
    "title": "mijndataset1",
    "notes": "De omschrijving van mijndataset1!",
    "tags":
    [
        { "name": "mijn" },
        { "name": "dataset" },
        { "name": "een" },
        { "name": "Den Bosch" },
        { "name": "CBS" }
    ],
    "metadata_language": "http://publications.europa.eu/resource/authority/
↔language/NLD",
    "theme":
    [
        "http://standaarden.overheid.nl/owms/terms/
↔Arbeidsomstandigheden_(thema)"
    ],
    "modified": "2018-04-11T09:00:00",
    "spatial_scheme":
    [
        "http://standaarden.overheid.nl/owms/4.0/doc/
↔waardelijsten/overheid.gemeente"
    ],
    "spatial_value":
    [
        "http://standaarden.overheid.nl/owms/terms/'s-
↔Hertogenbosch"
    ],
    "temporal_label": "Jaar 2017",
    "temporal_start": "2017-01-01T00:00:00",
    "temporal_end": "2017-12-31T23:59:00",
    "dataset_status": "http://data.overheid.nl/status/beschikbaar",
    "date_planned": "2018-01-11T13:29:00",
    "high_value": "True",
    "basis_register": "False",
    "referentie_data": "True"
}

```

5.1.3 Minimum resource creation request

(POST) {{CKAN_HOST}}/api/3/action/resource_create

```
{
  "package_id":          "{{ PACKAGE_ID }}",
  "name":                "myresource1",
  "url":                 "http://my.organization.com/mydataset/myresource1",
  "description":         "My dataset description",
  "metadata_language":  "http://publications.europa.eu/resource/authority/
↪ language/NLD",
  "format":              "http://publications.europa.eu/resource/authority/
↪ file-type/ZIP",
  "language":            "http://publications.europa.eu/resource/authority/
↪ language/NLD",
  "license_id":         "http://creativecommons.org/publicdomain/mark/1.0/
↪ deed.nl"
}
```

5.1.4 Full resource creation request

(POST) {{CKAN_HOST}}/api/3/action/resource_create

```
{
  "package_id":          "{{ PACKAGE_ID }}",
  "name":                "myresource1",
  "url":                 "http://my.organization.com/mydataset/myresource1",
  "description":         "My dataset description",
  "metadata_language":  "http://publications.europa.eu/resource/authority/
↪ language/NLD",
  "format":              "http://publications.europa.eu/resource/authority/
↪ file-type/ZIP",
  "language":            "http://publications.europa.eu/resource/authority/
↪ language/NLD",
  "license_id":         "http://creativecommons.org/publicdomain/mark/1.0/
↪ deed.nl",
  "linked_schemas":    "http://some.standard.nl/reference",
  "size":                1234567890,
  "download_url":       "http://my.organization.com/mydataset/myresource1.zip",
  ↪ ",
  "mimetype":           "https://www.iana.org/assignments/media-types/
↪ application/activity+json",
  "release_date":       "2017-12-31T15:16:00",
  "rights":              "",
  "status":              "http://purl.org/adms/status/Completed",
  "modification_date":  "2018-01-03T12:09:00",
  "hash":                "dfuyhdf;lgkjlwriyuwefhsdkf",
  "hash_algorithm":     "SHA1",
  "documentation":      "http://my.organization.com/mydataset/documentation"
}
```

5.2 Viewing a dataset or resource

When viewing a dataset or resource of a CKAN installation running the ckanext-dcatdonl extension additional fields will be shown to the consumer. These additional properties are in line with the schema provided in the CKAN Schema chapter of this documentation.

5.2.1 data.overheid.nl

Data.overheid.nl maintains several additional properties for datasets that may be encountered when viewing datasets and resources. These properties are detailed below:

Dataset.duplicate_resources States which resources have duplicates on data.overheid.nl

Resource.link_status States if the given download or accessUrl is available for consumption

Resource.link_last_checked_date States when the link_status was last updated

Resource.is_duplicate_of States the resource id that is a duplicate of this resource

5.3 Searching for datasets or resources

As this extension adds several new properties to datasets and resources, the CKAN search features can now also search for datasets or resources based on these new properties. Consult the CKAN Schema chapter of this documentation for a complete list of new properties that can be searched for.

The `ckanext-dcatdonl` extension extends the CKAN schema for datasets and resources in such a way that they are capable of holding metadata according to the DCAT-AP-DONL 1.1 metadata standard.

Not all the names of the properties used in this plugin match the vocabulary used in the DCAT-AP-DONL 1.1 metadata standard specification. Instead, the plugin tries to follow the mapping outlined in github.com/ckan/ckanext-dcat#rdf-dcat-to-ckan-dataset-mapping⁹.

6.1 DCAT to CKAN mapping

In the tables below the exact mapping of DCAT-AP-DONL properties to their CKAN schema counterparts is shown.

6.1.1 DCAT Dataset

DCAT Property	CKAN Property
Dataset.identifier	Dataset.identifier
Dataset.description	Dataset.notes
Dataset.title	Dataset.name
Dataset.language	Dataset.language
Dataset.license	Dataset.license_id
Dataset.modified	Dataset.modified
Dataset.contactPoint	Dataset.contact_point_name
	Dataset.contact_point_email
	Dataset.contact_point_website
	Dataset.contact_point_phone
	Dataset.contact_point_address
	Dataset.contact_point_title

Continued on next page

⁹ <https://github.com/ckan/ckanext-dcat#rdf-dcat-to-ckan-dataset-mapping>

Table 1 – continued from previous page

DCAT Property	CKAN Property
Dataset.distribution	Dataset.resources
Dataset.keyword	Dataset.tags
Dataset.publisher	Dataset.publisher
Dataset.theme	Dataset.theme
Dataset.landingPage	Dataset.url
Dataset.spatial	Dataset.spatial_scheme
	Dataset.spatial_value
Dataset.temporal	Dataset.temporal_label
	Dataset.temporal_start
	Dataset.temporal_end
Dataset.authority	Dataset.authority
Dataset.accessRights	Dataset.access_rights
Dataset.conformsTo	Dataset.conforms_to
Dataset.documentation	Dataset.documentation
Dataset.frequency	Dataset.frequency
Dataset.hasVersion	Dataset.has_version
Dataset.isVersionOf	Dataset.is_version_of
Dataset.otherIdentifier	Dataset.alternative_identifier
Dataset.provenance	Dataset.provenance
Dataset.relatedResource	Dataset.related_resource
Dataset.releaseDate	Dataset.issued
Dataset.sample	Dataset.sample
Dataset.source	Dataset.source
Dataset.version	Dataset.version
Dataset.versionNotes	Dataset.version_notes
Dataset.grondslag	Dataset.legal_foundation_ref
	Dataset.legal_foundation_uri
	Dataset.legal_foundation_label
Dataset.datasetStatus	Dataset.dataset_status
Dataset.datePlanned	Dataset.date_planned
Dataset.highValue	Dataset.high_value
Dataset.basisregister	Dataset.basis_register
Dataset.referentieData	Dataset.referentie_data

6.1.2 DCAT Distribution

DCAT Property	CKAN Property
Distribution.accessURL	Resource.url
Distribution.description	Resource.description
Distribution.format	Resource.format
Distribution.license	Resource.license_id
Distribution.byteSize	Resource.size
Distribution.checksum	Resource.hash
	Resource.hash_algorithm
Distribution.documentation	Resource.documentation
Distribution.downloadURL	Resource.download_url
Distribution.language	Resource.language
Distribution.linkedSchemas	Resource.linked_schemas
Distribution.mediaType	Resource.mimetype
Distribution.releaseDate	Resource.release_date
Distribution.rights	Resource.rights
Distribution.status	Resource.status
Distribution.title	Resource.name
Distribution.modified	Resource.modification_date

6.1.3 DCAT CatalogRecord

DCAT Property	CKAN Property
CatalogRecord.modified	Dataset.metadata_modified
CatalogRecord.conformsTo	Dataset.conforms_to
CatalogRecord.changeType	Dataset.changetype
CatalogRecord.listingDate	Dataset.metadata_created
CatalogRecord.description	Dataset.notes
CatalogRecord.language	Dataset.metadata_language
CatalogRecord.sourceMetadata	Dataset.source_catalog
CatalogRecord.title	Dataset.title

6.2 Dataset

Property	Description	Validation
identifier	A global identifier that identifies the dataset	Required, String, Is URI
alternate_identifier	Alternate identifiers that identify the dataset	Optional, List, Are URIs
language	The languages used for the data found in the dataset	Required, From donl:language
authority	Entity that is responsible for the contents of the dataset	Required, String, From donl:organizations
publisher	Entity responsible for maintenance and publication of the dataset	Required, String, From donl:organizations
contact_point_email	Email of the contact point	Optional, String
contact_point_address	Address of the contact point	Optional, String

Continued on next page

Table 2 – continued from previous page

Property	Description	Validation
contact_point_name	Name of the contact point	Required, String
contact_point_phone	Phonenumber of the contact point	Optional, String
contact_point_website	Webaddress of the contact point	Optional, String
contact_point_title	Title of the contact point, if it describes a person	Optional, String
access_rights	The level of openness of the dataset	Optional, String, From overheid:openhaardeidsniveau
url	Webpage that provides additional information about the dataset, its metadata or its authority	Optional, String, Is URL
conforms_to	Standards the dataset conforms to	Optional, List, Are URIs
related_resource	Resources related to the dataset	Optional, List, Are URIs
source	Dataset on which this dataset is based	Optional, List, Are URIs
version	The version of the dataset	Optional, String
version_notes	Version notes of the dataset	Optional, List, Strings
issued	Date and time on which the dataset was published	Optional, String, yyyy-mm-ddThh:mm:ss
has_version	References to datasets which are based on this dataset	Optional, List
is_version_of	References to datasets on which this dataset is based	Optional, List, Are URIs
legal_foundation_ref	specific reference of the legal_foundation	Optional, String
legal_foundation_uri	URI of the legal_foundation	Optional, String, Is URI
legal_foundation_label	Label of the legal foundation	Optional, String
frequency	How often the dataset is updated	Optional, String, From overheid:frequency
provenance	Webpages that describe how this dataset came to be	Optional, List, Are URLs
documentation	Webpages about the dataset	Optional, List, Are URLs
sample	Sample data of the dataset	Optional, List, Are URLs
license_id	The license that applies to the dataset	Required, From overheid:license
title	The title of the dataset	Required, String
notes	The description of the dataset	Required, String
tags	Keywords to describe the dataset	Optional, List
metadata_language	The language used in the metadata of the dataset	Required, List, From donl:language
theme	One or more themes that describe the dataset	Required, List, From overheid:taxonomiebeleidsgaenda
source_catalog	The original catalog of the dataset	Optional, From donl:catalogs
changetype	The latest action taken on the dataset	From adms:changetype, ckanext-dcatdonl will set the correct value for this property
modified	The date and time this dataset was last modified	Required, String, yyyy-mm-ddThh:mm:ss
spatial_scheme	The schemes of the spatial value	Optional, List, From overheid:spatial_scheme
spatial_value	Geographical locations based on the spatial_schemes provided	Optional, List, Validates against schemes defined in spatial_scheme
temporal_label	A name of a timeperiod	Optional, String

Continued on next page

Table 2 – continued from previous page

Property	Description	Validation
temporal_start	A point in time, together with temporal_end it describes a period in time	Optional, String, yyyy-mm-ddThh:mm:ss Must be smaller than temporal_end
temporal_end	A point in time, together with temporal_start it describes a period in time	Optional, String, yyyy-mm-ddThh:mm:ss Must be greater than temporal_start
dataset_status	State of the dataset, it describes the availability of the dataset	Optional, String, From overheid:datasetStatus, defaults to the URI for beschikbaar
date_planned	The date and time upon which it is planned that the dataset becomes available	Optional, String, yyyy-mm-ddThh:mm:ss
high_value	Indicates this dataset is considered of 'high value' by the Dutch government	Optional, Boolean, defaults to False
basis_register	Indicates this dataset is part of the Dutch 'basis-register'	Optional, Boolean, defaults to False
referentie_data	Indicates this dataset contains <i>highly</i> reusable data	Optional, Boolean, defaults to False

6.3 Resource

Property	Description	Validation
url	The URL used to access the resource	Required, String, Is URI
name	The name of the resource	Required, String
description	A description of the resource	Required, String
metadata_language	The language used in the metadata of the resource	Required, String, From donl:language
language	The languages used for the data found in the resource	Required, List, From donl:language
license_id	The license that applies to the resource	Required, From overheid:license
format	The format of the resource	Required, String, From mdr:filetype_nal
size	The size of the contents of the resource in kilobytes	Optional, Positive integer
download_url	List of URLs referring to downloadable variants of the resource	Optional, List, Are URLs
mimetype	Mimetype of the resource	Optional, String, From iana:mediatypes
release_date	The date the resource was released	Optional, String, yyyy-mm-ddThh:mm:ss
rights	Rights that apply to the resource	Optional, String
status	Distributionstatus of the resource	Optional, String, From adms:distributiestatus
modification_date	Date on which this resource was last modified	Optional, String, yyyy-mm-ddThh:mm:ss
linked_schemas	Standards the resource applies to	Optional, List, Are URIs
hash	A hash calculated based on the contents of the resource	Optional, String
hash_algorithm	The hash algorithm used to determine the hash	Optional, String
documentation	A list of URLs that refer to documentation of the resource	Optional, List, Are URLs

6.4 Schema validation

Outlined below are the possible validation messages that the ckanext-dcatdonl plugin can generate based on the input it is given. The standard CKAN validation messages are not included in this documentation.

6.4.1 Validation messages

website, email or phone is required for the contact_point

This error occurs when the given dataset does not contain a value for either `contact_point_website`, `contact_point_email` or `contact_point_phone`. At least one of these three properties must be provided in order for the dataset to be considered valid.

when hash is provided, has_algorithm must too be provided

when hash_algorithm is provided, hash must too be provided

This error occurs when either the property `hash` or `hash_algorithm` is present, but its counterpart is not. When either is provided, both are required.

legal_foundation_ref must be provided when providing any of the legal_foundation_* properties

legal_foundation_uri must be provided when providing any of the legal_foundation_* properties

legal_foundation_label must be provided when providing any of the legal_foundation_* properties

When any of the `legal_foundation_*` properties is given, all are required.

spatial_value cannot be validated without a corresponding spatial_scheme

spatial_scheme must be accompanied by a spatial_value

These errors may occur when the request body contains a `spatial_scheme` but not a `spatial_value` or vice versa. Both properties are required to provide spatial metadata. To resolve this, provide both properties in the request body.

Spatial validation

This complex validator spans the fields `spatial_scheme` and `spatial_value`. Both fields are optional. However when one is provided, the other must too be provided. Furthermore the value of `spatial_scheme` determines the validator that will be used on `spatial_value`. In the table below the `spatial_value` validators are shown for the possible values of `spatial_scheme`.

<code>spatial_scheme</code> (base= http://standaarden.overheid.nl)	<code>spatial_value</code> validation
<code>/owms/4.0/doc/waardelijsten/overheid.gemeente</code>	Required, String, From <code>overheid:spatial_gemeente</code>
<code>/owms/4.0/doc/waardelijsten/overheid.koninkrijksdeel</code>	Required, String, From <code>overheid:spatial_koninkrijksdeel</code>
<code>/owms/4.0/doc/waardelijsten/overheid.provincie</code>	Required, String, From <code>overheid:spatial_provincie</code>
<code>/owms/4.0/doc/waardelijsten/overheid.waterschap</code>	Required, String, From <code>overheid_spatial_waterschap</code>
<code>/owms/4.0/doc/syntax-codeerschemas/overheid.epsg28992</code>	Required, String, Regex match <code>^\d{6}(\.\d{3})?\d{6}(\.\d{3})?\$</code>
<code>/owms/4.0/doc/syntax-codeerschemas/overheid.postcodehuisnummer</code>	Required, String, Regex match <code>^[1-9]\d{3}([A-Z]{2}(\d+(\d\$+)?))?\$</code>

So when a list of spatial schemes is provided, e.g.

```
[  
  "http://standaarden.overheid.nl/owms/4.0/doc/waardelijsten/overheid.gemeente",  
  "http://standaarden.overheid.nl/owms/4.0/doc/waardelijsten/overheid.waterschap"  
]
```

Then the values in the list of `spatial_value` must be values validate against the validators defined in the table above. In this example the values must either be values of the valuelist `overheid:spatial_gemeente` or values of the valuelist `overheid:spatial_waterschappen`.

value {{{ value }}} is not a valid spatial according to the schemes provided

This error occurs when one of the values of the `spatial_value` property does not validate against the schemas provided in the `spatial_scheme` property. To correct this, either update the `spatial_value` or the `spatial_scheme` values so that they are in sync.

value must be a valid date (yyyy-mm-ddThh:mm:ss)

This error occurs when the temporal metadata is provided in the wrong datetime format. Ensure that all temporal metadata is provided in the `yyyy-mm-ddThh:mm:ss` format, e.g. `2017-12-31T13:15:00`.

temporal_start cannot be greater or equal to temporal_end

temporal_end cannot be smaller or equal to temporal_start

These errors occur when providing illegal temporal data. To resolve this, provide temporal data where the `temporal_start` property contains a date that is smaller than the date provided in `temporal_end`.

value must be a string

value must be a list

value must be a dictionary

These errors occur when providing the wrong datatype for a specific value. To resolve these errors the right datatype must be provided.

value is not a valid uri

This error occurs when an uri string is expected. To resolve this, provide a valid http uri.

only one license can be provided, list given

This error occurs when a list of licenses is provided. Only one license can be provided for a dataset or resource.

value must have a id property

This error occurs when the dictionary provided for the license property does not contain the key `id`. The license dictionary provided must have a `id` property to validate the given license.

no matching license id found for given value

This error occurs when the license id provided is not part of the list of valid licenses. Consult the license valuelist at `{{ URL }}` for the supported licenses.

value `{{ value }}` is not a valid `{{ valuelist }}`

This error occurs when providing values that are not part of the given valuelist. Consult the given valuelist to see the acceptable values.

values do not meet the minimum requirements

values do not meet the maximum requirements

These errors occur when providing too little or too many values in a given list for a given property. Consult the schema defined in 'CKAN Schema' chapter of this documentation for the expectations of the amount of values.

values must be unique

This error occurs when the list of values provided contains duplicates. Remove the duplicates to resolve this error.

6.5 Valuelists

The following valuelists (AKA controlled vocabularies) are used to validate parts of the CKAN schemas.

Name	Location
adms:changetype	https://waardelijsten.dcat-ap-donl.nl/adms_changetype.json
adms:distributiestatus	https://waardelijsten.dcat-ap-donl.nl/adms_distributiestatus.json
donl:catalogs	https://waardelijsten.dcat-ap-donl.nl/donl_catalogs.json
donl:language	https://waardelijsten.dcat-ap-donl.nl/donl_language.json
donl:organization	https://waardelijsten.dcat-ap-donl.nl/donl_organization.json
iana:mediatypes	https://waardelijsten.dcat-ap-donl.nl/iana_mediatypes.json
mdr:filetype_nal	https://waardelijsten.dcat-ap-donl.nl/mdr_filetype_nal.json
overheid:datasetStatus	https://waardelijsten.dcat-ap-donl.nl/overheid_dataset_status.json
overheid:frequency	https://waardelijsten.dcat-ap-donl.nl/overheid_frequency.json
overheid:license	https://waardelijsten.dcat-ap-donl.nl/overheid_license.json
overheid:openbaarheidsniveau	https://waardelijsten.dcat-ap-donl.nl/overheid_openbaarheidsniveau.json
overheid:spatial_gemeente	https://waardelijsten.dcat-ap-donl.nl/overheid_spatial_gemeente.json
overheid:spatial_koninkrijksdeel	https://waardelijsten.dcat-ap-donl.nl/overheid_spatial_koninkrijksdeel.json
overheid:spatial_provincie	https://waardelijsten.dcat-ap-donl.nl/overheid_spatial_provincie.json
overheid:spatial_scheme	https://waardelijsten.dcat-ap-donl.nl/overheid_spatial_scheme.json
overheid:spatial_waterschap	https://waardelijsten.dcat-ap-donl.nl/overheid_spatial_waterschap.json
overheid:taxonomiebeleidsagenda	https://waardelijsten.dcat-ap-donl.nl/overheid_taxonomiebeleidsagenda.json

6.5.1 Structure

All valuelists are served as `application/json`. The contents of these valuelist follows the pattern outlined below, which shows a sample of the `overheid:taxonomiebeleidsagenda` valuelist:

```
{
  "http://standaarden.overheid.nl/owms/terms/Afval_(thema)": {
    "labels": {
      "nl-NL": "Afval",
      "en-US": "Rubbish"
    }
  },
  "http://standaarden.overheid.nl/owms/terms/Arbeidsomstandigheden_(thema)": {
    "labels": {
      "nl-NL": "Arbeidsomstandigheden",
      "en-US": "Labour conditions"
    }
  },
  ...
}
```

When supplying a value that must be part of a valuelist, provide the key of the value. In the example above this would be `http://standaarden.overheid.nl/owms/terms/Arbeidsomstandigheden_(thema)`. The labels are provided so that front-end applications can provide a proper translation of the value.

The one exception to the format displayed above is for the valuelist of `overheid:license`. The format deviates here to accommodate CKAN, since CKAN requires its license source file to fit a specific format. This format is displayed below.

```
[
  {
    "domain_content": false,
    "domain_data": false,
    "domain_software": false,
    "family": "",
    "id": "notspecified",
    "is_generic": true,
    "maintainer": "",
    "od_conformance": "not reviewed",
    "osd_conformance": "not reviewed",
    "status": "active",
    "title": "License Not Specified",
    "url": ""
  },
  ...
]
```

6.5.2 Caching

The `ckanext-dcatdonl` plugin caches the contents of the valuelists for up to 24 hours. On the first request made after midnight each day, the cache will be invalidated and rebuild. This means that any changes made to the contents of the valuelists can take up to 24 hours to take effect.