

JHS 193 Unique identifiers of geographic data

Annex 1. URI generation process

Version: 1.0

Issued on: 2 September 2015

Validity: until further notice

Contents

1	Introduction.....	1
2	Creating a local identifier in the database and defining lifecycle rules for unique geographic feature types	2
3	Generating a URI.....	2
3.1	Minting a dataset identifier.....	3
3.2	Type of the feature to be identified.....	3
3.2.1	Geographic features (so).....	4
3.2.2	Real-world objects (id).....	4
3.2.3	Concepts (def).....	5
4	Defining a data product and offering it in web services.....	5
5	Setting redirections.....	6

1 Introduction

This annex describes the actions with which URIs in accordance with this recommendation can be generated and released. The actions are:

- Creating local identifiers in the database and defining lifecycle rules for geographic feature types (*Chapter 2*)
- Generating globally unique URIs for geographic features (*Chapter 3*)
- Defining a data product including data features identified using the identifiers and releasing it in web services (*Chapter 4*)
- Defining the redirection of the generated URIs in the national URI redirection service for geographic information (*Chapter 5*).

Through the adoption of URIs, data producers can promote the utilisation and reuse of their data more extensively and obtain user feedback on how the quality of data can be improved and on new needs for using data. Furthermore, the adoption of URIs promotes the adoption and utilisation of linked data.

Initially, the purpose of this recommendation is to support the generation and use of URIs in geographic datasets. However, the guidelines of this recommendation can also be applied to datasets and registers in which geographic information (geometry) about features is not recorded.

2 Creating a local identifier in the database and defining lifecycle rules for unique geographic feature types

A field is created in the database of the data-producing system for the unique local identifier of a geographic feature. For example, internal identifiers of systems or other standard identifiers can be used as the local identifier (localId), provided that they are:

- Unique
- Persistent
- Traceable
- Feasible

If no identifiers fulfilling the aforementioned conditions exist, a 128-character Universally Unique Identifier (UUID) in accordance with the *RFC 4122* standard can be used as the local identifier. Several data management systems and programming languages have modules for the generation of UUID.

To fulfil the requirements set for local identifiers, lifecycle rules must be defined for the geographic feature types identified by the identifiers. Lifecycle rules of a geographic feature represent changes in the geographic feature throughout its lifecycle. They are needed to conclude whether a change in a geographic feature causes changes in its identity and unique identifier or only a new version identifier. If lifecycle rules require the use of a version identifier, a field for it must be added to the database.

More detailed instructions on how to generate lifecycle rules are presented in *Annex 3* of this recommendation.

3 Generating a URI

The format of unique URIs in accordance with this recommendation is

`http://paikkatiedot.fi/{type}/{dataset identifier}/{local identifier}`, where

- type = the type of the object to be identified
 - type of geographic features = so
 - type of real-world objects = id
 - type of concepts = def
- dataset identifier = the dataset identifier of the Finnish discovery service Paikkatietohakemisto (GeoNetwork)
- local identifier = a unique local identifier generated in the database of the data-producing system

The figure below illustrates how URIs in accordance with this recommendation are generated.

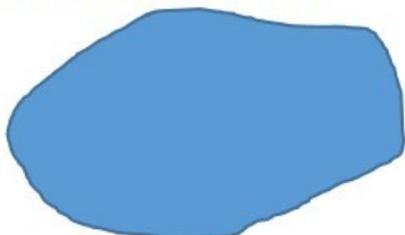
Standing water or stagnant water:

<http://paikkatiedot.fi/def/234567/vakavesi>



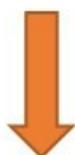
modelled as concepts

Real world: Lake Saimaa



<http://paikkatiedot.fi/id/123456/abc123>

INSPIRE data producer responsible for inland waters: SYKE



modelled as geographic features

SYKE creates an ID for the real-world object and provides its INSPIRE geographic feature with a corresponding /so/ identifier:

<http://paikkatiedot.fi/so/123456/hydrography/StandingWater/abc123>

3.1 Minting a dataset identifier

The dataset is described in the Finnish discovery service Paikkatietohakemisto (GeoNetwork, <http://www.paikkatietohakemisto.fi>). Paikkatietoikkuna and the operating instructions of GeoNetwork offer instructions on how to describe datasets. Public administration recommendation *JHS 158 Metadata for the geographic information* includes a summary of the requirements of the INSPIRE Directive and standards. The availability of data using URIs can be improved by recording metadata, including that of other than INSPIRE datasets, in accordance with the recommendation *JHS 158 Metadata for the geographic information*.

The dataset identifier of GeoNetwork can be obtained from metatieto@maanmittauslaitos.fi.

3.2 Type of the feature to be identified

In accordance with this recommendation, features to be identified include geographic features (so), real-world objects (id) and geographic concepts (def).

3.2.1 Geographic features (so)

The identifier of a geographic feature is expressed in the URI path using the /so/ path component (so = spatial object). The establishment of the /so/ component is mandatory in the JHS recommendation.

The general structure of the URI of a geographic feature is

`http://paikkatiedot.fi/so/{dataset identifier}/{local identifier}[/{version identifier}]`

For INSPIRE geographic features, the URI path must also include the INSPIRE theme and class:

`http://paikkatiedot.fi/so/{dataset identifier}/{theme}/{class}/{local identifier}[/{version identifier}]`

For example, `http://paikkatiedot.fi/so/123456/hy/StandingWater/abc123`

Both INSPIRE geographic features and corresponding geographic features of the original data source have separate URIs. The URI of the geographic feature of the original data source corresponding with the INSPIRE geographic feature does not include the semantic components "theme" and "class."

Datasets derived from the original dataset, such as INSPIRE data products, have a separate dataset identifier and, therefore, also a separate namespace. Objects of the derived dataset can have the same local identifier as the original dataset, if this is purposeful, for example, considering the updates of the derived dataset.

The version identifier of a geographic feature is used in accordance with the lifecycle rules of the geographic feature type. The version identifier is not mandatory. If no lifecycle rules have been defined, the version identifier can, for example, be the creation date of the geographic feature. (For example, a timestamp in accordance with *ISO 8601*: 2014-01-19 T12:38:31+03:00 or 2014-01-19.)

A geographic feature can have several forms of portrayal or one or more concepts can correspond with it. These are linked to the documentation identifier (doc). When establishing an identifier of a geographic feature (so), it is redirected to this documentation identifier (see *Chapter 4* and *Annex 2 of this recommendation*). The easiest way is to link geographic features to their corresponding concepts when generating the URI for them. By linking geographic features to concepts, it is easier to trace data in conjunction with open and linked data.

3.2.2 Real-world objects (id)

The identifier of a real-world object or phenomenon represented by a geographic feature is expressed in the URI path using the /id/ path component. The identifier of a real-world object acts as an identifier linking information between geographic features (/so/) representing the same real-world object and as a link bridging data features and resources of different domain names, enabling the extensive use of linked data.

The general URI of a real-world object is:

`http://paikkatiedot.fi/id/{dataset identifier}/{local identifier}`

For example, `http://paikkatiedot.fi/id/123456/abc123`

The authority responsible for an INSPIRE geographic feature establishes an identifier for the real-world object corresponding with the geographic feature while it establishes the URI of the instance of the geographic feature. To secure uniqueness, the same dataset identifier and local identifier are used in the identifier of the instance as are used in the identifier of the real-world object it models.

For objects other than INSPIRE geographic features, all providers can establish real-world identifiers, in which case the identifiers used most extensively for linking become de facto standards. Considering linked

data, identifiers of different providers referring to a single real-world object can be linked using, for example, owl:sameAs or skos:exactMatch linking.

One or more geographic features or concepts may correspond with a real-world object. These are linked to the documentation identifier. When establishing an identifier of a real-world object, it is redirected to this documentation identifier (see *Chapter 4* and *Annex 2 of this recommendation*).

3.2.3 Concepts (def)

The identifier of a concept is expressed in the URI path using the /def/ path component (def = definition). The general URI of a concept is:

```
http://paikkatiedot.fi/def/{dataset identifier}/{local identifier}
```

The concept source can be any vocabulary used by the data producer and provided with a URL in URI format. The vocabulary may be an ontology, data specification, schema, code list, taxonomy or thesaurus. The dataset identifier of the vocabulary used as the concept source can be obtained from metatieto@maanmittauslaitos.fi. For example, the dataset identifier of the Finnish Geospatial Domain Ontology is 1001000.

One or more formats may correspond with a concept. These are linked to the documentation identifier. When establishing an identifier of a concept, it is redirected to this documentation identifier (see *Chapter 4* and *Annex 2 of this recommendation*).

The Finnish Thesaurus and Ontology Service (Finto) is used as a distribution channel for concepts suitable for the service. More information: <http://finto.fi/fi/>.

4 Defining a data product and offering it in web services

The geographic dataset recorded in the database of the data producer is offered to users in the form of data products designed and modified for different purposes of use. A data product defined by the data product specification may follow the original data model of the database or, if necessary, data products corresponding better to the needs of users can be modified from the original data model. The recommendation *JHS 177 Specification of a geographic data product* instructs that geographic data products be defined as XML schemas and product documents as data product descriptions, and it presents a specification process aiming at these results.

INSPIRE data product specifications describe the content and structure of standardised European geographic data products. INSPIRE data product specifications require that URIs are released for the geographic features listed in *Chapter 5*. The URI of an INSPIRE geographic feature is released in the Identifier type, the sub-elements of which are defined as follows:

- localId = the local identifier of the geographic feature
- namespace = the first part of the URI of the geographic feature: <http://paikkatiedot.fi/so/{dataset identifier}/{theme}/{class}/>
- versionId (not mandatory) = the version identifier of the geographic feature

In general, the format of the Identifier element required by INSPIRE is:

```
<base:Identifier>
```

```
<base:localId>{local identifier}</base:localId>
```

```
<base:namespace>http://paikkatiedot.fi/so/{dataset identifier}/{theme}/{class}/</base:namespace>
```

```
<base:versionId>[version identifier]</base:versionId>
```

```
</base:Identifier>
```

It is recommended that the URI of a geographic feature in datasets other than INSPIRE datasets be released using the `gml:identifier` element of the GML markup language standardised by OGC. The mandatory attribute of the Identifier element is `codeSpace`, in which the first part of the URI, `http://paikkatiedot.fi/so/{dataset identifier}/`, is placed. The local identifier and any version identifier are entered as values of the `gml:identifier` element. In general, the format of the `gml:identifier` element is:

```
<gml:identifier codeSpace="http://paikkatiedot.fi/so/{dataset identifier}"/>{local identifier}[//{version identifier}]</gml:identifier>
```

The `gml:id` attribute is also mandatory for GML data objects.

Offering data products in viewing and download services enables the effective distribution of geographic information and makes the adoption of unique identifiers easier. The recommendation *JHS 180 Content services for geographic information* includes guidelines on the implementation of different web services for geographic information and INSPIRE requirements related to service quality and testing.

5 Setting redirections

Identifiers assigned to geographic features, real-world objects and concepts in accordance with this recommendation are located under the domain name `paikkatiedot.fi`. `Paikkatiedot.fi` acts as a national redirection service for HTTP URI identifiers of geographic information, and it redirects incoming requests to the URI service offered by the data producer. An example implementation of a URI service is presented in *Annex 2*.

The purpose of the issuer of an identifier is to define in the `paikkatiedot.fi` service the URL to which requests directed to specific dataset features are redirected. Redirections are managed using a web application. At the initial stage, data producers, however, send definitions of redirections to INSPIRE secretaries (`metatieto@nls.fi`) who forward them to the `paikkatiedot.fi` redirection service.

Redirections concerning identifiers should be made to the service offered by a data producer, the response received from which is in accordance with the responding practices defined in this recommendation. *Annex 2* of this recommendation offers an example of the technical implementation of a service offering responses in accordance with the responding practices.

The URL to which a query sent to the identifier of an instance, real-world object or concept is redirected forms the documentation identifier, and its general format is:

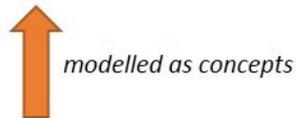
```
http://{domain name}/doc/{dataset identifier}/{local identifier}[//{version identifier}]
```

An example of redirection is offered below.

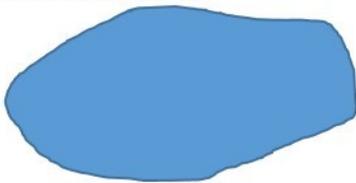
JUHTA – Advisory Committee on Information Management in Public Administration

Standing water or stagnant water (INSPIRE):
<http://inspire.ec.europa.eu/featureconcept/StandingWater>

Standing water or stagnant water redirected →
<http://paikkatiedot.fi/def/234567/vakavesi> redirected → [http://finto.fi/pto/fi/page/?uri=\[KÄSITTEEN_URI\]](http://finto.fi/pto/fi/page/?uri=[KÄSITTEEN_URI])
Lake Saimaa redirected →
<http://paikkatiedot.fi/def/234567/saimaa>

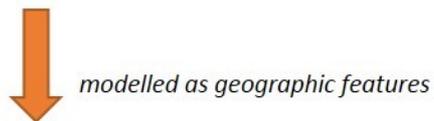


Real world: Lake Saimaa



<http://paikkatiedot.fi/id/123456/abc123>

INSPIRE data producer responsible for inland waters: SYKE
Other spatial data producer: City of Lappeenranta



SYKE creates an ID for the real-world object and provides its INSPIRE geographic feature with a corresponding /so/ identifier:

<http://paikkatiedot.fi/so/123456/hydrography/StandingWater/abc123>

redirected → <http://www.syke.fi/doc/123456/abc123>

The City of Lappeenranta has a geographic feature modelling Lake Saimaa in its dataset:

<http://paikkatiedot.fi/so/345678/qwe456>

redirected → <http://www.lpr.fi/doc/345678/qwe456>