

JHS 193 Unique identifiers of geographic data

Annex 4. Use case examples

Version: 1.0

Issued on: 2 September 2015

Validity: until further notice

Contents

1	Introduction.....	1
2	Use of open data.....	1
2.1	Use case to find and adopt open geographic data products.....	1
2.2	Viewpoint of data architecture for developers of open data.....	2
2.3	Stages.....	3
3	Links between data systems.....	5
3.1	Document links - an example of a service related to mineral resources.....	5
3.2	Synchronising maintenance using URIs - maintenance of bus stop data in the Digiroad2 system.....	5
4	Shared use using URIs - hydrography theme.....	7
4.1	Management of toxic emissions.....	7
4.2	Stages.....	7
5	Ontologies and linked data.....	8
5.1	Use of links at a metadata level.....	9
5.2	Use of links at a schema level.....	9
5.3	Use of links at metadata and schema levels.....	10
5.4	Use of links at a feature level.....	11

1 Introduction

This annex presents examples of use cases related to URIs. These use case examples have been used in the preparation of the recommendation.

2 Use of open data

2.1 Use case to find and adopt open geographic data products

The City of Tampere opened its first geographic data products under an open data licence in December 2012. These open geographic data products are distributed via the Web Feature Service interface (WFS) which currently offers 32 products. Open geographic datasets have several user groups, but this use case focuses to describe the use of open data from the point of view of an open data developer.

Open geographic datasets can be searched for from two services, the open data catalogue of Tampere or the Finnish discovery service Paikkatietohakemisto (GeoNetwork, <http://www.paikkatietohakemisto.fi>). Currently, GeoNetwork has metadata to represent only part of all open geographic datasets, but the data catalogue of Tampere includes all products. In the data catalogue of Tampere, metadata elements are not based on any metadata standard.

2.1.1 Searching for open geographic data products from GeoNetwork

A developer logs in to GeoNetwork at www.paikkatietohakemisto.fi. It starts to search for open datasets by entering, for example, "open data" in the search field, and the service lists all of the metadata descriptions in which "open data" appears. The metadata descriptions include the URLs of the open data licence and the physical Web Feature Service. The developer uses the Web Feature Service to download products.

However, the use of the Web Feature Service requires technical skills, which may form an obstacle for the use of the products. Currently, the metadata descriptions do not include any online references to Web Feature Service instructions. A metadata description includes the email address of the party responsible for the resource, from which the developer can ask for additional information regarding its open questions.

In GeoNetwork, a data provider identifies its products using a dataset identifier which, currently, is the only product-specific identifier. Products include unique identifiers maintained manually by the data producer and unique intra-system identifiers managed by data systems. Manually maintained object-specific identifiers involve the possibility of a human error and, therefore, present a risk from the viewpoints of persistence and the management of lifecycle rules. With regard to data systems, it should, however, be noted that keys produced by identification mechanisms are only intended for the internal operation of data systems. Only people can manage lifecycle rules of real-world objects and maintain these real-world keys in data systems.

The City of Tampere needs to start describing classified geographic datasets suitable for re-use in GeoNetwork, in which case the data producer should be able, in the update process, to link products and services to the Finnish Geospatial Domain Ontology on the basis of the URI mechanism. In addition, it should be possible in GeoNetwork to search for metadata descriptions of data products and services on the basis of licences (open data/service licence), which requires that products and services can be licenced in machine-readable format. Therefore, it would be important to link the licence service of geographic information (attribute provider) to the metadata of products and services.

These search combinations would enable a nationally effective search for open geographic information based on metadata and ontology.

2.1.2 Searching for open geographic data products from the open data catalogue of Tampere

A developer goes to www.tampere.fi/tampereinfo/avoindata.html. It starts to search for open data products by entering the name of the product being searched or a keyword in the search field. Alternatively, datasets can be browsed by clicking the "Seuraava (Next)" button at the bottom of the page.

The service lists search results that include the product name, description and a download link, through which the product can be downloaded in SHAPE-ZIP format. By clicking the product name, it is possible to view detailed product metadata. The metadata includes links to download the product in formats other than those offered by the Web Feature Service: JSON, GML2, GML32 and CSV. The developer downloads the product directly or uses the Web Feature Service as a data source for its application on the basis of the online query principle.

However, the use of the Web Feature Service requires technical skills, which may form an obstacle for the use of the products. Metadata descriptions include an online reference to Web Feature Service instructions. A metadata description includes the email address of the party responsible for the resource, from which the developer can ask for additional information regarding its open questions.

The future national open data catalogue project /VM is a linked project.

2.2 Viewpoint of data architecture for developers of open data

- Metadata descriptions and data product specifications of data products are important considering suitability evaluations.

JUHTA – Advisory Committee on Information Management in Public Administration

- Alternatively, attribute-specific metadata of products offers a quick and clear way to understand the content of each product.
- A permanent object-specific GML-featureID (fid) managed by the data provider must be a mandatory element of data products. It acts as a linking factor between the data producer and open data developer in the lifecycle management of the feature.
- Currently, JSON is the most popular format for product downloading.

2.3 Stages

T1: An open data developer has a need to have the bicycle network dataset of Tampere as part of its new service.

T2: The developer performs a free text search in the paikkatietohakemisto.fi service, and finds metadata of the dataset corresponding with its search. The metadata indicates that a data product entitled "Tampereen pyörätiet (Bicycle routes in Tampere)" with an open data licence is available. The metadata also indicates a physical URL for product adoption.

T3: The developer is also interested to see what other products are available regarding pedestrian routes. The user performs an ontology-based search using the ontology of a pedestrian route. The user also discovers data products "traffic light equipment", "traffic light-controlled junctions" and "traffic light indicators."

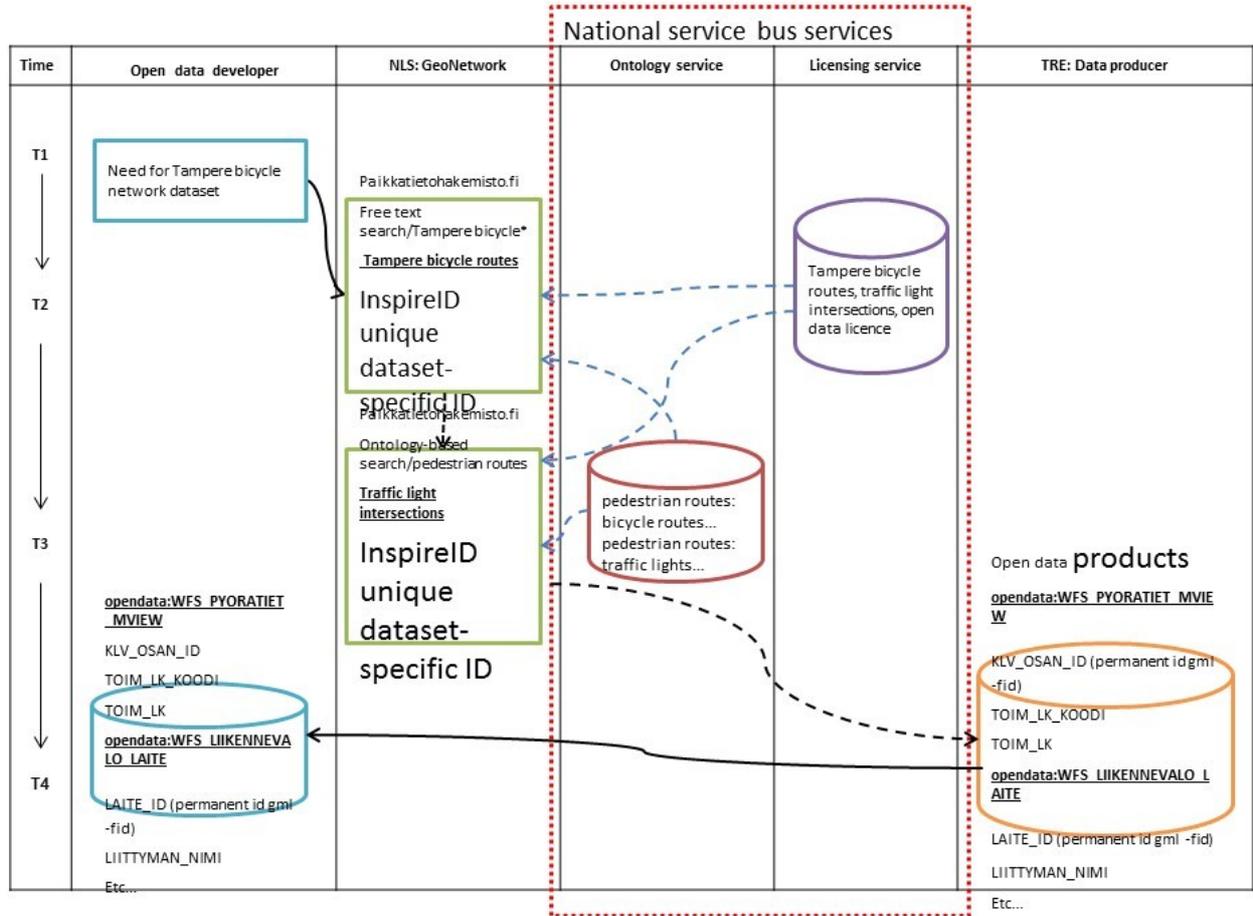
T4: The developer downloads the data products to its database via the Web Feature Service (WFS). In the future, the developer will update any changes in the data producer's data products according to a schedule on the basis of the permanent URI. Therefore, the permanent data system-independent URI is mandatory so that changes can be managed between data systems (see notes on the following page).

The data producer has linked the concepts of the ontology service to its products in GeoNetwork.

The data producer managers access rights to its products and services in the licence service, which has been linked to the metadata of the products and services in GeoNetwork.

In addition, replication requires an attribute in the product (e.g. date information), on the basis of which any changes can be identified without needing to download all data again. However, downloading all data may be the simplest and most sensible way to update any changes.

JUHTA – Advisory Committee on Information Management in Public Administration



3 Links between data systems

3.1 Document links - an example of a service related to mineral resources

A geographic feature, such as mineral deposit, also involves data other than that saved in the database, e.g. images, reports, scanned maps or other supplementary material, which has been saved in a directory.

The aim is to link the attached material to the geographic feature so that the link is saved in the database. The same attached material can be used in several different systems or services. The links saved in these different systems or services cannot change when, for example, servers are changed or directory structures are revised, because updating links in all locations is a laborious task.

If each image, report or other such attached material is provided with a unique identifier, through which linking is carried out at a level of systems and services, change management becomes easier.

This is only visible to users when they open attached material that interest them through a functional link.

3.1.1 Stages

A user zooms in on northern Finland on a map and uses the Info tool to open data about the features of mine XX in order to obtain more information related to, for example, estimated resources. The user is also interested in other material related to the mine, such as geological maps. In addition to attributes, the Info tool includes links to other material. By clicking the geological map link, a detailed geological map of the area opens up for the user.

T1: The user finds the desired feature using search tools of a map application or by browsing a map.

T2: The user opens the Info tool of the application and clicks a feature, the attributes of which the user is interested in.

T3: The application returns the attributes of the feature to the user in an information window. This information includes a link to attached material (redirection from the identifier of the geographic feature (so) to the documentation identifier (doc)).

T4: The service user clicks the link to open the document (e.g. PDF file) located behind the link.

T5: The attached document saved on a server opens for the user.

3.2 Synchronising maintenance using URIs - maintenance of bus stop data in the Digiroad2 system

A user from the City of Tampere wishes to transfer data from their operational bus stop system (Winbus) to the national Digiroad2 system.

3.2.1 Stages

T1: Information about bus stops in Tampere is maintained in the Winbus system (SQL server).

1. A new bus stop is created in the Digiroad2 system where it is assigned a new unique DIGIROAD_ID.
2. A new bus stop is created in the Winbus system, and the bus stop is assigned the new DIGIROAD_ID generated in Digiroad2. Other attributes are added to the feature.

T2: Tampere ETL process, once every 24 hours:

JUHTA – Advisory Committee on Information Management in Public Administration

Reads Winbus system data from the SQL server database.

Converts it according to the Digiroad2 schema.

Makes a coordinate system conversion in the ETRS-GK24 system.

Copies the converted data to the Oracle server.

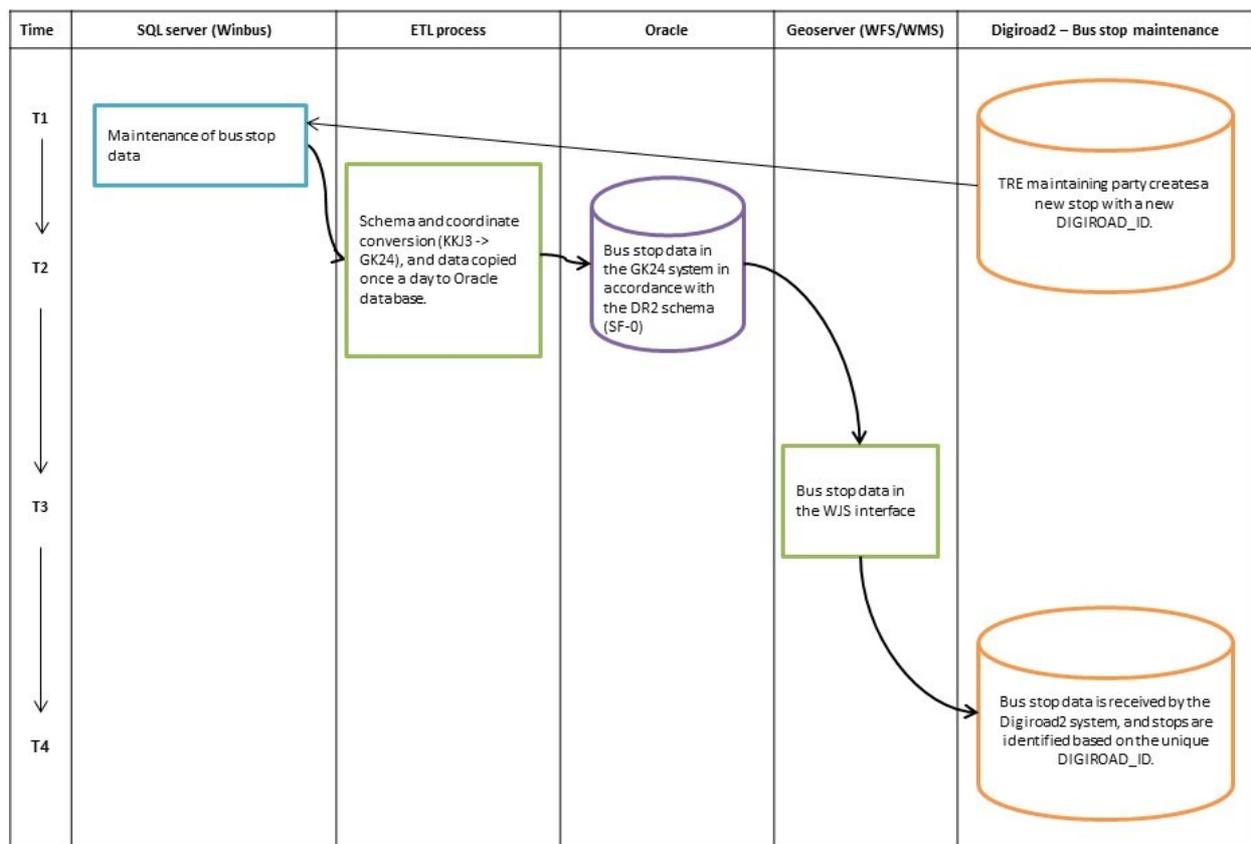
T3: The Oracle table is linked to the WFS interface for distribution in accordance with the Digiroad2 schema.

T4: The user requests new and changed data from the WFS interface to the Digiroad2 system in the TM35 coordinate system.

The identification of bus stops is based on the unique DIGIROAD_ID. Changes are saved in the Digiroad2 database.

T5: Use of data

Users use the Digiroad2 system as a primary system as intended, not the source system of the data producer. Users can request additional attributes for Digiroad2 features from the URI service of the data producer using the DIGIROAD_ID.



4 Shared use using URIs - hydrography theme

Data content related to the hydrography theme is maintained in several different organisations. The Topographic Database (MTK) of the National Land Survey of Finland (NLS) includes the most detailed and nationwide description of the physical characteristics of water systems. For example, the dataset describes lakes, ponds, rivers, ditches, rapids, dams and basins. The dataset focuses on the description of the geometry of water features.

The Finnish Environment Institute (SYKE) maintains an extensive dataset system which includes volumes of information about water-related features. SYKE uses geometries in accordance with MTK as source data for its datasets. SYKE has modified the data obtained from MTK so that the object-specific management of water-related features is possible. In addition, SYKE has generated a consistent channel network model representing the water system where a pseudo channel has been added for lakes, for example, due to the consistency requirement of the network.

The Finnish Transport Agency (FTA) maintains hydrographic information, particularly in relation to waterways and traffic.

When transferring data between organisations that maintain the data and different data systems, it would be necessary that water features could be processed using mechanisms based on permanent unique identifiers. This would be especially necessary when transferring updated data about the geometries of features from NLS to SYKE. It would also be useful to transfer other feature-specific attribute data from SYKE to NLS.

Feature identifiers would be particularly necessary from the viewpoint of smart applications that use different data sources. Below is an example of a hydrography-related use case where data related to a specific geographic feature is required from different organisations.

4.1 Management of toxic emissions

A user wishes to use hydrography data in order to simulate the impact of hazardous emissions from a plant located alongside a watercourse or to attempt to react to actual emissions. Important background information includes water areas and information about their depth, water channels and water flows, ports, swimming locations, existing boat routes, etc. The channel network of SYKE, including its water flow data, offers a starting point for modelling the spread of the emission. Channel-related features representing the physical environment, including their geometries, define the impact area of the emission more closely. The significance of the emission considering evacuation is emphasised in boating and swimming locations. The task is a success because the user can link relevant SYKE channels to up-to-date information of NLS about the dimensions of water areas on the basis of identifiers. Information about the water temperature can be linked to the analysis of the SYKE observation network using watercourse-specific identifiers. In addition, boat routes and swimming locations need to be linked to these water features, which can reliably be carried out on the basis of identifiers.

4.2 Stages

T1: A user (a person from the rescue department responsible for operations performing an evacuation plan on the toxic emission) locates the point of the emission using a map-based application. The user searches for the URIs of hydrographic features (lakes, rivers, etc.) related to the location from the background system using an area-based limitation. The most critical features can be found by navigating in the channel network in the direction of the water flow.

T2: Using the identifiers of the real-world objects (id) corresponding with the aforementioned geographic features (so), the user starts a search which returns all available identifiers of geographic features related to the specific real-world object.

T3: From among the returned identifiers, the user selects relevant features concerning the task at hand, such as swimming locations and ports.

T4: Using the identifiers of the individual geographic features discovered, the user starts a search which adds detailed information about these features to the application.

T5: The information indicates, for example, details of parties responsible for swimming locations and ports and, in this way, the user is able to issue the alarms required.

Note: T2 requires that the identifier of the real-world object can be found from each geographic feature related to the specific feature.

5 Ontologies and linked data

In addition to linking geographic features to real-world objects, geographic features should also be linked to concepts that describe them. Ontologies are sets of concepts presented in machine-readable format. Links to these can provide users of information with more information about the semantics of different geographic features. Links to ontologies can be made at three levels:

1. At a metadata level: a keyword in metadata is linked to its corresponding concept in the Finnish Geospatial Domain Ontology, such as a keyword concept in GeoNetwork or a concept of ontologised INSPIRE data product specifications.
2. At a schema level: features and attributes defined by a schema are linked to their corresponding concepts in an application ontology generated from schemas (e.g. INSPIRE data product schemas), and these concepts are linked to the Finnish Geospatial Domain Ontology.
3. At a feature level: links to ontologies are made directly from geographic features (so). In other words, geographic features are linked to their corresponding concepts in the aforementioned ontologies.

OGC 08-167r2

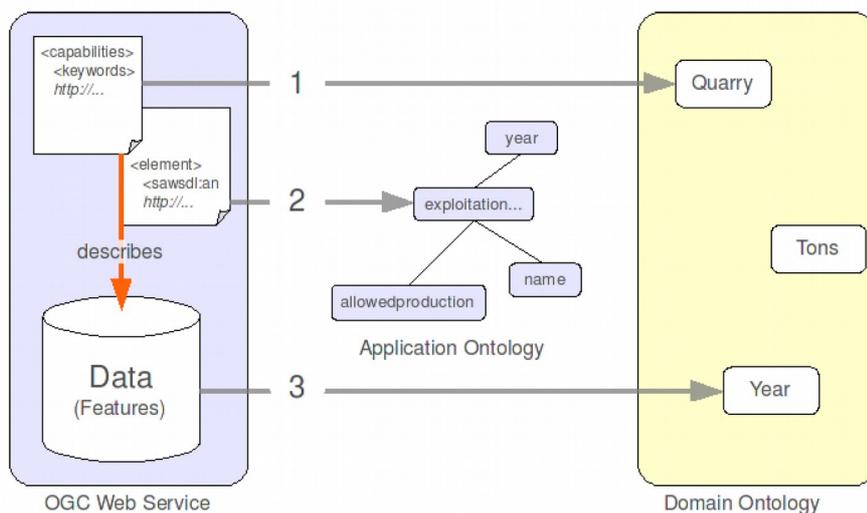


Figure 3 - Semantic annotations at three different levels

In the figure

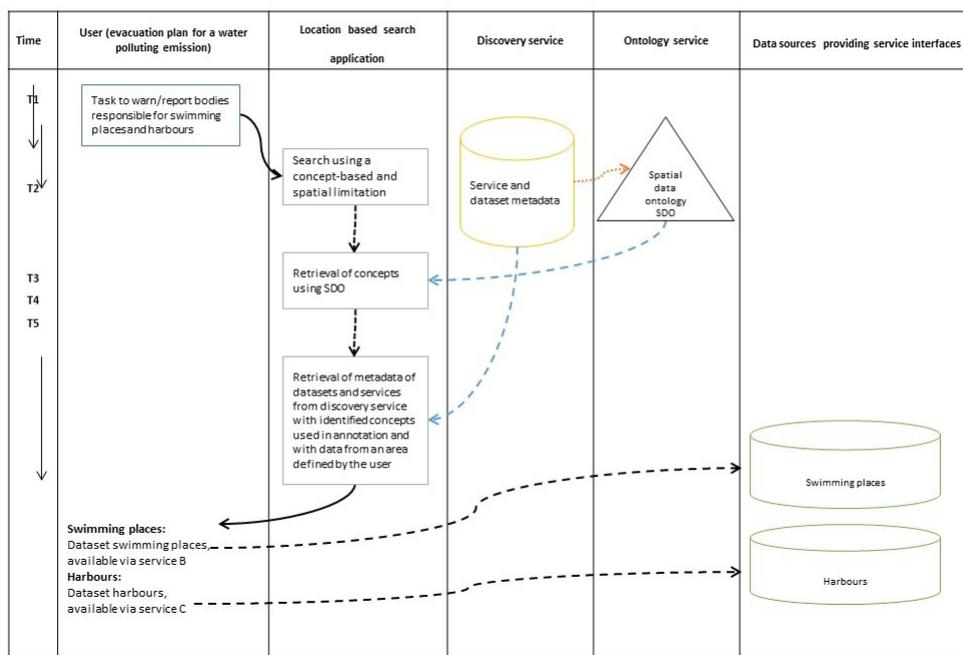
- Application Ontology could correspond to an application ontology generated automatically from a data product schema in accordance with INSPIRE data specifications.
- Domain Ontology would, in practice, be the Finnish Geospatial Domain Ontology located in the Finto service.

In this recommendation, it is recommended that identifiers of concepts corresponding with geographic features are linked to the identifier of documentation concerning the geographic features. This corresponds with level 3 presented above (Section 5.4 below). However, this chapter presents a data combination model in accordance with all of the three levels presented above to illustrate how the recommended procedure simplifies the use of data.

5.1 Use of links at a metadata level

The ontology service indicates the services and data sources which include swimming locations and ports.

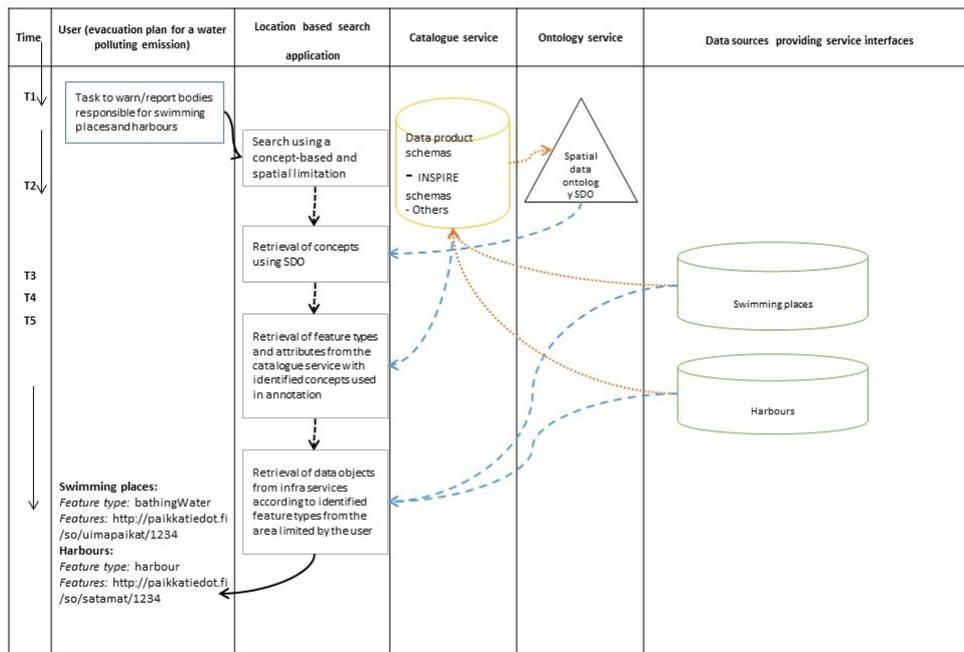
Use of annotations at a metadata level



5.2 Use of links at a schema level

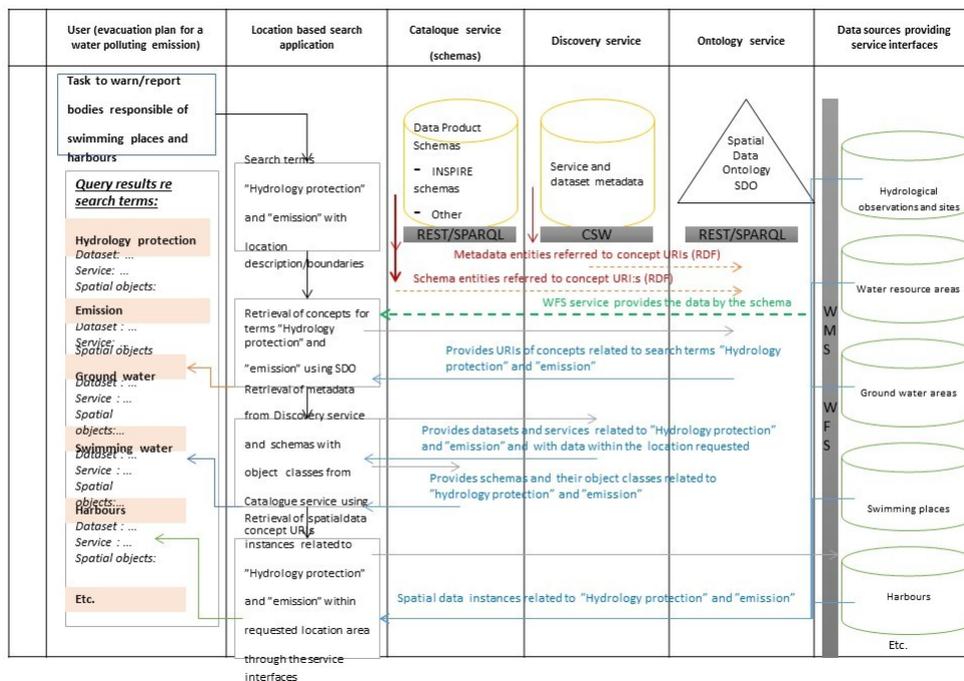
Data product schemas, including swimming locations and ports, are identified from a catalogue service. Schemas can be used in applications, e.g. WFS download services, to retrieve features directly from data sources via interfaces. This can be utilised in terms of open data.

Use of annotations at a schema level



5.3 Use of links at metadata and schema levels

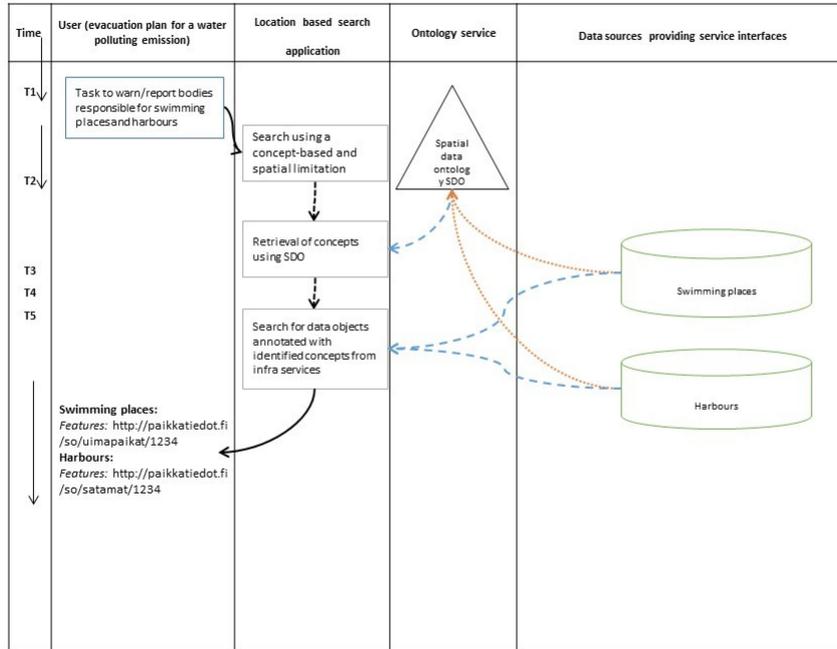
Data features in accordance with the feature types identified from services found via a discovery service are searched for over the area limited by the user (swimming locations and ports). This example presents the exchange of information between services in more detail than in the other examples.



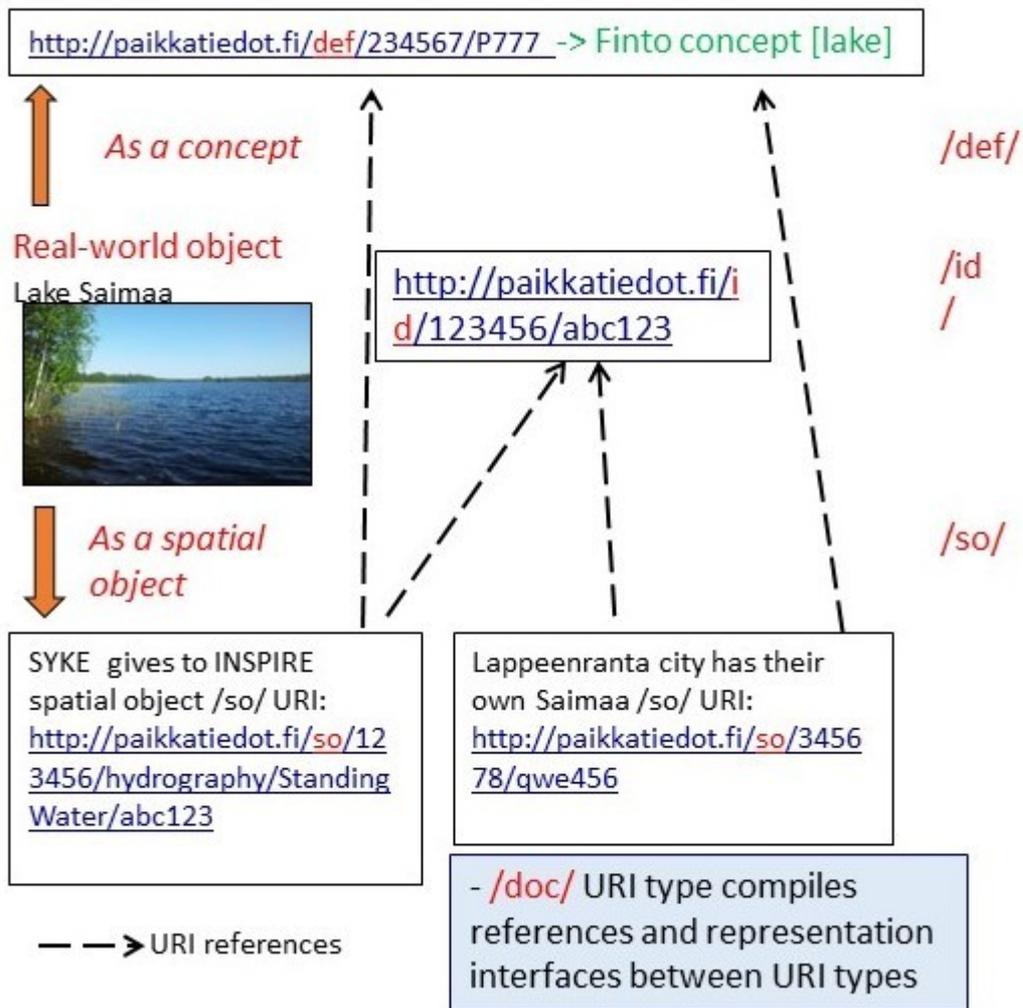
5.4 Use of links at a feature level

Features, i.e. swimming locations and ports in this example, can be searched for by limiting the concept and location areas.

Use of annotations at an object level



The figure below presents how linked data is networked between different data features in accordance with this recommendation.



The figure indicates that the release of URIs

- at a concept level enables that geographic information related to concepts can be searched for using concepts
- for real-world objects enables that information representing a single feature can be combined from different data sources
- Networking URIs to Finto's concept hierarchy allows geographic information to be searched for and discovered through it using Finto searches.