Co-financed by the Connecting Europe Facility of the European Union

# Activity 5: Interoperability framework for Environmental Sensing

**Due date of deliverable:** 31/05/22

**Actual submission date:** 30/06/22

**Version:** Version 1

**Main Authors:** imec: Harm Delva , Brecht Van de Vyvere, Pieter Colpaert , Philip Leroux

Digitaal Vlaanderen: Annelies De Craene, Bert Van Nuffelen, Geraldine Nolf, Eveline Vlassenroot, Geert Thijs, Raf Buyle, Tom Callens

**Abstract** In this deliverable, we will provide an intermediate overview of the output of the different tasks within Activity 5.

**Keywords** NGSI-LD, Data Models, OSLO standardisation, Linked Data Event Streams, Air Quality, Water Quality

# Table of content

Introduction to activity 5 ....... 3
Task 5.1.: Publishing interoperable air quality and water quality data for maximum reuse ....... 6
Introduction to OSLO models: 'Why have these models been designed?' ....... 6
The European Interoperability Framework ....... 7
OSLO, the Flemish instantiation of the EIF ....... 8
How does an OSLO model fit within the NGSI-LD and Smart Data Models ecosystem? ....... 8
What is the process behind the creation of novel OSLO models? ....... 9
Description of the OSLO Environmental models of ODALA ....... 14
OSLO Toolchain ....... 17
Practical guidelines on how these models should be adopted ....... 19
**OSLO Building registry** ....... 19
**Data snippet** ....... 20
**Adding properties** ....... 22
**NGSI-LDify** ....... 24
**rdfs2ngsild.js** ....... 27
**What is the difference between an RDFS object and NGSI-LD object?** ....... 27
**What it does** ....... 27
Task 5.2.: Facilitate the discoverability of environmental datasets via European standards ....... 28
Step 1: Definition of metadata standard ....... 28
Step 2: Implementation in a metadata management system ....... 31
Step 3: Describe the published datasets and developed data services ....... 34
Step 4: Connect dataservices and datasets ....... 42
Step 5: Connect with the data models defined in WP5.1 ....... 43
Step 6: Share the datasets and dataservices throughout National and European (Open) data Portals ....... 45
Task 5.4.: An Open Data interface for NGSI-LD ....... 48
Linked Data Event Streams (LDES) for open data publishing ....... 48
How to map the NGSI-LD interface conform to the LDES principles? ....... 49
How to deploy NGSI-LDES in practice? ....... 51

Task 5.3.: Tooling for NDSI-LDES - Extending comunica with time series and geospatial queries supported by NGSI-LD ........ 52

Future Work ........ 54

Annex I: UML Model OSLO Air & Water ........ 55

# INTRODUCTION TO ACTIVITY 5

Activity 5 of the CEF ODALA project has the overall ambition to elaborate on those components currently needed within the domain of Smart Cities to work with and share data that relates to environmental use case scenarios. The ambition of this activity is not to design an end-to-end framework to do so, but rather focus on those domain specific components (specifications and tools) that are needed on top of and compliant with the overall ODALA architecture, as defined within ODALA Activity 2. This ambition translated to a main focus within this activity on:

1. How to model such environmental data in a uniform way and to provide the tools and guidelines to work with such models in compliance with the CEF ODALA design principles (CEF CB, NGSI-LD Compatibility, etc.).
2. How to make such data available to different (public) shareholders in a resource efficient manner. Again, with a focus on CEF ODALA compliancy i.c.w. providing the necessary tools and guidelines.

4 main tasks were identified in order to successfully complete these activity goals:

- **Task 5.1: Publishing interoperable air quality and water quality data for maximum reuse.**

  Design of sustainable vocabularies and application profiles for Air quality which bridge between existing international standards and NGSI.

- **Task 5.2: Facilitate the discoverability of environmental datasets via European standards.**

  Run OSLO trajectory (Open Standards for Linked Organizations) on metadata for data services and time series by extending DCAT-AP 2.0 and map this OSLO metadata standard to NGSI-LD context semantics into DCAT-AP 2.0.

- **Task 5.3: Extending comunica with time series and geospatial queries supported by NGSI-LD**.

  Enable more flexible client-side querying of open data, allowing cities to put a lower-cost Open Data interface on top of data brokers for sensor data that still is able to answer any query. Comunica is an open source imec component that currently solves graph queries.

- **Task 5.4: An Open Data interface for NGSI-LD**

  Build a Web API that enables the query performance of Comunica to be optimised, while still aiming for a high cacheability. This can be the basis for a Public Open Data interface for sensor data.

The PoC demonstrator of Activity 5 had the ambition to publish air quality and water quality/quantity data from the region of Flanders on the European Data Portal (via the Flemish Open Data Portal) , while being compliant with all the principles and standards as defined within the different ODALA activities.

A high-level overview of this PoC demonstrator is also shown in Figure 1, detailing:

1. Converting Air and Water Quality data from the Flemish region to OSLO data models and ingest this data into an NGSI-LD context broker.
2. Retrieve the data streams from the NGSI-LD context broker and convert them into discoverable LDES data streams.
3. Make these data streams available on the Flemish/European data portal.
4. Develop dedicated client tools to query these LDES streams and extend the larger Comunica framework with specific querying functionality.

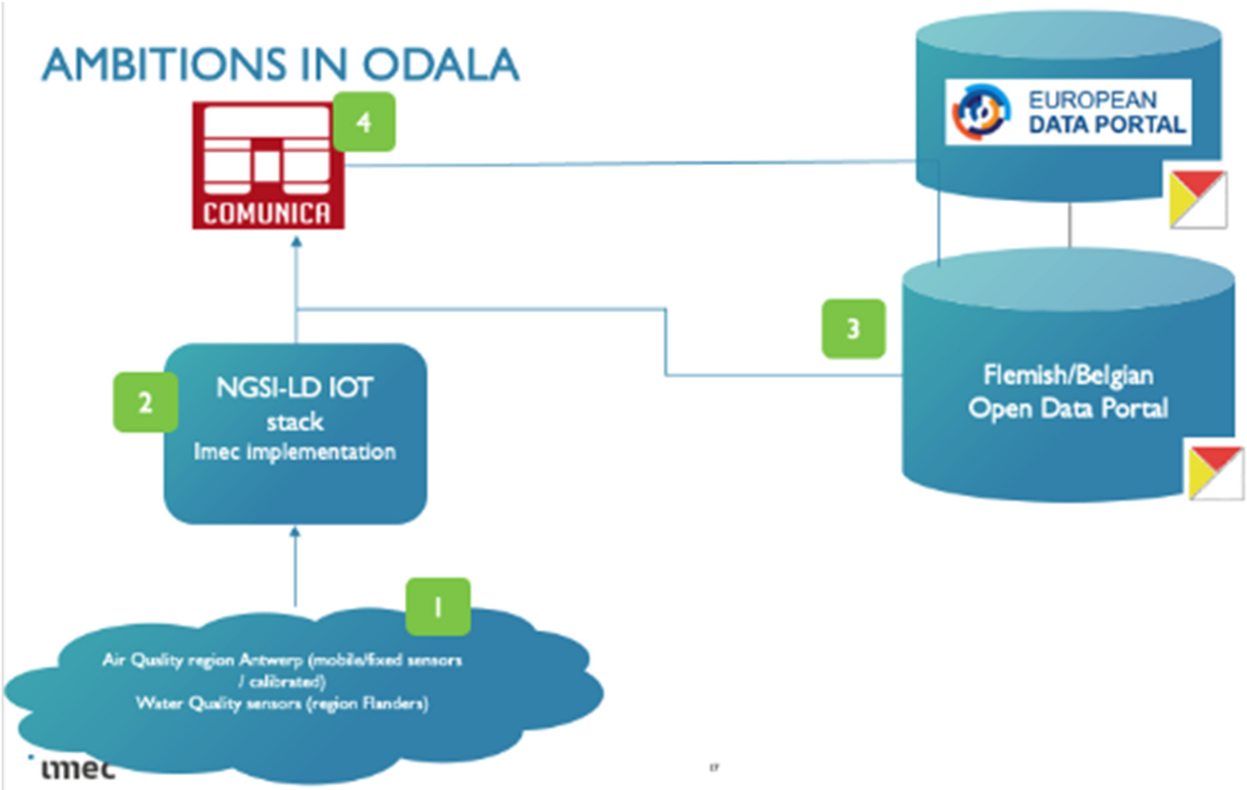



**Figure 1**: High-level overview of the ODALA Activity 5 tasks and PoC demonstrator.

During project execution, novel insights were obtained which positively impacted the approach on how to achieve some of these goals and on how to implement some of the planned tasks. For example, while focussing on Task 5.4, the initial approach was to design the open data interface on top of the NGSI-LD context broker via the Linked Data Fragments principles[1]. However, during project execution, novel insights were obtained and a

[1] https://linkeddatafragments.org/

new set of principles and specifications were developed, i.e. LDES, the Linked Data Events Streams[2]. As a result, also the focus within Task 5.3 shifted towards the development of LDES compliant tools that would enable cities to adopt these principles in combination with the ODALA architecture while complying with the data models and specifications of Tasks 5.1 and 5.2. More details about how the obtained insights led to novel tools and specifications and how we were able to comply with and extend on the original ambitions are described within this deliverable.

The overall aim of this deliverable is to be both an informative guide as well as a hands-on tutorial for smart cities and other environmental stakeholders (e.g water companies) on how they can open up and publish environmental data while being compliant with the emerging smart city standards (cfr. NGS-LD specI, CEF Context Broker, etc.). Furthermore, it should be noted that while the focus within this deliverable is on environmental data, all described principles and processes can be easily mapped towards other (smart city) application scenarios. Finally, it should also be mentioned that the work of CEF ODALA Activity 5 is continued within the CEF project GreenMov.

[2] https://w3id.org/ldes/specification

# TASK 5.1.: PUBLISHING INTEROPERABLE AIR QUALITY AND WATER QUALITY DATA FOR MAXIMUM REUSE

## Introduction to OSLO models: 'Why have these models been designed?'

Open Standards for Linked Organisations - OSLO - laid the basis for an open semantic information standard.

As suggested by the name, the aim of OSLO is to achieve interoperability for the exchange of data in a broad sense by facilitating and recognising semantic and technical data standards.

The objectives of OSLO are to:

- Facilitate semantic and technical standardisation through an open process;
- Maintain existing standards;
- Ensure rules and governance are respected;
- Provide a publication platform ;
- Promote standards; and
- Provide training and support for the adoption of data standards.

The main philosophy behind the OSLO-approach is to ensure the reusability of existing international standards such as W3C, EU ISA Core Vocabularies, INSPIRE… (see Fig 2) and to seek the involvement of a multitude of different stakeholders such as government, industry and academia. The latter is often referred to as the Triple Helix approach.

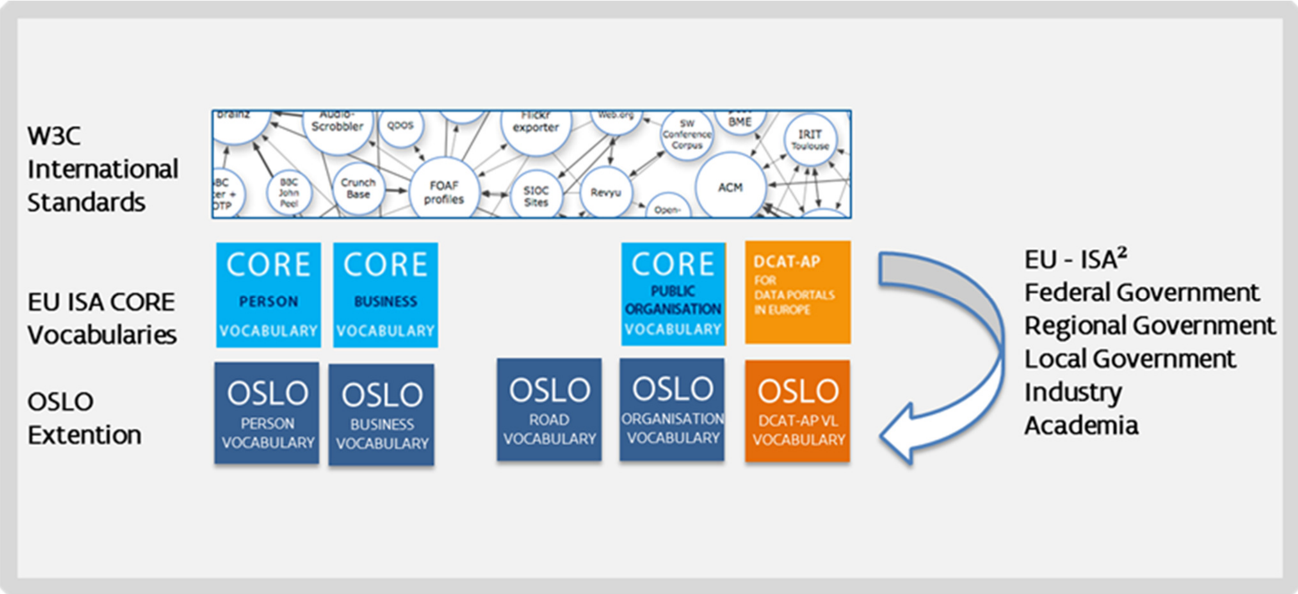



**Figure 2**: OSLO process & methodology: maximum reuse of existing standards

## The European Interoperability Framework

Connecting digital systems with each other is often a non-trivial task hindered by small and larger differences resulting from the design decisions of the involved digital systems. To have smooth data exchange, not solely a technical bridge has to be built, but all data interoperability challenges have to be resolved. The European Commission has recognized that getting a grip on data interoperability is crucial for public governments as they offer digital services throughout many different systems. To aid in streamlining the processes supporting end-to-end digital services within and across member states the European Interoperability Framework (EIF) has been established. The EIF categorizes the data interoperability challenges into 4 main categories: legal, organizational, semantic, and technical interoperability (see Fig.3). It also provides a set of principles that can be used to drive towards a more interoperable digital environment (European Commission, 2017).

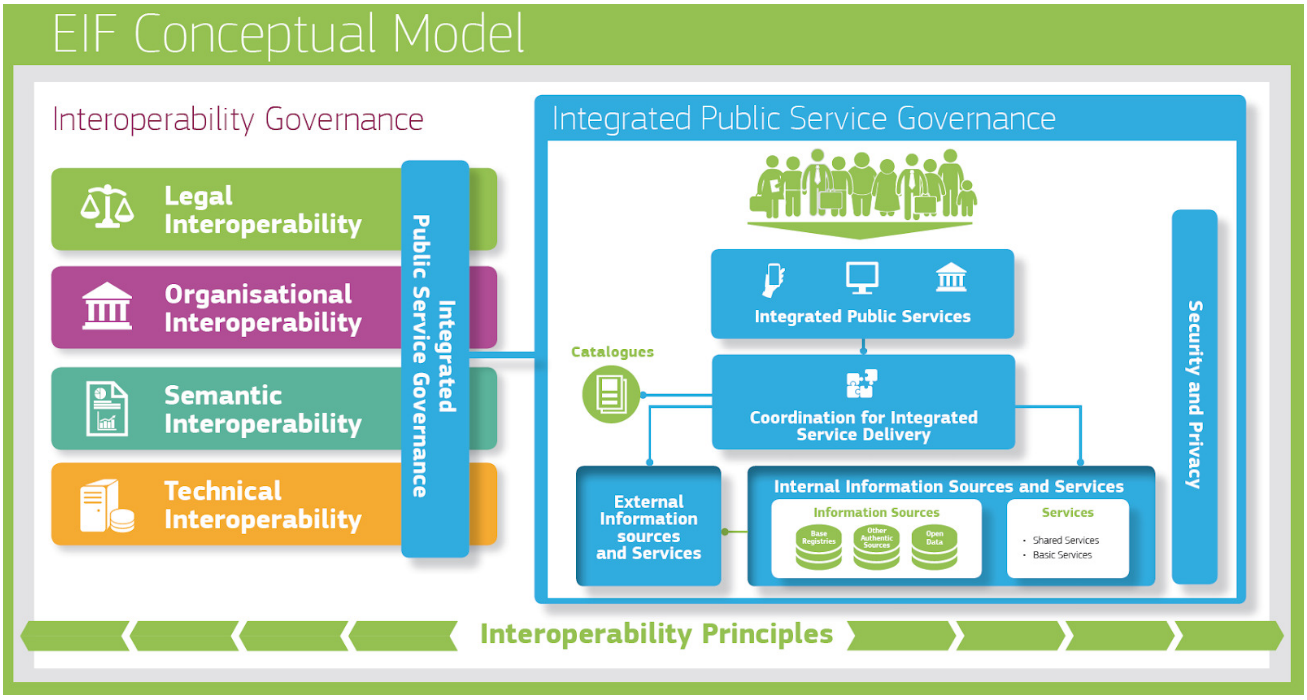



**Figure 3**: Model that categorizes interoperability in 4 groups: legal, organizational, semantic and technical[3]

## OSLO, the Flemish instantiation of the EIF

The Flemish government is committed to unambiguous standards for the exchange of information. The public services to citizens in Flanders are supported by various specialized applications from different software providers. With Open Standards for Linked Organizations (OSLO), the aim is to ensure greater coherence, better comprehensibility, and findability of data and services.

## How does an OSLO model fit within the NGSI-LD and Smart Data Models ecosystem?

The NGSI-LD information model structure provides a core metamodel for property graph capabilities and a cross-domain ontology for common properties. With the Smart Data Models (SDM) governance model, domain-specific ontologies can be contributed in a use case driven setting. By providing among others an example file of the data and a JSON schema, a new data model can be created. Such a data model is created per type of entity contrary to the OSLO initiative. A data model in OSLO describes multiple types of entities in a certain domain (Application Profile) which can be represented with a UML diagram. Reusers of an OSLO data model can then select the classes and properties they need for their implementation. Overall, SDM is more implementation driven using examples per type of entity while OSLO focuses on providing an overview of how entities are

[3] Source: https://ec.europa.eu/isa2/eif_en

semantically linked across domains.

In ODALA and extended within the Greenmov project, we investigated how data following the OSLO principles can flow through an NGSI-LD Context Broker (see Fig. 4). On the input side, we have an OSLO-compliant object composing one or more types of entities, which uses RDF Schema (RDFS) as grounding. First, to be able to ingest the object into the broker, the object needs to use the NGSI-LD core metamodel as grounding. Second, although this would already work, the object needs to be split per type of entity. Otherwise, it would not be possible to manage the context of the deeper nested entities. These two steps have been automatized with a Javascript library[4]. On the output side, a broker is able to publish its entities with an NGSI-LD *normalized* format and a *keyvalues* format. The latter means that the property graph structure introduced by the core metamodel is removed. Using the keyvalues format, the OSLO-based entities are re-aligned with its original OSLO structure.


**Figure 4**: OSLO-compliant data should be able to flow through an NGSI-LD Context Broker

## What is the process behind the creation of novel OSLO models?

The focus of this section takes the process of raising and implementing semantic and technical agreements in the Open Standards for Linked Organisations (OSLO) program into account and is built upon a peer reviewed method[5]. OSLO is an interoperability program in the region of Flanders, which brings together expertise from different business domains and governmental levels, independent of specific thematic use cases. The Flemish Government developed several domain models in line with international standards, including Interoperable Europe[1] (ISA²) and INSPIRE[6] enriched by data extensions to comply with the local (European) context[2]. The formal specification is published at data.vlaanderen.be[7] or purl.eu[8]. The thematic working groups, with over 500 authors from the public sector, private sector and academia, demonstrated that it is possible to raise the interoperability and foster the harmonization of data coming from different use cases..

The applied method to raise interoperability on the technical and semantic level is based on the  principles of

[4] https://github.com/brechtvdv/rdfs2ngsi-ld
[5] https://joinup.ec.europa.eu/collection/interoperable-europe/interoperable-europe
[6] http://inspire.ec.europa.eu/
[7] http://data.vlaanderen.be/ns/
[8] https://purl.eu/

Linked Data[3]. The method includes an implementation framework that describes how to make authoritative data self-describing[3]. The semantic agreements are traceable and aligned to match the different stakeholders: policymakers, domain experts, analysts, and developers. The Resource Description Framework (RDF) can in particular facilitate the semantic agreements and JSON-LD allows developers to work with Linked Data without a high entry barrier[3]. This facilitates the implementation of the semantic agreements across different use cases..

Our peer reviewed[1] approach to raising interoperability combines the process to reach technical and semantic agreements by broad consensus and an end-to-end method based on the principles of Linked Data to maintain the semantic agreements within an operational public sector context. This can be applied in four steps (see Fig.5):

1. **Set up a formal governance** by anchoring the standardisation process at an existing governance body or initiating a new governance body. This step is crucial for the trust of the various stakeholders and supports the adoption of data standards.

2. **Agree on a transparent process** to reach semantic and technical agreements. The process outlines the roles of the different actors and specifies how consensus can be reached among stakeholders. Reference implementations of this process are applied and documented in Flanders[9] and on the Belgian interfederal level[10].

3. **Install an end-to-end method** based on the principles of Linked Data. This implies that all records of decisions, discussions and models are publicly accessible online; the latter is documented using a formal language based on RDF. The method must include an implementation framework that ensures semantic agreements are traceable and aligned to match the different stakeholders: policy makers, domain experts, analysts, and developers. Reference implementations of this process are applied and documented in Flanders[11] and on the Belgian interfederal level[12].

4. **Cocreate data standards**: starting from existing international standards, vocabularies and datasets in the European Data Portal, the semantic agreements are reached in open thematic working groups, consisting of domain experts from the public sector, private sector and academia. These working groups follow the process and method within a formal governance framework.

---

[9] https://data.vlaanderen.be/standaarden/erkende-standaarden/proces-methode-ontwikkeling-standaarden/proces-methode-ontwikkeling.html *(Dutch)*
[10] https://github.com/belgif/review/blob/master/Process/201906-ICEG%20-%20process%20and%20method.docx *(English)*
[11] https://data.vlaanderen.be/standaarden/erkende-standaarden/proces-methode-ontwikkeling-standaarden/proces-methode-ontwikkeling.html *(Dutch)*
[12] https://github.com/belgif/review/blob/master/Process/201906-ICEG%20-%20process%20and%20method.docx *(English)*

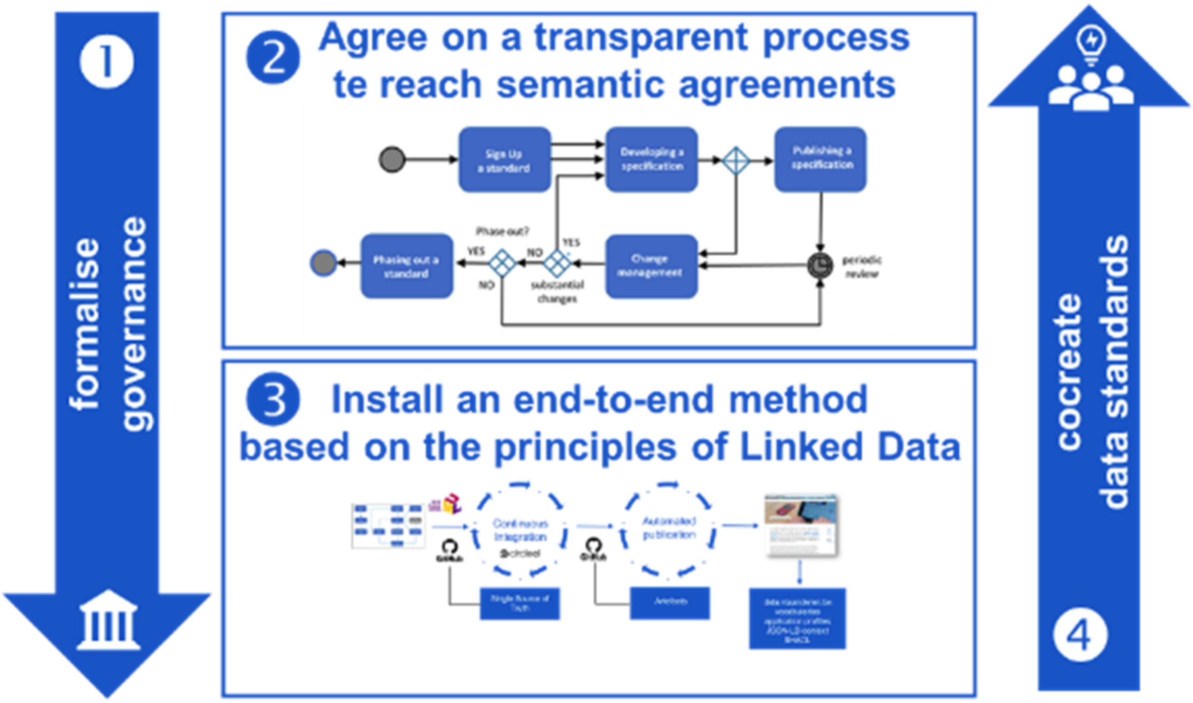



**Figure 5**: Raising interoperability in the public sector

In this specific OSLO Air & Water trajectory, the agreements are laid down for air and water quality in such a way that the information about air and water can be easily exchanged. The intention is to model the data flow semantically and to standardise the structure of the data. This will make it possible to semantically enrich data streams, which will make air and water data easier to find and understand, as well as beter combinable with other data sources. To achieve this, agreements were needed that describe the information in a technology and platform independent manner and at the same time are applicable in the (most) used technological contexts and platforms.

These agreements, or data standards were developed in co-creation during several multi-stakeholder workshops. The timeline of these co-creation workshops is shown in Figure 6. This involved representatives from the departments and agencies of the Flemish government, local authorities, universities but also technical and business experts from the private sector.

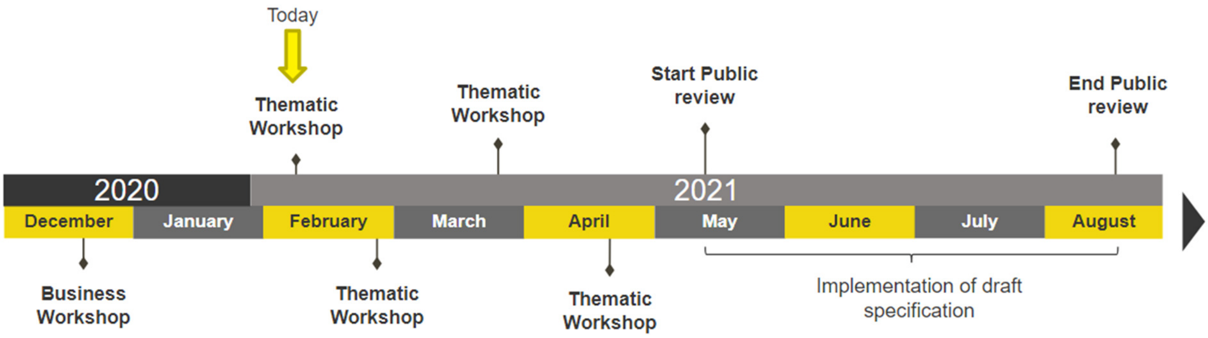



**Figure 6**: Timeline co-creation workshops for OSLO Air & Water

During the business workshop we inventarized the existing (international and European) standards and agreements, identified the information needs and use cases and aligned the scope of the trajectory (what level of detail is needed?). In the first thematic workshop the use cases were discussed and central concepts were defined, starting from the insights of the business workshop. Subsequently, this input was then processed internally and presented in the following workshops where we also had discussions on the definitions of the various concepts and attributes. In this way the support among the participants could grow. In the second workshop the focus was on the core concepts “Observation” / ”OberservableProperty” / ”FeatureOfInterest” / ”SamplingFeature” / ”Result” / ”Sensor” / ”Procedure” / ”Metadata” / ”Platform” and making them more concrete by trying to fit examples from the participants into the model. Here, further modelling and subtypes were discussed such as XXX and the interrelationships between these concepts and attributes were examined. During the third and fourth thematic workshop the data model was further refined, based on the use cases of the participants.

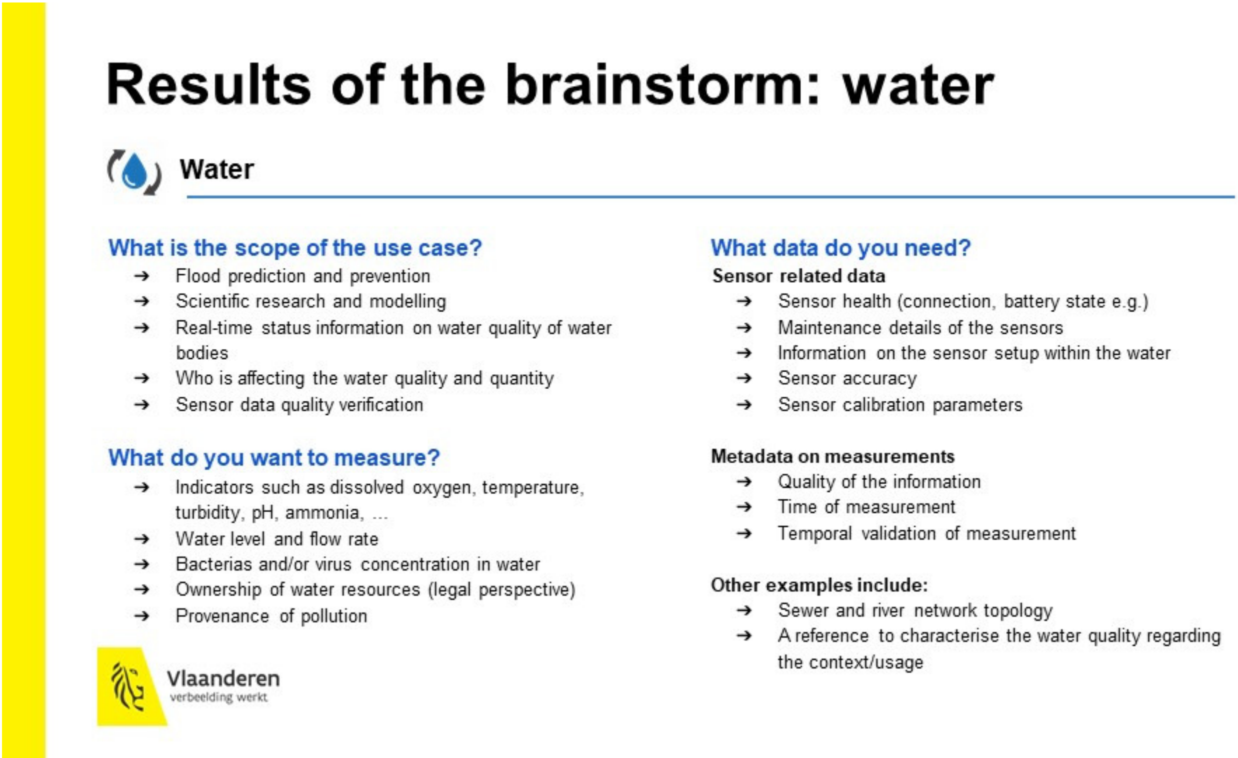



**Figure 7**: Use cases ”water” from first workshop

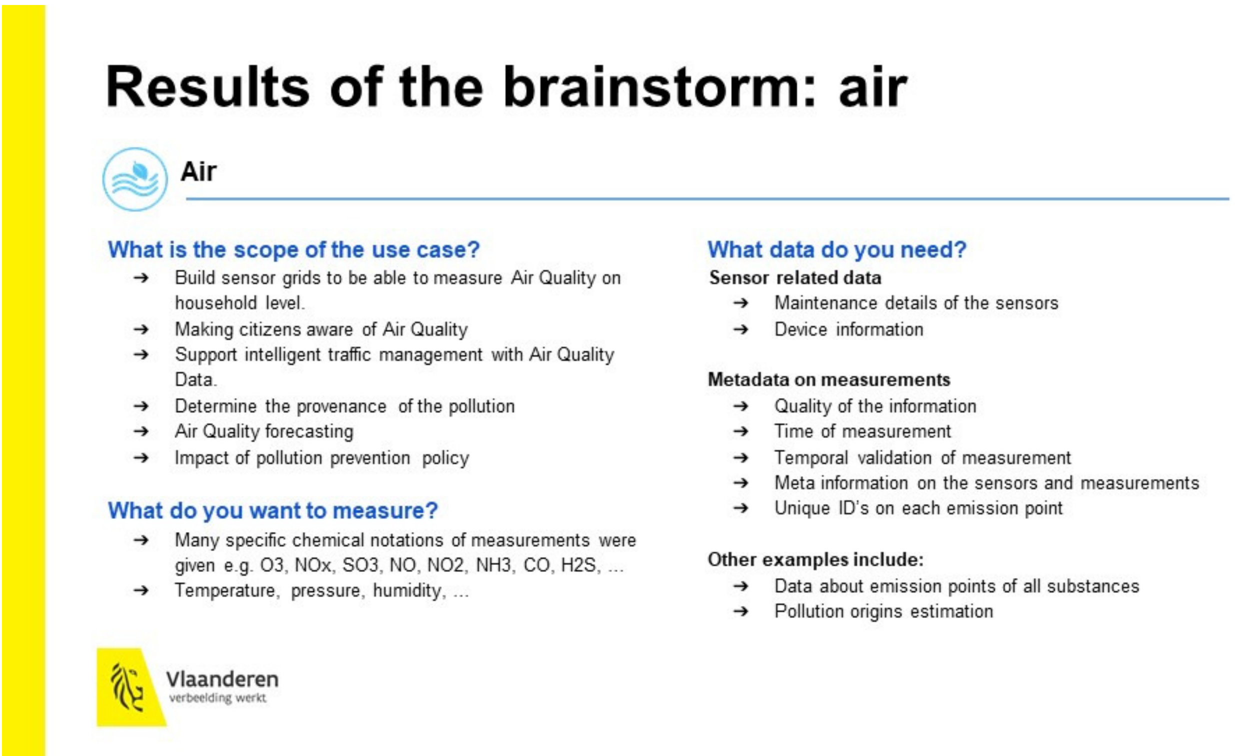



**Figure 8**: Use cases ”air” from first workshop

These workshops resulted in three application profiles:

- ODALA Air & Water - Core (Application Profile)[13]
- ODALA Air & Water - Air Quality (Application Profile)[14]
- ODALA Air & Water - Water Quality (Application Profile)[15]

## Description of the OSLO Environmental models of ODALA

During the workshops, four main building blocks were identified: measurement, measurement device, context of the measurement and administrative data. These are the core of the models, in what follows we have listed the concepts related to these main building blocks.

| **Data about the measurement itself:** | **Data about the measurement device:** |
|---|---|
| • All elements available in ISO O&M<br>• Chromatogram - Liquid chromatography-mass spectrometry<br>• Waterplants<br>• Fauna (e.g. # of blue mussels)<br>• Measure pollutant<br>• Identification of what you measured (which parameter)<br>• Environmental parameters during the measurement<br>• Water pollutants<br>• Type of measurement (chemical, physical, biological status)<br>• Associations between observations<br>• Identification of the object you observe (or proxy object = sample or sampling spot...)<br>• Unit of measurement<br>• Kind of aggregation (1-hour avg., 1 hour maximum, daily maximum) | • Date since device is active<br>• Health status (up/down/faulty)<br>• Calibration<br>• Reliability of the data<br>• All elements available in SOSA/SSN<br>• Admin status (device active or not)<br>• Sensor calibration parameters<br>• Biological observation methods like 'transect'<br>• Time of the last measurement<br>• Location<br>• Station code<br>• Asset management (when it was installed, where, for how long,...)<br>• Time of deployment<br>• Fair chip design<br>• Age - brand - type - material<br>• Manufacturer or model of the device |

[13] https://purl.eu/doc/applicationprofile/AirAndWater/Core
[14] https://purl.eu/doc/applicationprofile/AirAndWater/Air/
[15] https://purl.eu/doc/applicationprofile/AirAndWater/Water

| | |
|---|---|
| • Connected to the water smart-meter | |
| **Data about the context of the measurement:**<br><br>• Type of installation<br>• Measurement unit<br>• date/time<br>• Location<br>• Identification of the river at which the device is installed<br>• Information about the matrix (air, water, sediment)<br>• Coordinates<br>• Type of emission source (industry, …)<br>• Must be connected to the schematic (a plan for the workers that will use it to repair the device/pipe)<br>• Location in the building<br>• Typical values/threshold values<br>• Waterbody and river reference<br>• Must be connected to an object (pipe)<br>• How the observation was made (sensor, procedure)<br>• Distinctions between fixed stations and measurements along transects<br>• Sampling methodology<br>• Lab report<br>• Method of analysis in the lab | **Administrative metadata:**<br><br>• Contact address<br>• Quality assurance steps already performed<br>• License (open, limited, closed)<br>• The shape of the observation object<br>• Date of oldest and newest observation<br>• Owner and maintainer of the device |

The main building blocks that were proposed based on existing international standards are represented in the UML diagram below (see Figure 7). The full diagram can be found in Annex 1.

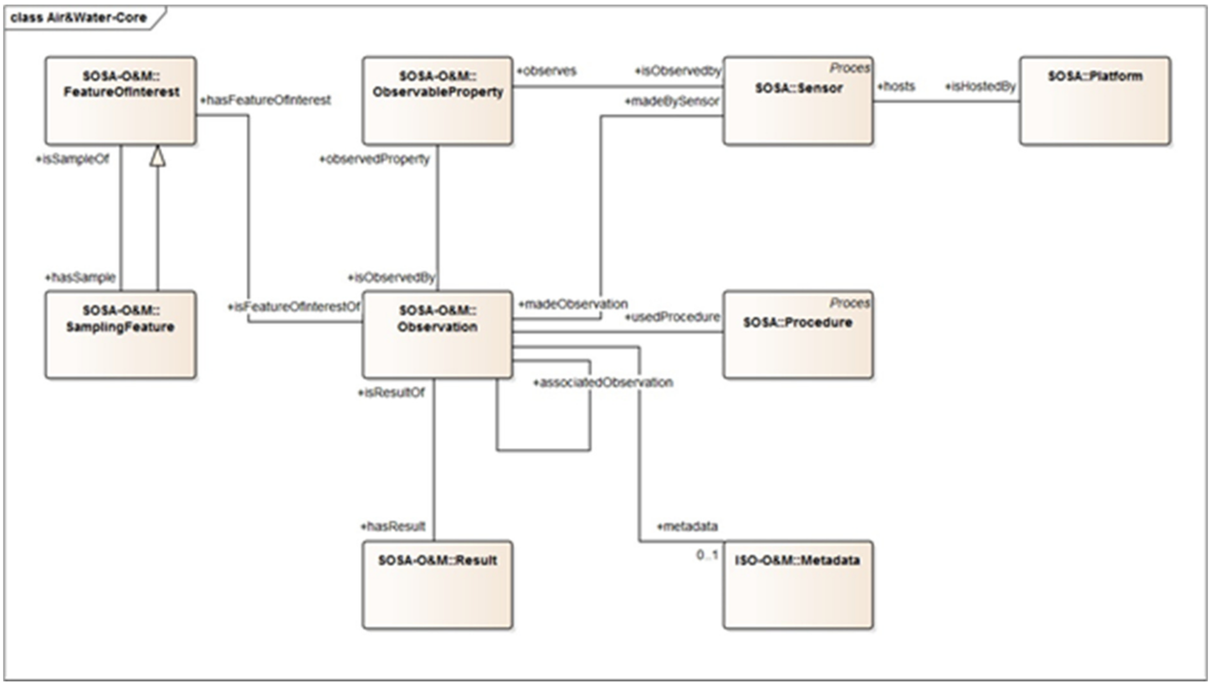



**Figure 7**: UML Model OSLO Air & Water

Links to the vocubalary:

- ODALA Air & Water - Core (Vocabulary)
- ODALA Air & Water - Air Quality (Vocabulary)
- ODALA Air & Water - Water Quality (Vocabulary)

## OSLO Toolchain

Digital Flanders has developed an ecosystem of tools, processes, and governance that allow the designing of data standards in a decentral approach, embedded in global governance. This ecosystem relies on Open Source software and is executed by open continuous integration systems requiring minimal costs for the organization. The setup uses “a separation of concerns” approach: modeling, provenance, and the publication with a maximal of automation to facilitate expectations for humans and machines are decoupled.

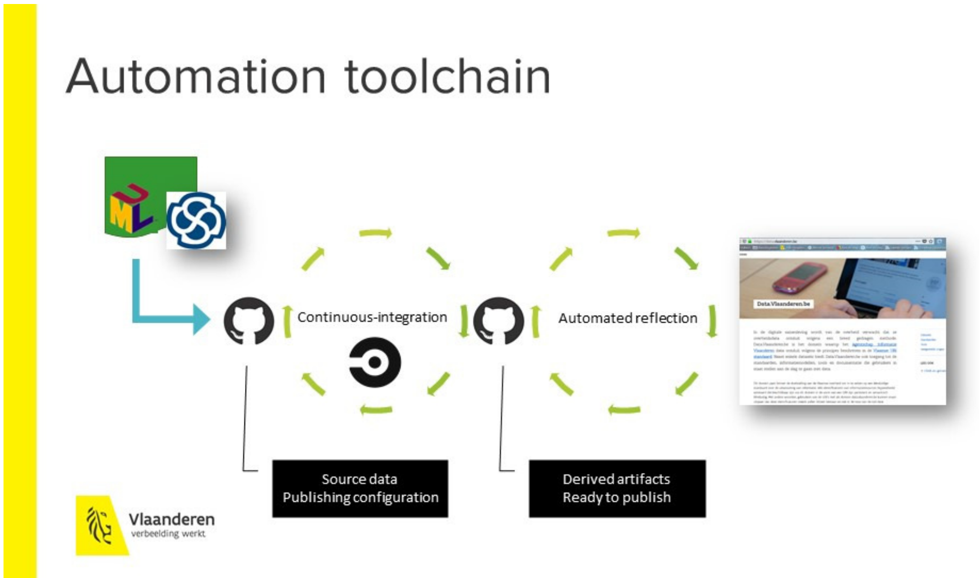



**Figure 10**: Automation toolchain

The core process, usually referred to as OSLO Toolchain, works as follows. First, the semantic data modeling starts with an annotated UML document which is then converted to human machine-readable and machine-processable artifacts of which HTML pages, JSON-LD context files, and SHACL files are examples. This whole flow is described as a circle-ci configuration, which allows continuous integration. Initiating the process from an annotated UML document is a well-motivated design choice: it allows to have a uniform graphical representation with a coherent semantic web representation. The other way around is much more difficult to achieve.

In 2020, a new version of the Toolchain, Toolchain 2.0, was made available and supports *versioning*. The main driver for Toolchain 2.0 was the need for improved decentralization and versioning support without the loss of overall coherency. Today, each data standard has its own repository, which functions as the single source of truth for the data standard and also preserves the lifecycle of the data standard. By registering publication-ready commits of their repository in a central repository, a coherent view is created.

To facilitate the adoption of data standards, the core process is surrounded by tools and other guidelines:

- A URI strategy for persistent identifiers, together with an implementation example on the domain data.vlaanderen.be
- A compliance checking set up to check if data is compliant with an OSLO data standard. This setup uses SHACL and is based on the ISA testbed validator.
- Process provenance through a registry of data standards
- Tooling for the publication of code lists with persistent identifiers
- A playground to create your OSLO compliant payloads.

The methodology and setup have been applied in real-world projects where the data standards, vocabularies for broad reuse, and application profiles for usage in a generic application context, are turned into implementation models using the same Toolchain approach (Buyle et al., 2019).

***Note**:*

*The objective of the OSLO Toolchain is not to become a toolchain for publishing the ultimate data standard, but a toolchain enabling that data standards are published using the same foundation, i.e. the Semantic Web so that differences and correspondences are made visible. It relies on elements in common to most data modeling languages: Classes, Properties, and Instances, identified with a URI.*
*Each of these entities is described with an interpretation in the real world (definitions) and preferred label. Only a few data modeling constructs are supported: subclass, subproperty, domain, and range. No additional “logical inference” statements like rules, OWL formal semantics, negation, ... are used.*

*Exploiting additional formal semantics to enable automated reasoning is possible when these do not violate the real-world interpretation and are not imposed as mandatory when using the data standard. This lightweight formal semantics approach creates a highly reusable base layer for sharing information between systems.*

## Practical guidelines on how these models should be adopted

We created a tutorial[16] that demonstrates the creation of a data snippet that is aligned with an OSLO Application Profile. Then it converts the snippet towards NGSI-LD by introducing the NGSI-LD metadata model (NGSI-LDify). While OSLO defines how different types of entities are interrelated in a certain domain using the Uniform Modelling Language (UML), Smart Data Models provide a tangible description (JSON Schema) per type of entity. Therefore, an OSLO-structured entity should not only be converted to the NGSI-LD metadata model, but also should be split per type of entity. The transformation steps have been automated with a Javascript library[17], called rdfs2ngsi-ld, which can be reused in the pipeline between Context Producers and Context Brokers. In the Greenmov project, an OSLO model is being implemented as a Smart Data Model, which bridges the gap between data publishers that want to be compliant with both OSLO and SDM.

Table 1: tutorial how an OSLO object can be transformed to NGSI-LD

In this tutorial, we will a create a simple data snippet, which represents a building, in accordance with the OSLO Application Profile (AP) "Building registry". In the next tutorial, We will ingest this snippet in a context broker.

### OSLO Building registry

On https://data.vlaanderen.be, all available vocabularies and APs that are governed by the Flemish government are listed:


The Building registry AP (Dutch: Gebouwenregister) is used to describe a building. In the

[16] https://github.com/linkedtimeseries/tutorials.NGSI-LDF/blob/master/tutorials.Data-Snippet.md
[17] https://github.com/brechtvdv/rdfs2ngsi-ld

UML diagram we see that "Building" (Dutch: Gebouw) is the core class and contains a list of attributes, and relations to other classes. For our data snippet, we will only provide two properties: geometry (Dutch: geometrie) and buildingName (Dutch: gebouwNaam):

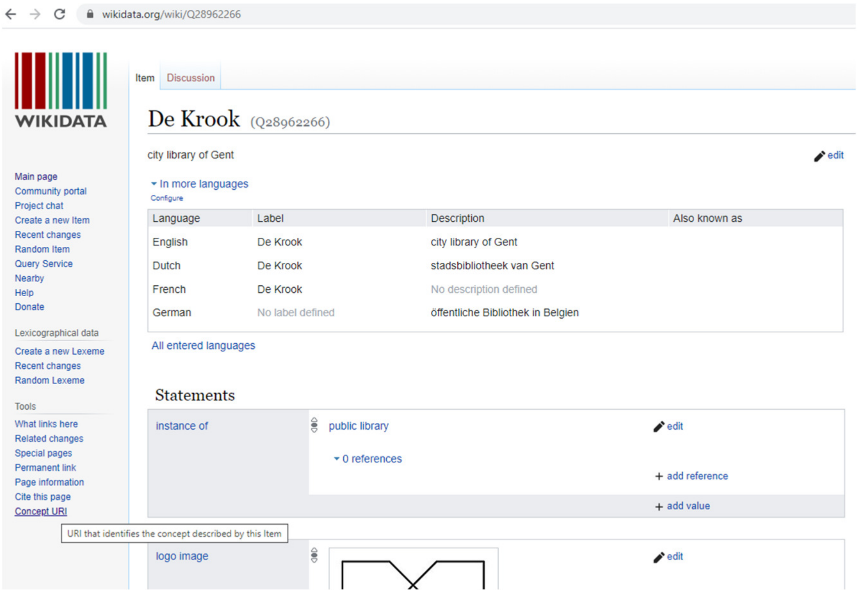

Our data snippet will not be compliant with the AP, because we will not provide the mandatory property "status".

## Data snippet

Every AP has a dedicated JSON-LD context, which we can reuse. You can find the context at the bottom of the AP specification: http://data.vlaanderen.be/context/gebouwenregister.jsonld This is the starting point of our JSON-LD snippet:

```
{
  "@context": "http://data.vlaanderen.be/context/gebouwenregister.jsonld"
}
```

As example, we will publish data about the public library of Ghent "De Krook". First, following the principles of Linked Data, we must reuse an existing identifier of this building. We will use its Wikidata identifier, which can be found by copying the link at "Concept URI":




```
{
  "@context": "http://data.vlaanderen.be/context/gebouwenregister.jsonld",
  "@id": "http://www.wikidata.org/entity/Q28962266"
}
```

To describe that this is a building, we need to look in the OSLO specification for "Building" (Gebouw). When you hover over it, you can see the URI that is linked with this term:

**Gebouw** https://data.vlaanderen.be/ns/gebouw#Gebouw

**Beschrijving**

Een gesloten en/of overdekt, bovengronds of ondergronds bouwwerk, dat dient of bestemd is, ofwel om mensen, dieren en voorwerpen onder te brengen, ofwel om economische goederen te vervaardigen of diensten te verstrekken. Een gebouw verwijst naar gelijk welke structuur die op blijvende wijze op een terrein opgetrokken of gebouwd wordt.

**Eigenschappen**

Voor deze entiteit zijn de volgende eigenschappen gedefinieerd: aantal bovengrondse verdiepingen, aantal ondergrondse verdiepingen, adres, bestaat uit, bouwjaar, gebouwnaam, gebouwstatus, geometrie, laatste renovatie, ligt op.

Make sure to check in the context what the correct term is to use in your snippet:

```
{
 - @context: {
      Adres: "https://data.vlaanderen.be/ns/adres#Adres",
      AdresseerbaarObject: "https://data.vlaanderen.be/ns/adres#AdresseerbaarObject",
      Gebouw: "https://data.vlaanderen.be/ns/gebouw#Gebouw",
      Gebouweenheid: "https://data.vlaanderen.be/ns/gebouw#Gebouweenheid",
      Perceel: "https://data.vlaanderen.be/ns/perceel#Perceel",
    - Gebouw.geometrie: {
          @id: "https://data.vlaanderen.be/ns/gebouw#Gebouw.geometrie",
          @type: "https://data.vlaanderen.be/ns/gebouw#2DGebouwgeometrie"
      },
```

We can now safely add the '@type' property:

```
{
  "@context": "http://data.vlaanderen.be/context/gebouwenregister.jsonld",
  "@id": "http://www.wikidata.org/entity/Q28962266",
  "@type": "Gebouw"
}
```

Of course, we can add a translation in the context if English is preferred:

```
{
  "@context": [ "http://data.vlaanderen.be/context/gebouwenregister.jsonld", {
    "Building": "Gebouw"
  },
  "@id": "http://www.wikidata.org/entity/Q28962266",
  "@type": "Building"
}
```

## Adding properties

For the geometry property, we need to create a 2D-Gebouwgeometrie object. This an intermediary object and also contains a geometry property that will contain a WKT literal for the exact coordinates. Again, it is important to hover in the specification for every property what its URI is and then use this to find out the correct term in the JSON-LD context. For "geometrie" of a building, the URI is "https://data.vlaanderen.be/ns/gebouw#Gebouw.geometrie". In the context, we see that we need to use "Gebouw.geometrie" as term in our snippet. Note that not all terms are mapped in the context, such as "2DGebouwgeomtrie". These terms need to be added manually to the context of the snippet.

```
{
  "@context": ["http://data.vlaanderen.be/context/gebouwenregister.jsonld", {
    "2DGebouwgeometrie": "https://data.vlaanderen.be/ns/gebouw#2DGebouwgeometrie",
    "Geometry": "http://www.w3.org/ns/locn#Geometry"
  }],
  "@id": "http://www.wikidata.org/entity/Q28962266",
  "@type": "Gebouw",
  "Gebouw.geometrie": {
    "@type": "2DGebouwgeometrie",
    "geometrie": {
      "@type": "Geometry",
      "wkt": {
        "@value": "POINT(3.7288391590118404, 51.04909701806207)",
        "@type": "http://www.opengis.net/ont/geosparql#wktLiteral"
      }
    }
  },
  "gebouwnaam": {
    "@value": "De Krook",
    "@lang": "nl"
  }
}
```

We used the WKT playground to determine the WKT location of De Krook. The building name expects a Geografic name (GeografischeNaam). When you hover it, you see that this is a language-tagged string. This is solved in JSON-LD with the `@lang` property.

Finally, use the JSON-LD playground to check if all triples are correctly returned:

JSON-LD Input | Options | Document URL

```
 "@context": ["http://data.vlaanderen.be/context/gebouwenregister.jsonld", {
   "2DGebouwgeometrie": "https://data.vlaanderen.be/ns/gebouw#2DGebouwgeometrie",
   "Geometry": "http://www.w3.org/ns/locn#Geometry"
 }],
 "@id": "http://www.wikidata.org/entity/Q28962266",
 "@type": "Gebouw",
 "Gebouw.geometrie": {
   "@type": "2DGebouwgeometrie",
   "geometrie": {
     "@type": "Geometry",
     "wkt": {
       "@value": "POINT(3.7288391590118404, 51.04909701806207)",
       "@type": "http://www.opengis.net/ont/geosparql#wktLiteral"
     }
   }
 },
 "gebouwnaam": {
   "@value": "De Krook",
   "@lang": "nl"
 }
}
```

Expanded | Compacted | Flattened | Framed | N-Quads | Normalized | Table | Visualized | Signed with RSA | Signed with Bitcoin

| Subject | Predicate | Object | Language | Datatype |
|---|---|---|---|---|
| _:b0 | http://www.w3.org/1999/02/22-rdf-syntax-ns#type | https://data.vlaanderen.be/ns/gebouw#2DGebouwgeometrie | | |
| _:b0 | https://data.vlaanderen.be/ns/gebouw#geometrie | _:b1 | | |
| _:b1 | http://www.w3.org/1999/02/22-rdf-syntax-ns#type | http://www.w3.org/ns/locn#Geometry | | |
| _:b1 | http://www.opengis.net/ont/geosparql#asWKT | POINT(3.7288391590118404, 51.04909701806207) | | http://www.. |
| http://www.wikidata.org/entity/Q28962266 | http://www.w3.org/1999/02/22-rdf-syntax-ns#type | https://data.vlaanderen.be/ns/gebouw#Gebouw | | |
| http://www.wikidata.org/entity/Q28962266 | https://data.vlaanderen.be/ns/gebouw#Gebouw.geometrie | _:b0 | | |
| http://www.wikidata.org/entity/Q28962266 | https://data.vlaanderen.be/ns/gebouw#gebouwnaam | De Krook | | |

## NGSI-LDify

Our data snippet is conform with OSLO. However, to ingest this in an NGSI-LD context broker, we need to apply the NGSI-LD metamodel to our data snippet. The snippet will not have the semantic meaning it is intended by OSLO, but we will be able to process it in a context broker and fix this later on with 1) the keyValues option or 2) NGSI-LDF.

NGSI-LDifying means that we will add an intermediary ngsi:Property object when a triple's object is a Literal, and an intermediary ngsi:Relationship object when the triple's object is another resource. We added the NGSI core context for these terms.

The NGSI-LD data snippet looks as follows:

```
{
  "@context": ["http://data.vlaanderen.be/context/gebouwenregister.jsonld",
    "https://uri.etsi.org/ngsi-ld/v1/ngsi-ld-core-context.jsonld", {
    "2DGebouwgeometrie": "https://data.vlaanderen.be/ns/gebouw#2DGebouwgeometrie",
    "Geometry": "http://www.w3.org/ns/locn#Geometry"
  }],
```

```
  "@id": "http://www.wikidata.org/entity/Q28962266",
  "@type": "Gebouw",
  "Gebouw.geometrie": {
    "@type": "Relationship",
    "object": {
      "@type": "2DGebouwgeometrie",
      "geometrie": {
        "@type": "Relationship",
        "object": {
          "@type": "Geometry",
          "wkt": {
            "@type": "Property",
            "value": {
              "@value": "POINT(3.7288391590118404, 51.04909701806207)",
              "@type": "http://www.opengis.net/ont/geosparql#wktLiteral"
            }
          }
        }
      }
    }
  },
  "gebouwnaam": {
    "@type": "Property",
    "value": {
      "@value": "De Krook",
      "@lang": "nl"
    }
  }
}
```

If we want our building to be retrievable with geographic queries using the temporal query API, we need to add the `ngsi:location` property to the building. This location requires a GeoJSON-LD value, so we also add the GeoJSON-LD context to add context to "Point" and "coordinates":

```
curl --location --request POST 'localhost:9090/ngsi-ld/v1/entities' \
```

```
--header 'Content-Type: application/ld+json' \
--data-raw '{
  "@context": ["http://data.vlaanderen.be/context/gebouwenregister.jsonld",
    "https://uri.etsi.org/ngsi-ld/v1/ngsi-ld-core-context.jsonld",
    "https://geojson.org/geojson-ld/geojson-context.jsonld", {
    "2DGebouwgeometrie": "https://data.vlaanderen.be/ns/gebouw#2DGebouwgeometrie",
    "Geometry": "http://www.w3.org/ns/locn#Geometry"
  }],
  "@id": "http://www.wikidata.org/entity/Q28962266",
  "@type": "https://data.vlaanderen.be/ns/gebouw#Gebouw",
  "Gebouw.geometrie": {
    "@type": "Relationship",
    "object": {
      "@id":                                                "http://localhost:9090/ngsi-
ld/v1/entities/http%3A%2F%2Fwww.wikidata.org%2Fentity%2FQ28962266#2DGebouwgeometrie",
      "@type": "2DGebouwgeometrie",
      "geometrie": {
        "@type": "Relationship",
        "object": {
          "@id":                                            "http://localhost:9090/ngsi-
ld/v1/entities/http%3A%2F%2Fwww.wikidata.org%2Fentity%2FQ28962266#geometrieClass",
          "@type": "Geometry",
          "wkt": {
            "@type": "Property",
            "value": {
              "@value": "POINT(3.7288391590118404, 51.04909701806207)",
              "@type": "http://www.opengis.net/ont/geosparql#wktLiteral"
            }
          }
        }
      }
    }
  },
  "gebouwnaam": {
    "@type": "Property",
    "value": {
      "@value": "De Krook",
      "@language": "nl"
    }
  },
  "location": {
      "type": "GeoProperty",
      "value": {
          "type": "Point",
          "coordinates": [3.7288391590118404, 51.04909701806207]
      }
    }
}'
```

Table 2: Readme of the rdfs2ngsi-ld.js tool to transform OSLO objects into multiple NGSI-LD compliant objects

## rdfs2ngsild.js

JS library to convert RDFS objects (in JSON-LD format) to NGSI-LD compliant JSON-LD objects

## What is the difference between an RDFS object and NGSI-LD object?

An example of an RDFS object:

```
{
  "@context": "https://brechtvdv.github.io/demo-data/OSLO-airAndWater-Core-ap.jsonld",
  "@id": "https://lodi.ilabt.imec.be/odala/data/observations/16584343831",
  "@type": "Observation",
  "Observation.observedProperty": "http://www.wikidata.org/entity/Q48035511",
  "Observation.hasSimpleResult": "8.10 ug/m3"
}
```

Should be transformed to an NGSI-LD compliant object:

```
{
  "@context": [
    "https://uri.etsi.org/ngsi-ld/v1/ngsi-ld-core-context.jsonld",
    "https://brechtvdv.github.io/demo-data/OSLO-airAndWater-Core-ap.jsonld"
  ],
  "@id": "https://lodi.ilabt.imec.be/odala/data/observations/16584343831",
  "@type": "Observation",
  "Observation.observedProperty": {
    "@type": "Relationship",
    "object": "http://www.wikidata.org/entity/Q48035511"
  },
  "Observation.hasSimpleResult": {
    "@type": "Property",
    "value": "8.10 ug/m3"
  }
}
```

## What it does

- Loop over the input object when it's an array
- Copy the JSON-LD context and add it with NGSI-LD's context
- loop over all the properties of the JSON object
- Based on the value of the property:
    - if the value is a string and starts with http or has a @value key, add as NGSI-LD Property
    - else add as NGSI-LD relationship
        - create an identifier if not provided, based on the subject id, relation and index

# TASK 5.2.: FACILITATE THE DISCOVERABILITY OF ENVIRONMENTAL DATASETS VIA EUROPEAN STANDARDS

In addition to the alignment of water and air quality data - described in previous section task 5.1, alignment on the descriptions of the dataset and data services distributing the water and air quality data is an important step to increase the discoverability and reuse of this data. In the past decades the discoverability and reuse of data has been stimulated by initiatives on legislation, metadata standards and platforms. Important initiatives to mention are INSPIRE, ISO, PSI, Open Data and FAIR. The main technical activity driving this objective is the design of metadata standards for describing datasets and data services. Where INSPIRE for the geospatial community has adopted in the early 2000's the ISO standards on metadata, the Open Data community adopted a W3C standard DCAT. Over the past years both have grown towards each other. In particular DCAT has evolved/is evolving towards capturing the needs and use cases that drove the design of the ISO standards. In order to make metadata descriptions fostering reuse it is important that similar situations are documented in a similar way. Metadata standards offer a framework for structuring information, allowing the to-be-documented entities to be described in various ways. Therefore it is important that technical communities or thematic ecosystems agree on how to encode or document the entities. In this section the result of discussions on describing air and water quality datasets in the DCAT standard are presented. This agreement is part of a larger discussion happening today in the community around the nature of datasets and data services. And thus the presented results not only contribute to the direct objectives of ODALA, but also have sustainable impact.

In the next paragraphs the reader can follow how the process that was taken to facilitate the discoverability was made, step by step. Steps 1 and 2 describe the groundwork namely defining the metadata standard as well the building of an implementation. The next 3 steps show a data provider in action, highlighting the key decision making points data providers must consider when publishing Linked Data Event Streams. We conclude the description of this task with showing how the work of previous steps contributes to discoverability, and an outlook of future work.

## Step 1: Definition of metadata standard

This section aims to represent the OSLO (Open Standards for Linked Organisations) trajectory which delivers a canonical data model and a process to arrive at an agreement for facilitating better semantic interoperability.

The first activity in the task WP5.1 is the creation of a data standard for describing a catalogue of datasets and dataservices. It should be possible to describe air and water quality datasets, and the various ways the data can be accessed throughout file downloads or online (data)services. In the context of ODALA particular attention will be given to Linked Data Event Stream service descriptions.

To achieve this objective an OSLO trajectory for metadata has been initiated. The setup of this trajectory is represented in Figure 8 below.

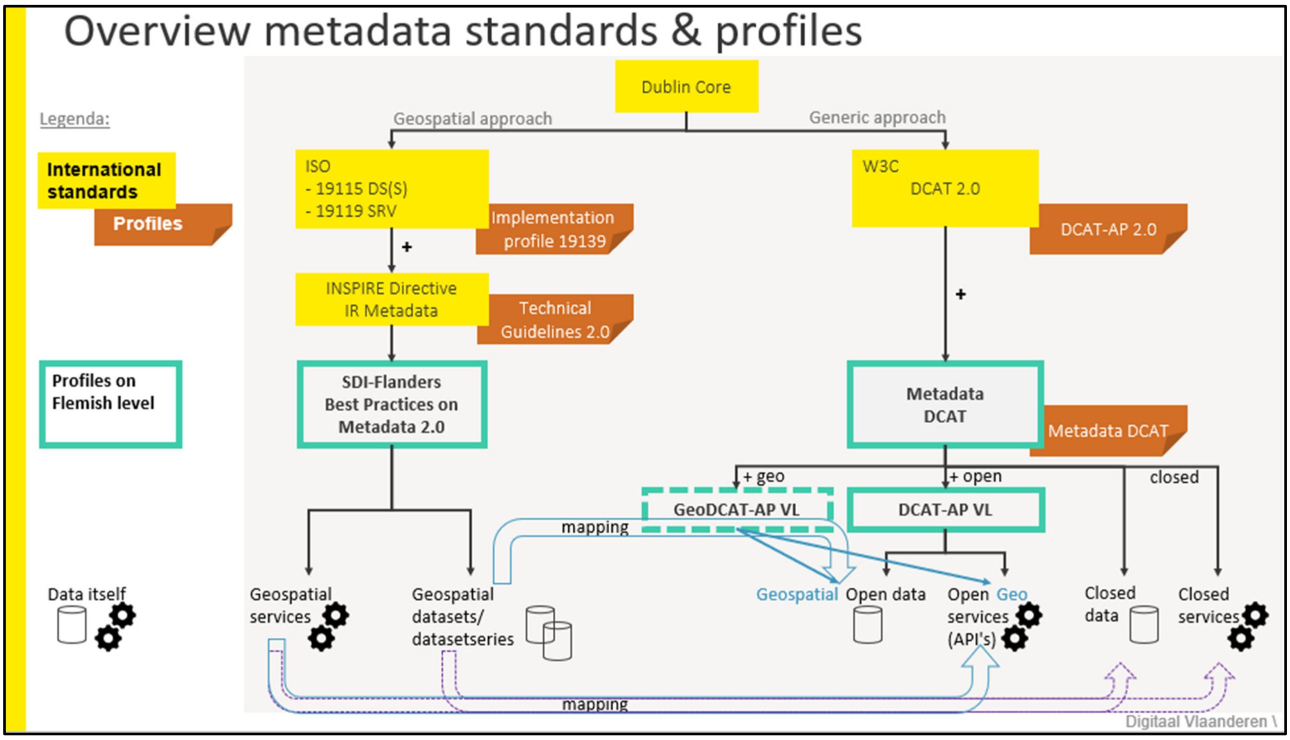



***Figure 8**: Overview metadata standards & profiles*

The goal is not to create a novel independent data standard, but to create a profile for existing metadata standards enforcing existing infrastructures for the discovery of datasets and data services. The yellow blocks in the figure 8 show the international agreements which form the foundations of the work. The interconnection between the data standards for the geospatial domain are shown on the left of the figure. On the right, the domain agnostic track is shown. It is in this track that the OSLO trajectory for metadata takes place. Both DCAT and its European profile for Open Data DCAT-AP are very open standards: it is technically possible to encode the same dataset and data services in various ways. The OSLO trajectory for metadata provides additional agreements and more concrete guidelines to aid the data providers in creating the metadata description in a more harmonious way.

The OSLO trajectory has been divided in subsequent phases. The **first phase** is the creation of a more concrete umbrella DCAT profile, called Metadata DCAT, that is more elaborated than the international standards. Key concerns are (a) the clarification of the entities Dataset, Distribution and Dataservice, (b) strengthen the correctness of the aggregation of catalogues and (c) identify the key properties and their usage.

In a **second phase** the Metadata DCAT profile is further profiled to fit the Open Data community, in a profile called DCAT-AP VL. **Thirdly** a profile for the geospatial domain is designed, GeoDCAT-AP VL. This last is strongly connected with the geospatial metadata standards in the context of INSPIRE, based on the ISO metadata standards. It should be possible to automatically convert an INSPIRE/ISO metadata description into GeoDCAT-AP VL. These agreements were developed in co-creation during several multi-stakeholder workshops starting from July 2020. Representatives from the departments and agencies of the Flemish government, the federal and Dutch government, local authorities, universities but also technical and business experts from the private sector were involved. The workshops for phase 1 resulted in a vocabulary[18] an application profile[19] both ratified by the 'Steering Committee of Flemish Information and ICT-policy', which is empowered by a decree[20], on the 22nd of April 2021. The second phase is currently a public review, the status right before ratification. Phase 3 has been formally initiated and the public consultation has been announced for Q4 of 2021.

Despite these activities happening under the formal governance 'Steering Committee of Flemish Information and ICT-policy' which is a local Flemish, Belgium activity, the results are disseminated towards the international community. Digitaal Vlaanderen is a W3C member (maintainer of DCAT) and it is an active contributor to the EU ISA action SEMIC (maintainer of DCAT-AP and GeoDCAT-AP); also some participating experts are active contributors to these standards. This enables and ensures that agreements are shared and get an uptake in or are supported by the international standards.

In **phase 1, the OSLO Metadata DCAT trajectory** (ran from 20/7/2020 till 22/04/2021), the agreements are laid down for metadata of personal services in such a way that they can be easily merged into the central catalog. To achieve this, agreements were needed that describe the data services in a technology and platform independent manner and at the same time are applicable in the (most) used technological contexts and platforms. On the basis of international standards (DCAT-AP v2.0) we are building the vocabulary which will be an extension of the OSLO standard DCAT-AP with the possibility of also describing APIs. These agreements, or data standards, were developed in co-creation during several multi-stakeholder workshops. This involved representatives from the departments and agencies of the Flemish government, local authorities, universities but also technical and business experts from the private sector.

[18] https://data.vlaanderen.be/ns/metadata-dcat/
[19] https://data.vlaanderen.be/doc/applicatieprofiel/metadata-dcat/
[20] http://docs.vlaamsparlement.be/pfile?id=1213278

During the first business workshop we inventarized the existing standards and agreements, identified the information needs and use cases and aligned the scope of the trajectory (what level of detail is needed?). In the first thematic workshop the use cases were discussed and central concepts were defined, starting from the insights of the business workshop. Subsequently, this input was then processed internally and presented in the following workshops where we also had discussions on the definitions of the various concepts and attributes. In this way the support among the participants could grow. In the second workshop the focus was on the core concepts “Dataset”/”Dataservice”/”Distribution” and making them more concrete by trying to fit examples from the participants into the model. Here, attributes such as “Beschrijving” (=Description) and “Landingspagina” (=LandingPage) were discussed and the interrelationships between these concepts and attributes were examined. During the third workshop the data model was further refined, based on the use cases of the participants.

### Step 2: Implementation in a metadata management system

Simultaneously with step 1, the standard “OSLO Metadata DCAT” was implemented in the reference implementation as part of GeoNetwork 3.8. It enabled us to test the eligibility of the agreements at an early stage.

The reference implementation[21] is based on a pre-existing GeoNetwork DCAT-AP 2.0 plug-in, which has been adapted to fit the agreements of the data standard Metadata DCAT. The work consists of creating suited templates for the online editor to enable data providers to enter the datasets and data services, appropriate validation rules and help texts.

This first implementation of the OSLO Metadata DCAT was realized through the development of templates, namely to help data providers describe the metadata of their environmental (open) datasets and their Linked Data Event Stream data services. In Figures 9 and 10 the templates are demonstrated that help editors to document their datasets or data services.

[21] URL to the reference implementation: https://geonetwork-oslo-demo.gim.be/geonetwork/ based on https://github.com/GIM-be. This implementation will be pull-requested to the core-implementation of GeoNetwork: https://github.com/metadata101/dcat2.

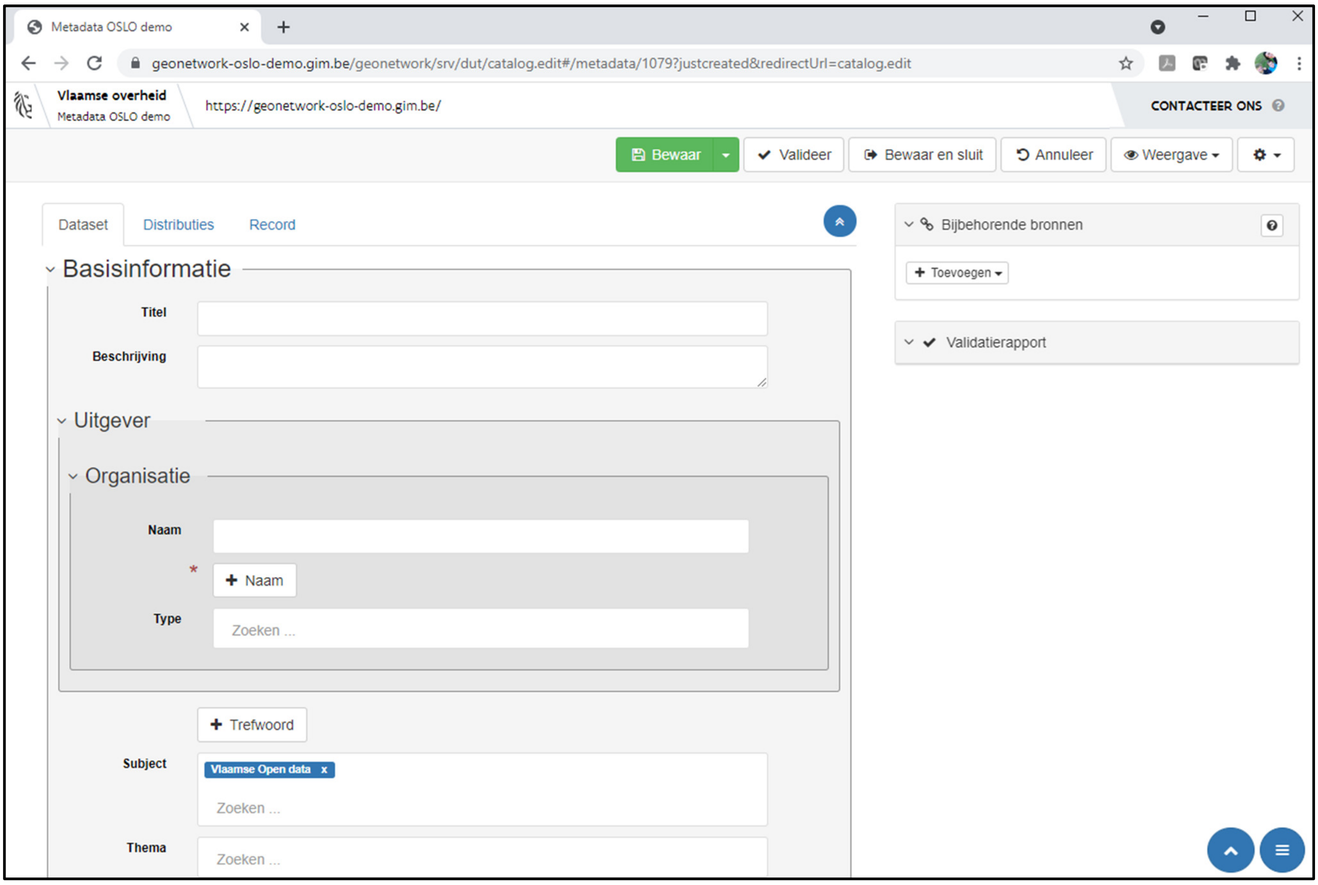



***Figure 9**: An example of the dataset template in the reference implementation*

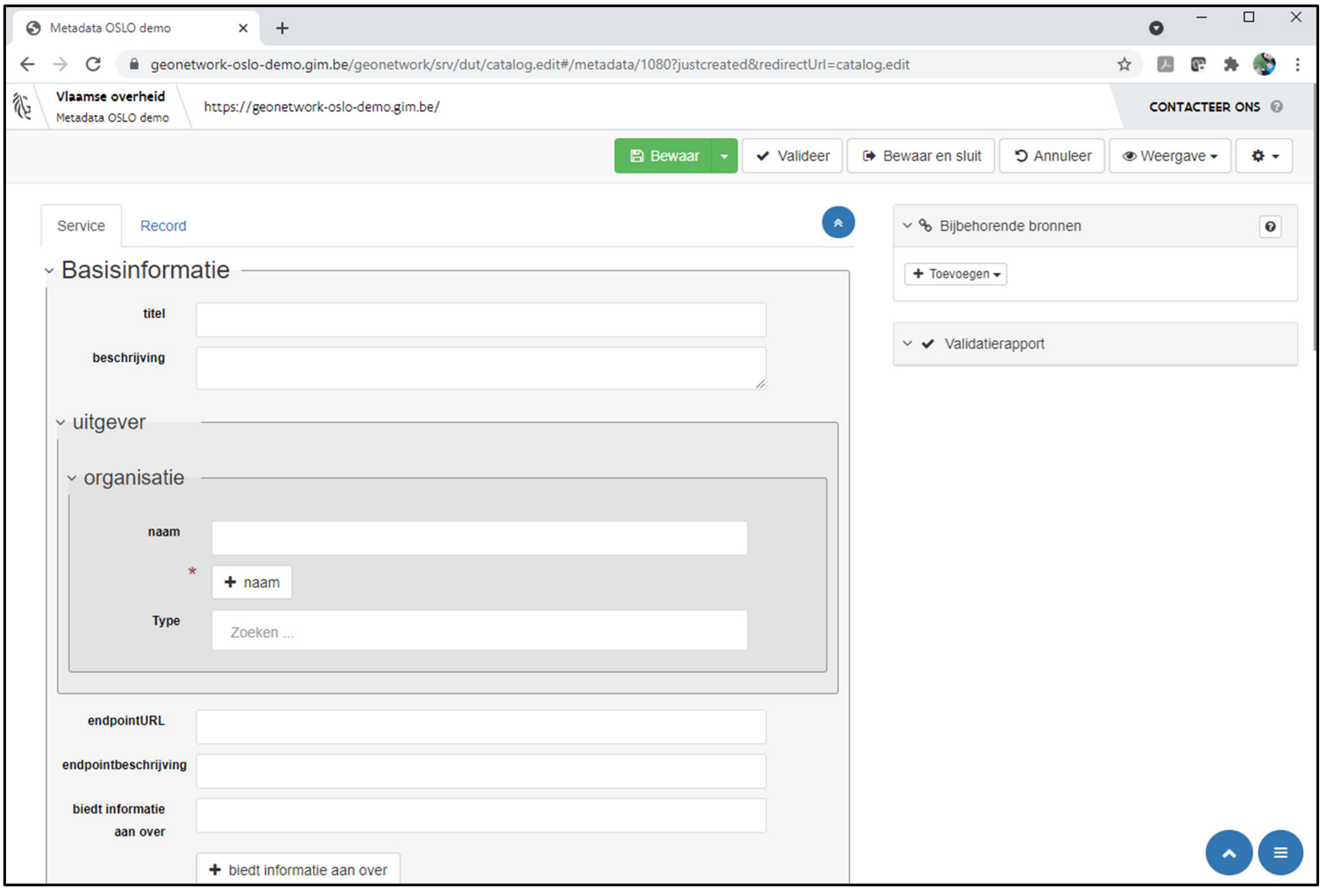



***Figure 10:*** *An example of the dataservice template in the reference implementation*

Next to the templates that were developed, also help texts and validation rules are implemented in the reference implementation. These help texts and validation rules (see Figure 11) make it possible that data providers are guided in correctly handling their editor work to document the datasets and data services.

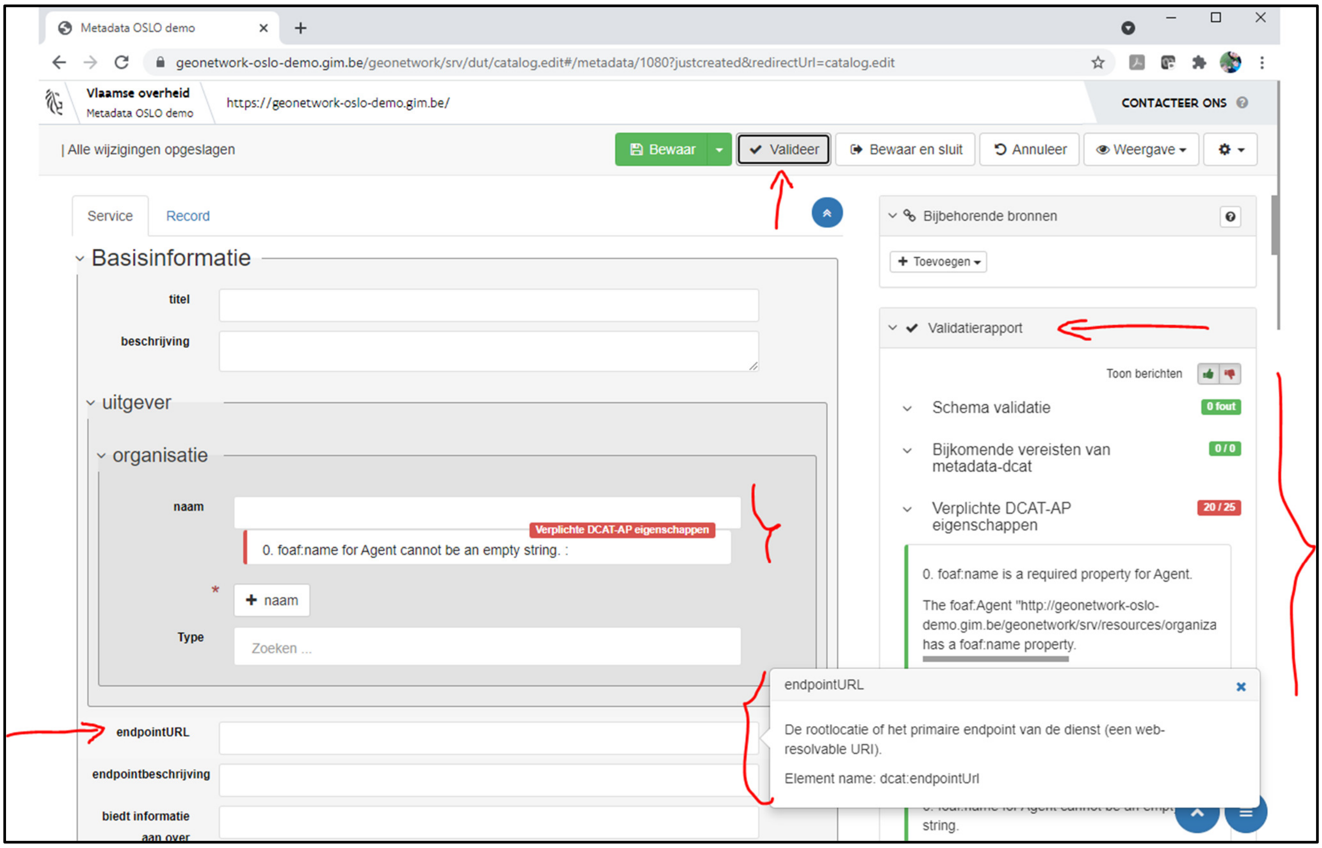



***Figure 11**: An example of the dataset template in the reference implementation, with hover over help textes and validation guidelines in the template and on the validation panel (right hand sided)*

## Step 3: Describe the published datasets and developed data services

This step and subsequent steps 4 and 5 document the activity a data provider has to perform to describe air quality or water quality Linked Data Event Streams as a combination of datasets and data services. Figure 12 shows the UML diagram depicting the Metadata DCAT agreements. Datasets are collections of data. They are conceptual entities. The actual downloadable data can be retrieved via the information described in the associated Distributions. Data Services are a collection of operations that provides access to one or more datasets or data processing functions. The distinction between data services and distributions, both entities with the objective to provide access to the actual data, has been a key discussion throughout the design of the standard Metadata DCAT. The introduction of the entity dataservice in DCAT 2.0 is recent and represents a major shift of mind with respect to the practice of DCAT 1.0. To our knowledge, the OSLO trajectory has been among the very first implementations of this new perspective. The use case provided by ODALA was input in this discussion and acted as a test case to verify whether the usage guidelines can be applied.

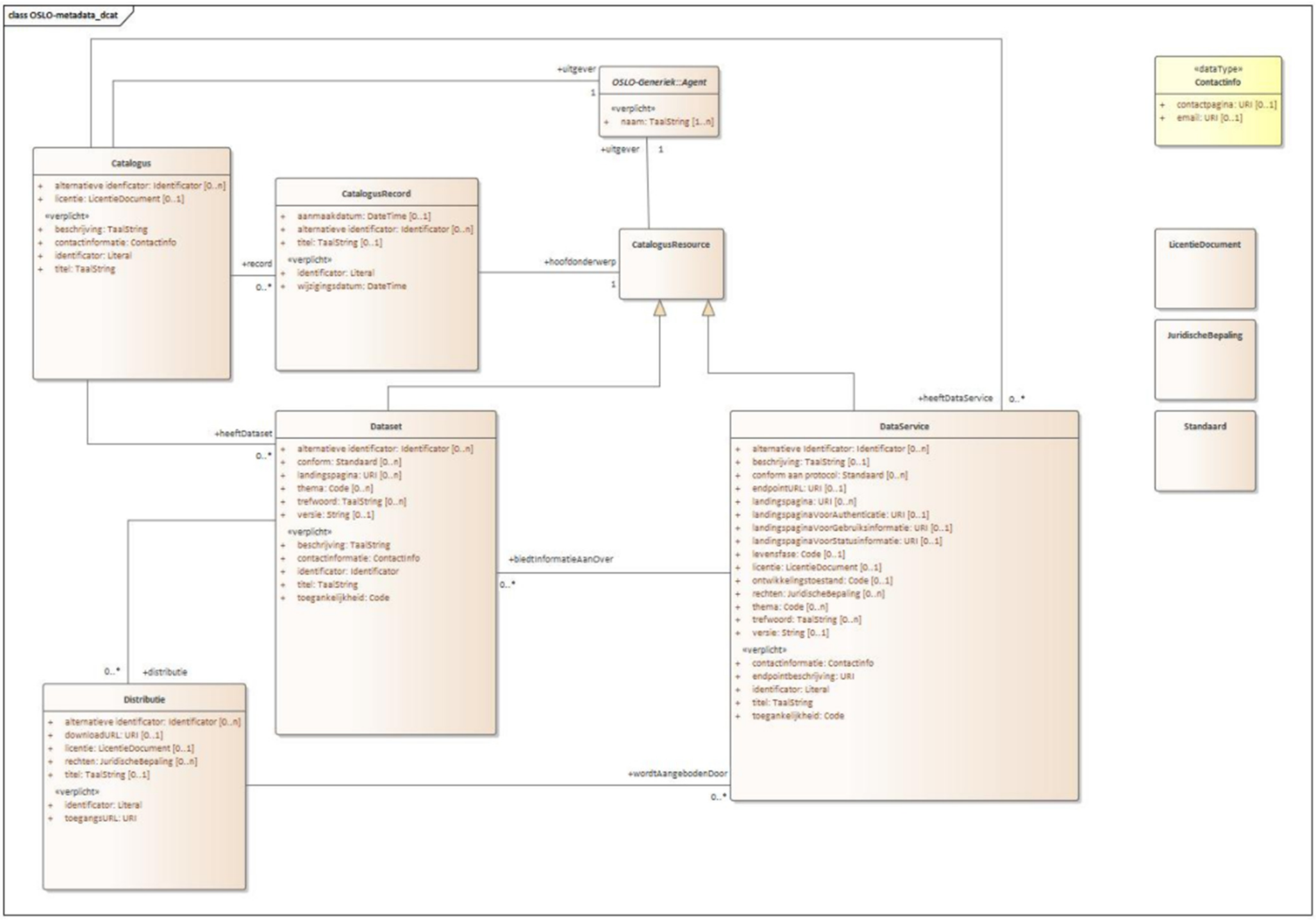



***Figure 12**: An overview of the Metadata DCAT application profile, where the Catalogue handles Datasets and/or Dataservices since DCAT 2.0*

In short the guidelines are:

- Distributions are entities that cannot exist on their own. They are always connected with a Dataset. Because they represent a downloadable representation of the dataset. Via a Distribution, reusers get the ability to obtain all data that the associated Dataset circumscribes in a simple, easy way. The reuser cannot change the obtained content, it has been fixed by the publisher. Distributions are mainly supporting the reuse use case, where a reuser collects data in its own environment before processing it.
- Dataservices are anything else. Their intent is to provide reusers an intelligent, smart access to the data of the dataset. A dataservice aims to provide a more active interaction with data reusers. Not only by providing the most recent data, but also by providing optimized support for frequent access patterns.

Depending on the access pattern one wants to stress, the encoding as distribution or dataservice can be felt more appropriate. For Linked Data Event Streams the guidelines are depicted in Figure 13.

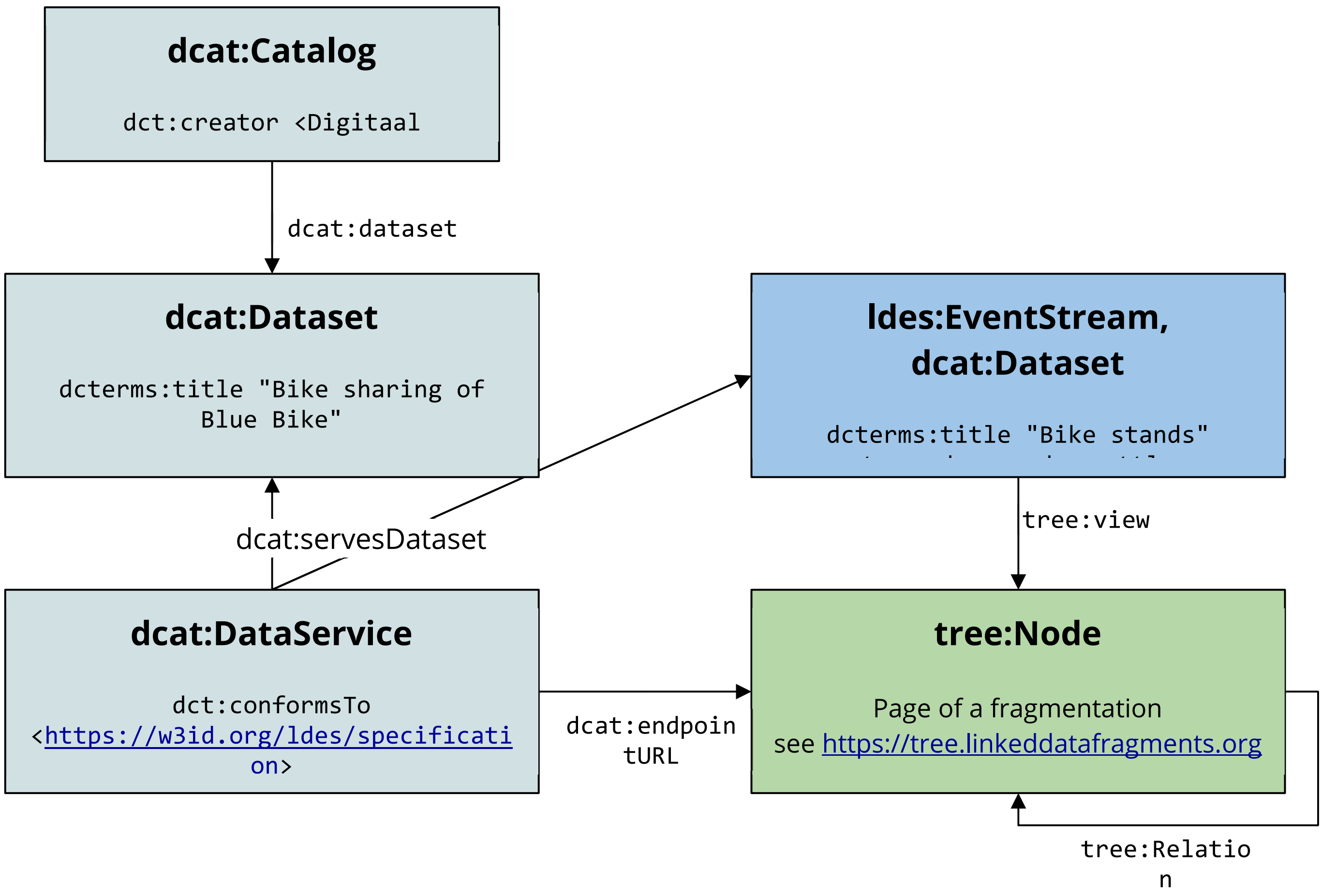



***Figure 13**: Guidelines for Linked Data Event Streams*

A LDES dataservice can serve one or more (LDES) datasets. These datasets provide a LDES tree view on the events that change the data. Therefore, are these datasets the most fine grained description of the data provided by the LDES dataservice. The figure also depicts a broader closely related dataset description (on the left hand side). The broader dataset and the (LDES) dataset are closely related, but differ in the kind of data they publish. This is one of the area's where the distinction between dataset and distribution become vague. Because in some contexts the LDES dataset could be considered as a distribution of the broader one. Unless the international community (to which we also proposed the above approach) imposes another guideline, the above agreement will be applied.

This setup is a balance between the human discoverability and the machine processability of the metadata. A human is capable of finding datasets that are the result of a LDES publishing activity, or to find the LDES dataservices that are offering data. A machine can discover dataservices (dcat:servesDataset) that offer data about some topic via the LDES dataset (dct:conformsTo). And then connect with the retrieved dataservices to the endpoint (dcat:endpointURL). It is assumed that the LDES dataservice maintainers ensure that the metadata descriptions are in sync with the LDES metadata descriptions. In this step we illustrate the metadata description for datasets and dataservices independently. In subsequent steps we highlight the relationships between them and the schema of the data.

Below are concrete examples of the initial editorial work data providers have to do, namely describing the LDES dataservice and the LDES datasets.

First we illustrate the dataset metadata descriptions for air quality and water quality data.

- Dataset Water Quality: https://geonetwork-oslo-demo.gim.be/geonetwork/srv/dut/catalog.search#/metadata/a6df529b-fa45-417f-b923-40784379ef06

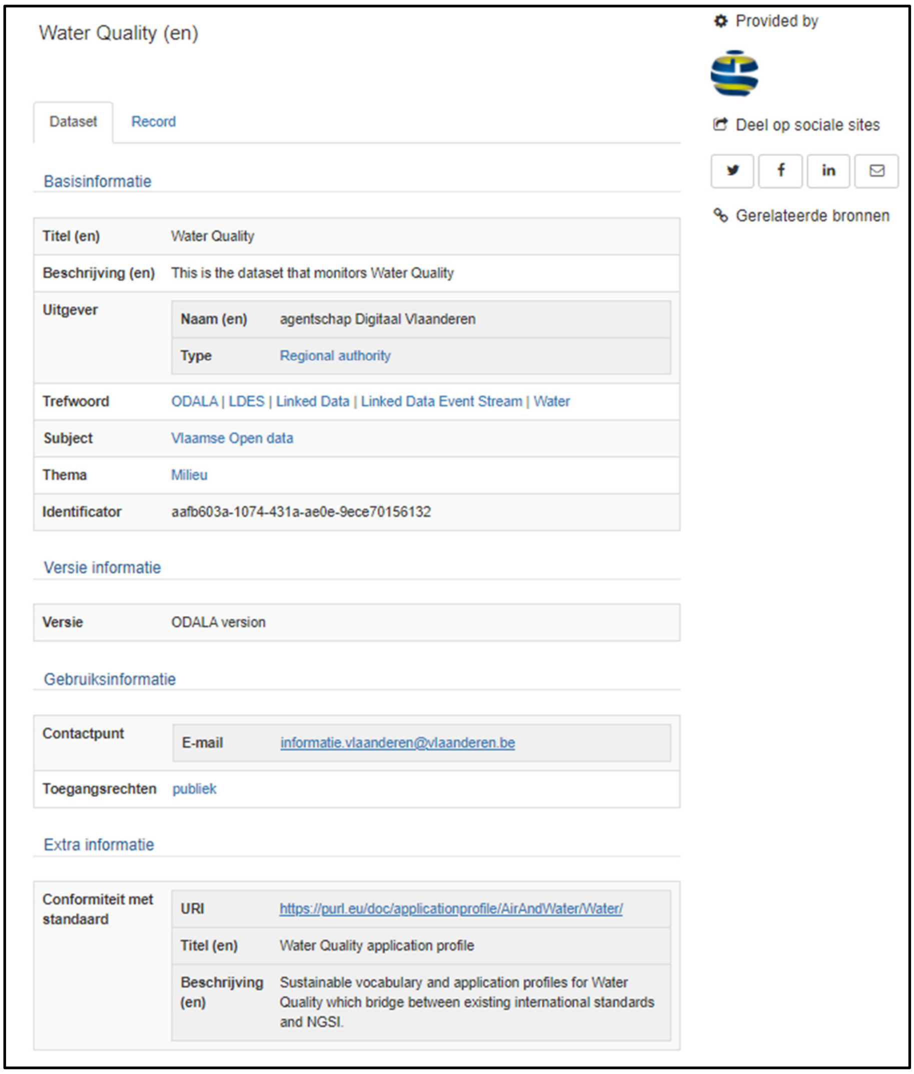
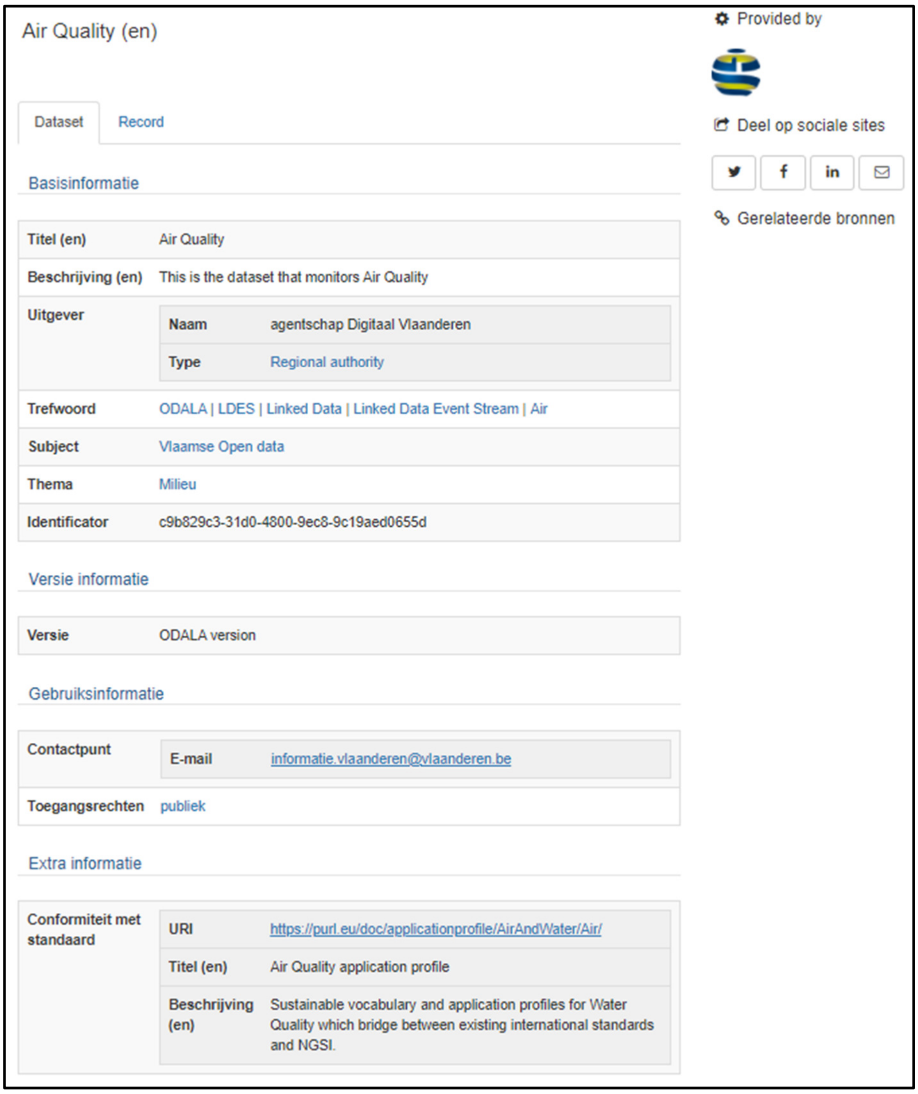


***Figure 14**: Illustration of the metadata description of the dataset Water Quality*

- Dataset Air Quality*:* [*https://geonetwork-oslo-demo.gim.be/geonetwork/srv/dut/catalog.search#/metadata/b4aa6493-c55f-49d9-bcf7-bf0467f37ef4*](https://geonetwork-oslo-demo.gim.be/geonetwork/srv/dut/catalog.search#/metadata/b4aa6493-c55f-49d9-bcf7-bf0467f37ef4)



Air Quality (en)

Provided by

Dataset | Record

Deel op sociale sites

Gerelateerde bronnen

Basisinformatie

| | |
|---|---|
| Titel (en) | Air Quality |
| Beschrijving (en) | This is the dataset that monitors Air Quality |
| Uitgever | Naam: agentschap Digitaal Vlaanderen<br>Type: Regional authority |
| Trefwoord | ODALA \| LDES \| Linked Data \| Linked Data Event Stream \| Air |
| Subject | Vlaamse Open data |
| Thema | Milieu |
| Identificator | c9b829c3-31d0-4800-9ec8-9c19aed0655d |

Versie informatie

| | |
|---|---|
| Versie | ODALA version |

Gebruiksinformatie

| | |
|---|---|
| Contactpunt | E-mail: informatie.vlaanderen@vlaanderen.be |
| Toegangsrechten | publiek |

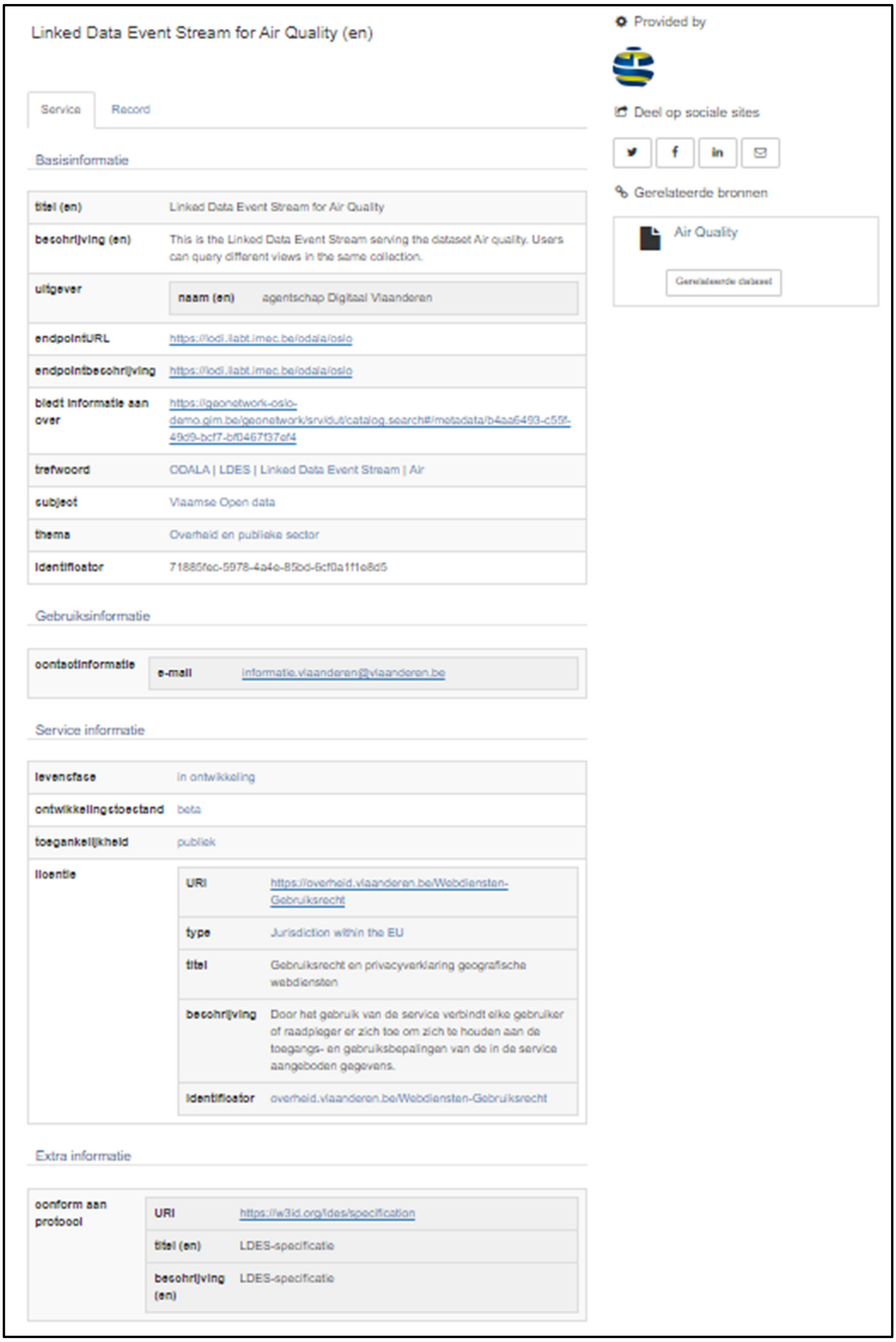
Extra informatie

| | |
|---|---|
| Conformiteit met standaard | URI: https://purl.eu/doc/applicationprofile/AirAndWater/Air/<br>Titel (en): Air Quality application profile<br>Beschrijving (en): Sustainable vocabulary and application profiles for Water Quality which bridge between existing international standards and NGSI. |

***Figure 15:** Illustration of the metadata description of the dataset Air Quality*

Secondly the dataservices are illustrated:

- Linked Data Even Stream for Air Quality: [https://geonetwork-oslo-demo.gim.be/geonetwork/srv/dut/catalog.search#/metadata/cb3d3c25-ec26-46e1-95f7-e8ddcbb8eec9](https://geonetwork-oslo-demo.gim.be/geonetwork/srv/dut/catalog.search#/metadata/cb3d3c25-ec26-46e1-95f7-e8ddcbb8eec9)



Linked Data Event Stream for Air Quality (en)

Service | Record

Basisinformatie

| | |
|---|---|
| titel (en) | Linked Data Event Stream for Air Quality |
| beschrijving (en) | This is the Linked Data Event Stream serving the dataset Air quality. Users can query different views in the same collection. |
| uitgever | naam (en) agentschap Digitaal Vlaanderen |
| endpointURL | https://lodi.ilabt.imec.be/odala/oslo |
| endpointbeschrijving | https://lodi.ilabt.imec.be/odala/oslo |
| biedt informatie aan over | https://geonetwork-oslo-demo.gim.be/geonetwork/srv/dut/catalog.search#/metadata/b4aa6493-c55f-49d9-bcf7-bf0467f37ef4 |
| trefwoord | ODALA \| LDES \| Linked Data Event Stream \| Air |
| subject | Vlaamse Open data |
| thema | Overheid en publieke sector |
| identificator | 71885fec-5978-4a4e-85bd-6cf0a1f1e8d5 |

Gebruiksinformatie

| | |
|---|---|
| contactinformatie | e-mail informatie.vlaanderen@vlaanderen.be |

Service informatie

| | |
|---|---|
| levensfase | in ontwikkeling |
| ontwikkelingstoestand | beta |
| toegankelijkheid | publiek |
| licentie | URI: https://overheid.vlaanderen.be/Webdiensten-Gebruiksrecht<br>type: Jurisdiction within the EU<br>titel: Gebruiksrecht en privacyverklaring geografische webdiensten<br>beschrijving: Door het gebruik van de service verbindt elke gebruiker of raadpleger er zich toe om zich te houden aan de toegangs- en gebruiksbepalingen van de in de service aangeboden gegevens.<br>identificator: overheid.vlaanderen.be/Webdiensten-Gebruiksrecht |

Extra informatie

| | |
|---|---|
| conform aan protocol | URI: https://w3id.org/ldes/specification<br>titel (en): LDES-specificatie<br>beschrijving (en): LDES-specificatie |

Provided by

Deel op sociale sites

Gerelateerde bronnen

Air Quality

Gerelateerde dataset

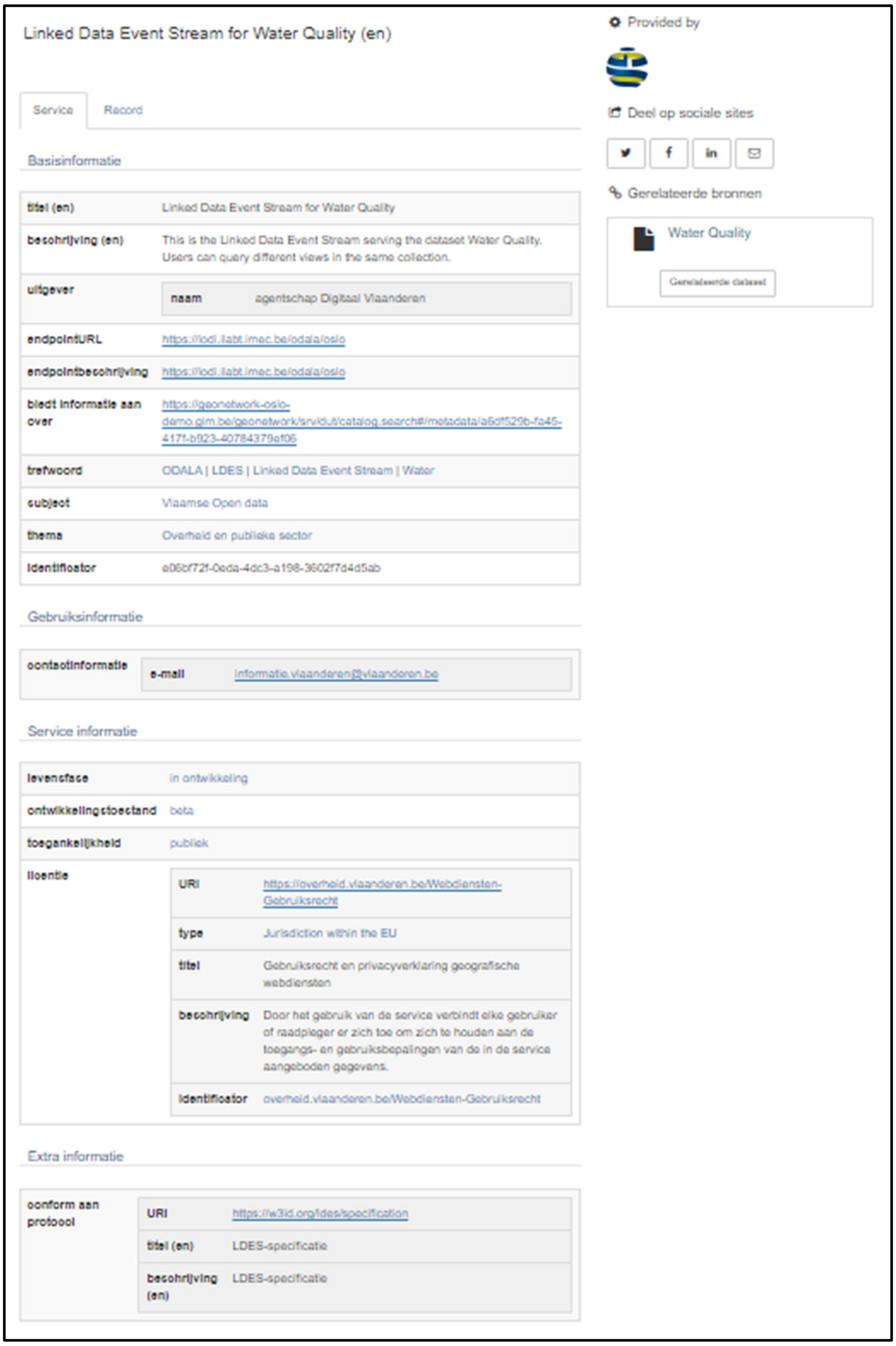
***Figure 16****: Illustration of the metadata description of the dataservice LDES for Air Quality*

lodi.ilabt.imec.be/odala/oslo


```
 },
- @included: [
  - {
        @id: "https://lodi.ilabt.imec.be/odala/data/observations/10839640967",
        @type: "ObservationCollection",
      - ObservationCollection.hasFeatureOfInterest: {
            @type: "SpatialSamplingFeature",
            SamplingFeature.sampledFeature: "http://www.wikidata.org/entity/Q56245086",
          - http://www.w3.org/ns/locn#geometry: {
                @type: "Geometry",
                Geometry.asWkt: "<http://www.opengis.net/def/crs/EPSG/0/4979> POINT(50.866 4.324 46.1)"
            }
        },
      - ObservationCollection.madeBySensor: {
          - @type: [
                "Sensor",
                "Device"
            ],
            @id: "https://lodi.ilabt.imec.be/odala/data/sensors/71858",
            Device.manufacturerName: "Plantower",
            Device.modelName: "PMS5003"
        },
        ObservationCollection.resultTime: "2022-06-09T11:56:47Z",
      - ObservationCollection.hasMember: [
          - {
                type: "Observation",
                @id: "https://lodi.ilabt.imec.be/odala/data/observations/24070007688",
                Observation.observedProperty: "http://www.wikidata.org/entity/Q48035511",
              - Observation.hasSimpleResult: {
                    @type: "http://w3id.org/lindt/custom_datatypes#ucum",
                    @value: "8 ug/m3"
                }
            },
```


***Figure 17**: Illustration of the dataservice LDES for Air Quality*

- Linked Data Even Stream for Water Quality: https://geonetwork-oslo-demo.gim.be/geonetwork/srv/dut/catalog.search#/metadata/6fcd805c-2857-422a-9562-c14bac3450d6



Linked Data Event Stream for Water Quality (en)

Service | Record

Basisinformatie

| | |
|---|---|
| titel (en) | Linked Data Event Stream for Water Quality |
| beschrijving (en) | This is the Linked Data Event Stream serving the dataset Water Quality. Users can query different views in the same collection. |
| uitgever | naam: agentschap Digitaal Vlaanderen |
| endpointURL | https://lodi.ilabt.imec.be/odala/oslo |
| endpointbeschrijving | https://lodi.ilabt.imec.be/odala/oslo |
| biedt informatie aan over | https://geonetwork-oslo-demo.gim.be/geonetwork/srv/dut/catalog.search#/metadata/a6df529b-fa45-417f-b923-40784379ef06 |
| trefwoord | ODALA \| LDES \| Linked Data Event Stream \| Water |
| subject | Vlaamse Open data |
| thema | Overheid en publieke sector |
| identificator | e06bf72f-0eda-4dc3-a198-3602f7d4d5ab |

Gebruiksinformatie

| | |
|---|---|
| contactinformatie | e-mail: informatie.vlaanderen@vlaanderen.be |

Service informatie

| | | |
|---|---|---|
| levensfase | in ontwikkeling | |
| ontwikkelingstoestand | beta | |
| toegankelijkheid | publiek | |
| licentie | URI | https://overheid.vlaanderen.be/Webdiensten-Gebruiksrecht |
| | type | Jurisdiction within the EU |
| | titel | Gebruiksrecht en privacyverklaring geografische webdiensten |
| | beschrijving | Door het gebruik van de service verbindt elke gebruiker of raadpleger er zich toe om zich te houden aan de toegangs- en gebruiksbepalingen van de in de service aangeboden gegevens. |
| | identificator | overheid.vlaanderen.be/Webdiensten-Gebruiksrecht |

Extra informatie

| | | |
|---|---|---|
| conform aan protocol | URI | https://w3id.org/ldes/specification |
| | titel (en) | LDES-specificatie |
| | beschrijving (en) | LDES-specificatie |

Provided by

Deel op sociale sites

Gerelateerde bronnen

Water Quality

Gerelateerde dataset

***Figure 18***: *Illustration of the metadata description of the dataservice LDES for Water Quality*

```
lodi.ilabt.imec.be/odala-oslo-water-quantity/hierarchical?type=http%3A%2F%2Fdef.isotc211.org%2Fiso19156%2F2011%2FObservation%23OM_Observation&timeAt=2022-06-09T00:00:05.097Z&endTimeAt=2022-06-09T00:00:0
- @included: [
  - {
      id: "urn:ngsi-ld:observation:1654732140:60960042_0",
      type: "http://def.isotc211.org/iso19156/2011/Observation#OM_Observation",
    - http://def.isotc211.org/iso19156/2011/Observation#OM_Observation.metadata: {
          type: "https://purl.eu/doc/applicationprofile/AirAndWater/Core/#Metadata",
          description: "P04_020",
          title: "Massemen_P"
      },
      http://def.isotc211.org/iso19156/2011/Observation#OM_Observation.observedProperty: "http://purl.obolibrary.org/obo/CHEBI_28112",
      http://def.isotc211.org/iso19156/2011/Observation#OM_Observation.resultTime: "2022-06-09T01:01:00.000+02:00",
    - http://def.isotc211.org/iso19156/2011/Observation#Observation.result: {
          type: "https://schema.org/QuantitativeValue",
          unitText: "millimeter per hour",
          value: 0,
          observedAt: "2022-06-09T01:01:00.000+02:00",
          unitCode: "mm/h"
      },
      @context: "https://uri.etsi.org/ngsi-ld/v1/ngsi-ld-core-context.jsonld",
      memberOf: "https://lodi.ilabt.imec.be/odala-oslo-water-quantity/dataset?type=http%3A%2F%2Fdef.isotc211.org%2Fiso19156%2F2011%2FObse
  },
  - {
      id: "urn:ngsi-ld:observation:1654732140:60960042_1",
      type: "http://def.isotc211.org/iso19156/2011/Observation#OM_Observation",
    - http://def.isotc211.org/iso19156/2011/Observation#OM_Observation.metadata: {
          type: "https://purl.eu/doc/applicationprofile/AirAndWater/Core/#Metadata",
          description: "P04_020",
          title: "Massemen_P"
      },
      http://def.isotc211.org/iso19156/2011/Observation#OM_Observation.observedProperty: "http://purl.obolibrary.org/obo/CHEBI_28112",
      http://def.isotc211.org/iso19156/2011/Observation#OM_Observation.resultTime: "2022-06-09T01:02:00.000+02:00",
    - http://def.isotc211.org/iso19156/2011/Observation#Observation.result: {
          type: "https://schema.org/QuantitativeValue",
          unitText: "millimeter per hour",
          value: 0,
          observedAt: "2022-06-09T01:02:00.000+02:00",
          unitCode: "mm/h"
```


*Figure 19: Illustration of the dataservice LDES for Water Quality*

The order in which they are created is irrelevant. A good editorial system will support any order.

## Step 4: Connect dataservices and datasets

Next, the LDES datasets and LDES dataservices are coupled. This is shown in Figure 20 below. Data wise is the connection trivial, but it is more complicated for an editor. There is no preferred route to connect them as metadata editors can start from a dataservice or from a dataset. Because a dataset can be served by many dataservices and a dataservice can serve multiple datasets and both collections should not be identical, creating a UI that supports the management of this (possibly complex) network of relationships requires some work.

Last step in documenting the dataservices (3), is to couple the data services to the correct datasets, so it is clear which datasets they serve.

The element in the DCAT-standards to use for the coupling to the correct datasets is the 'servesDataset' element. The Linked Data Event Stream dataservice can be coupled via ‘servesDataset’ to the correct datasets it serves. Data providers can link to the dataset collections, or to the dataset fragmentations.

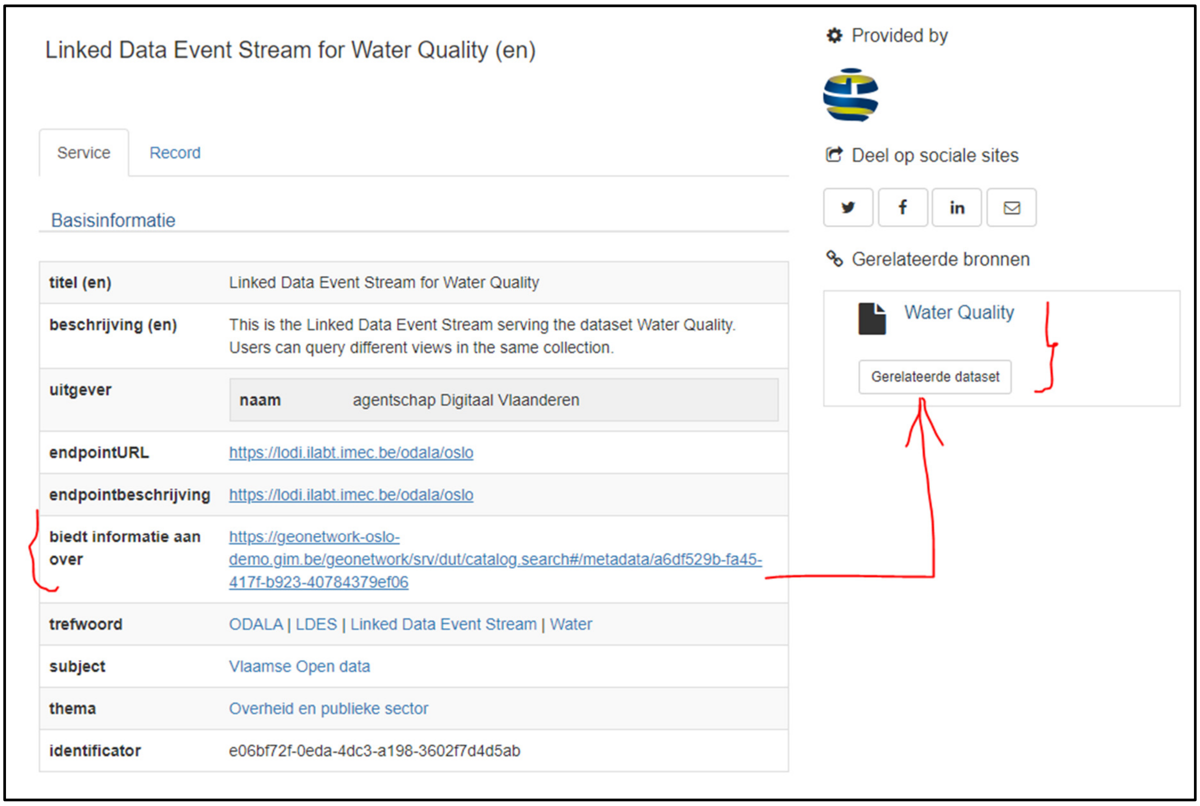



***Figure 20**: A print-screen of this coupling of Dataservices to Datasets in the reference implementation*

## Step 5: Connect with the data models defined in WP5.1

Step 5 highlights the coupling with the vocabularies and application profiles that are defined in WP5.1. By connecting LDES datasets with these, the knowledge graph interconnects the high level perspective (metadata) with the (semantics of the) actual data. This creates new discovery possibilities such as “finding a LDES dataservice that provides information about an air quality observation”.

The connection with the data models is supported in DCAT via the dct:conformsTo[22] property. The usage of this property has been intensively discussed as its semantics slightly differ when connected with a dataset, a distribution, a dataservice, a catalogue record or a catalogue.  Also the property is/can be used to capture different perspectives aggregating all in one property. This is often not desirable when these perspectives should be presented to humans in a UI. Since it is difficult to assess the full impact of this aggregation, the decision has

[22] https://w3c.github.io/dxwg/dcat/#Property:resource_conforms_to

been taken to use for now the dct:conformsTo property. And evaluate during the development if it is necessary to split out the perspectives (which can be done using subproperties).

Below in Figure 21 an example is shown for the Water Quality dataset.

Dataset | Record

### Basisinformatie

| | | |
|---|---|---|
| **Titel (en)** | Water Quality | |
| **Beschrijving (en)** | This is the dataset that monitors Water Quality | |
| **Uitgever** | **Naam (en)** | agentschap Digitaal Vlaanderen |
| | **Type** | Regional authority |
| **Trefwoord** | ODALA \| LDES \| Linked Data \| Linked Data Event Stream \| Water | |
| **Subject** | Vlaamse Open data | |
| **Thema** | Milieu | |
| **Identificator** | aafb603a-1074-431a-ae0e-9ece70156132 | |

### Versie informatie

| | |
|---|---|
| **Versie** | ODALA version |

### Gebruiksinformatie

| | | |
|---|---|---|
| **Contactpunt** | **E-mail** | informatie.vlaanderen@vlaanderen.be |
| **Toegangsrechten** | publiek | |

### Extra informatie

| | | |
|---|---|---|
| **Conformiteit met standaard** | **URI** | https://purl.eu/doc/applicationprofile/AirAndWater/Water/ |
| | **Titel (en)** | Water Quality application profile |
| | **Beschrijving (en)** | Sustainable vocabulary and application profiles for Water Quality which bridge between existing international standards and NGSI. |

***Figure 21**: An example in the reference implementation of the coupling of the dataset Water Quality to the used Application Profile for Water Quality, defined in WP.5.1*

## Step 6: Share the datasets and dataservices throughout National and European (Open) data Portals

This step is about reaching out to potential (re)users of the data. Over the past decades Europe has invested in a network of data portals to make the available data visible and discoverable for the European (re)users. Like in any other member state of the European Union, Belgium has created a national network where metadata is aggregated from the sources (local governments like cities or agencies). This is shown in figure 18. For the ODALA, the created LDES dataservices are set up and maintained by the agency Digitaal Vlaanderen in DCAT, therefore the metadata has been entered directly in its metadata catalogue, called *Metadata Vlaanderen*.

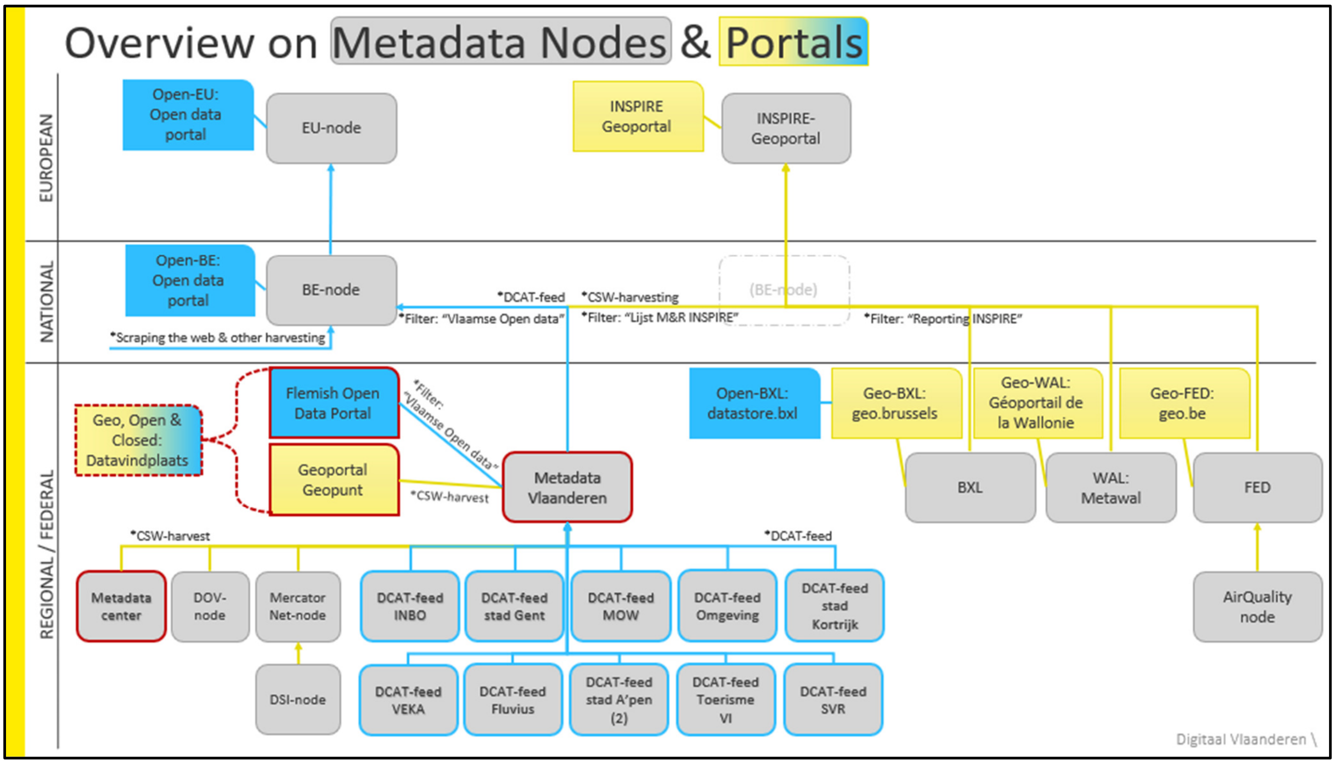



***Figure 22**: An overview of the harvesting landscape from regional metadata nodes and portals into national and European nodes and portals, based on the AS IS production environments.*

To ensure the correct aggregation and propagation of the metadata descriptions throughout this network, the

DCAT profile Metadata DCAT also imposes guidelines on identifiers. By promoting (and imposing) strong identifier management at the source the setup of a correct functioning aggregation process is much simpler.

Recently a new initiative, called *Datavindplaats*, has started creating a new end-user portal combining the Flemish Open Data Portal and Geopunt. The *Datavindplaats* will be the place for finding data (datasets and dataservices) offered by the Flemish Public sector. Therefore it will not only provide information about public accessible data but also about restricted or closed access data. The work presented for enabling the discoverability of LDES services, is thus a specific case in the *Datavindplaats’* objectives. A preview of the new portal is shown below in the illustrations of Figures 19 and 20 in particular for the search on the LDES-topic.

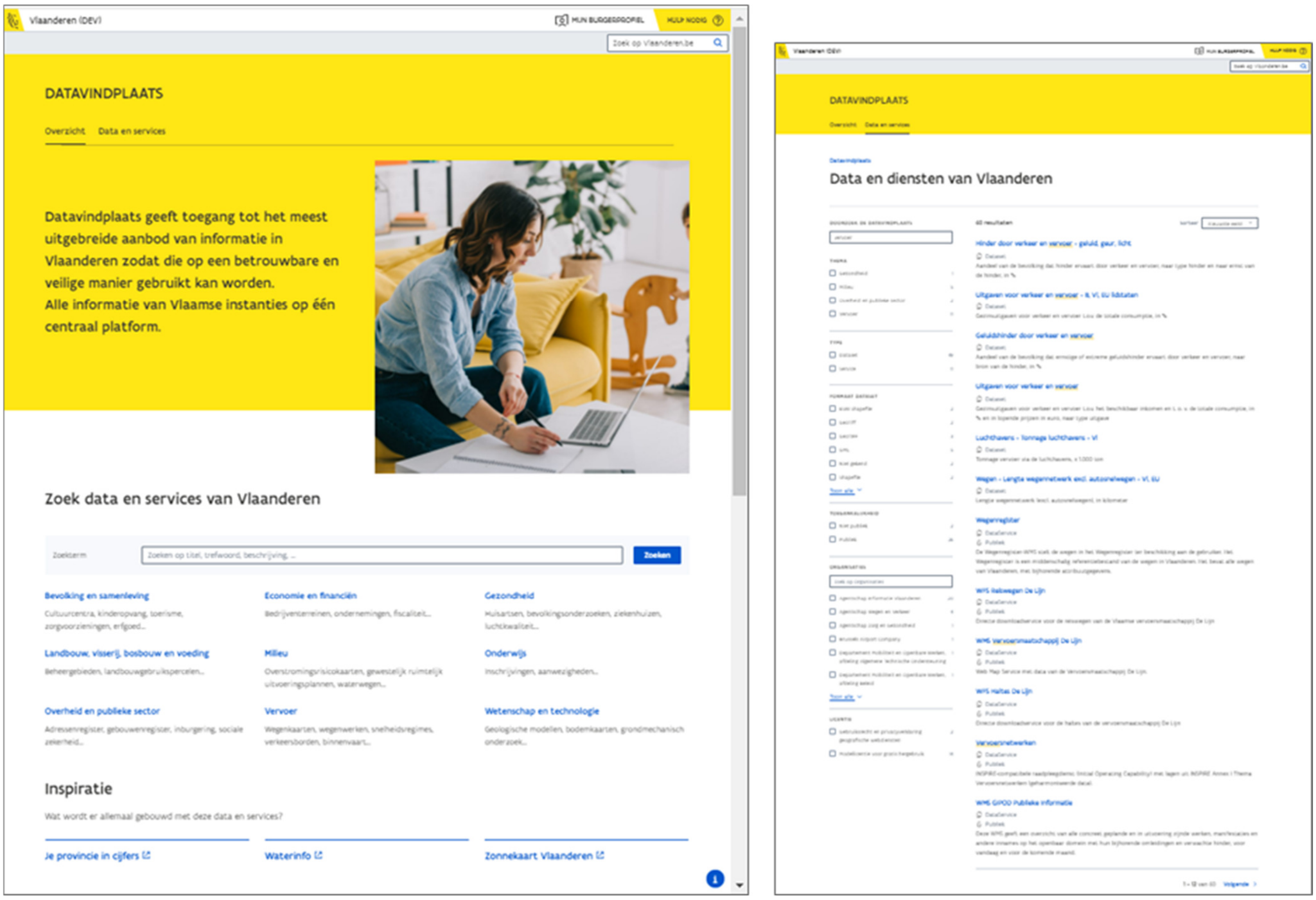



***Figure 23***: *Print-screens of the Datavindplaats, the new umbrella concept for the user-friendly portal, where all data and information is gathered in one catalogue, combining Open, Geospatial, Restricted and Closed Data and Dataservices.*

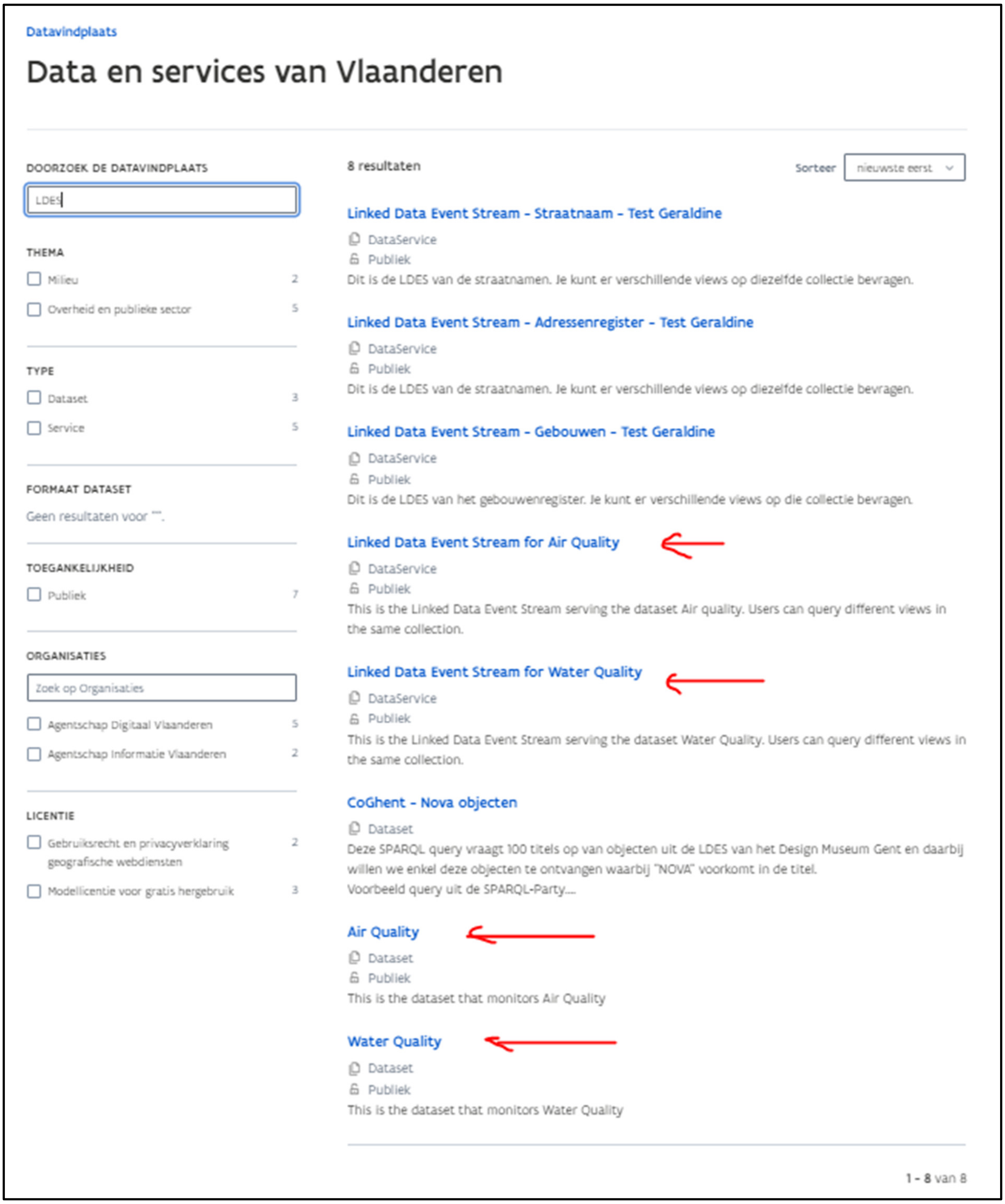



***Figure 24:*** *A print-screen of today's Datavindplaats, filtered on a 'LDES'-search, where the 4 results of the documented ODALA-datasets and dataservices are shown*

# TASK 5.4.: AN OPEN DATA INTERFACE FOR NGSI-LD

## Linked Data Event Streams (LDES) for open data publishing

Modern cities maintain numerous data resources such as real time air quality observation and detailed descriptions of the road network. However, these resources are primarily used by the cities themselves to perform their core duties such as city planning. Making the data publicly available is often a nice-to-have, and this is reflected in which datasets are made public, and how they are published. Concretely, open data publishing is often limited to providing an open data portal of static data dumps. Although this is certainly better than not offering any open data at all, this method is ill-suited to rapidly changing datasets such as air quality observations. Nevertheless, cities often express their desire to make such datasets publicly available - as long as it does not interfere with their internal systems that are needed to perform their duties. One common solution to this problem is to provide data access APIs, but with some form of authentication and rate limiting to avoid overloading the systems. Unfortunately this makes the data harder to reuse, and smaller cities do not have the resources to build and maintain such APIs.

Luckily, since we focus on Linked Data within this project, there are other ways of publishing data. All ways of publishing Linked Data (data dumps, query endpoints, subject pages, …) expose a fragment of the whole dataset. In the case of a data dump, there is simply one fragment containing all data. A SPARQL endpoint on the other hand exposes all fragments that match any graph pattern. This idea forms the conceptual framework of Linked Data Fragments[23] (LDF): each publishing method exposes certain fragments of the data, and each method includes hypermedia controls to access the fragments. The core insight of this framework is that no single publishing method is great at everything, and that there are alternatives to the conventional data dumps or query APIs - and that their strengths should be explored as well.

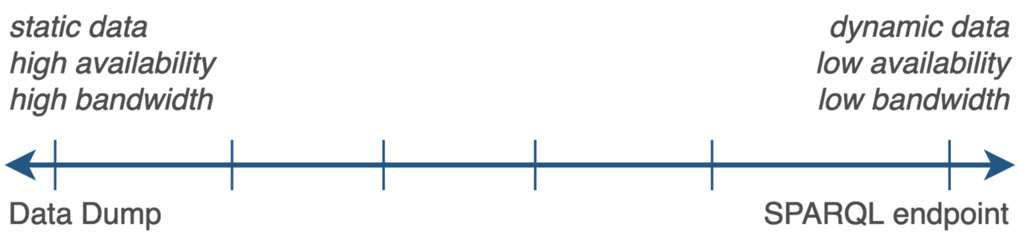



***Figure 25:*** The two extremes of the Linked Data Fragments axis are static data dumps on one hand, and query APIs on the other. Both approaches have their own characteristics, with no single method being the indisputable best approach for all use cases. For example, maintaining a data portal of static data dumps is easier to maintain

[23] https://linkeddatafragments.org/

than a feature-rich query API, but these APIs can more easily return realtime data. There are many in-between approaches on this axis though, notably including Triple Pattern Fragments[24] and Linked Data Event Streams.

As part of this project, we refined the definitions and tooling of one such alternative LDF interface: the Linked Data Event Stream[25] (LDES). Intuitively, an LDES is an append-only collection of immutable objects. Everything that has ever been added to an LDES collection remains part of the collection forever, and the individual data objects never change. A Basic LDES is a special case where each data object contains a (creation) timestamp, and the collection is fragmented by grouping objects from the same time interval in the same fragment. This simple restriction entails that at any given time, only one data fragment can still change - all the others contain all the data they will ever contain and can be persisted to disk or to a cloud-based storage system, as if they were small static data dumps.

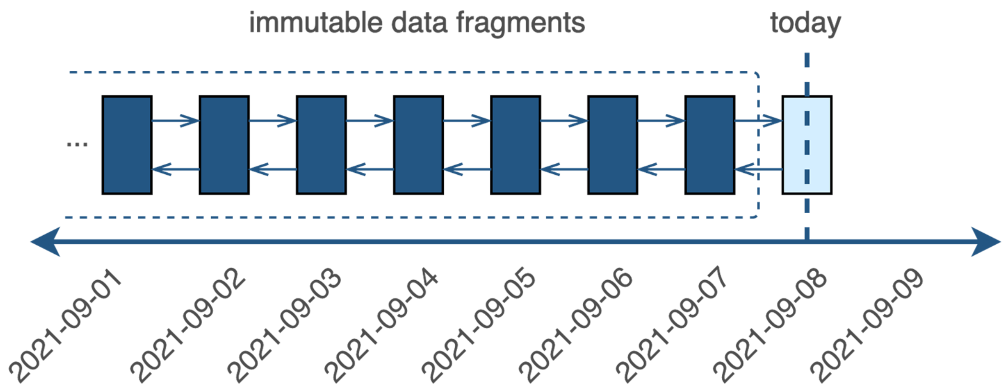



***Figure 26:*** As an example of a Linked Data Event Stream, suppose a city wants to publish the occupancy rates of a car park. Each observation is an immutable data object that contains an observation timestamp, making this dataset a good fit for a LDES API. All observations that happened on the same day can be grouped into the same data fragment, with hypermedia controls linking subsequent fragments together in a sort of linked list. Only today's data fragment can still change, making all previous fragments permanently cacheable.

## How to map the NGSI-LD interface conform to the LDES principles?

Even most popular NGSI-LD context brokers such as Scorpio and Orion-LD do not currently provide built-in rate limiting. As a consequence, exposing these systems to the public falls under the undesirable scenario that was described earlier; cities that do want to publish the underlying data need to deploy their own data access APIs to protect their underlying systems. However, the NGSI-LD specification handles the temporal evolution of entities - which should make an LDES API a good fit for these systems. Moreover, because all spec-compliant

[24] https://linkeddatafragments.org/specification/triple-pattern-fragments/
[25] https://w3id.org/ldes/specification

implementations implement a common set of API HTTP Binding interfaces, reusable software components can be developed so that even cities with small technical divisions can set up these APIs.

The NGSI-LD specification guarantees that all entities, relations, and properties contain several useful timestamps: createdAt, modifiedAt, and observedAt. Of these three, the first two are the most applicable to be used in the creation of a Basic LDES. The latter property does not guarantee that old data fragments will never receive new data, notably in the case when data from external systems is periodically imported. The createdAt property is to be used for entities that do not change after being created, such as AirQualityObserved entities from Smart Data Models. If the entities can change after creation, the modifiedAt property should be used - but this requires the creation of so-called version objects. Suppose a sensor has been moved so that its location property is updated. According to the semantics of RDF, the collection of a whole now states that the sensor has multiple locations. This problem can be avoided by giving each version of an entity a new identifier - often by appending the timestamp to the identifier - and by adding a new relation from the version object to the original entity.

NGSI-LD data meets the requirements for publishing an LDES, however, any non-trivial dataset requires fragmentation. The specification mandates a Temporal Query language that can return all data from a specified time interval. This API can be used to generate the fragments, the only remaining challenge is determining which time intervals should form a fragment. A naive solution is to use fixed time intervals, but even this solution needs to know what the first and last objects are - lest we publish an infinite collection of empty fragments. Moreover, overfull fragments may induce bandwidth issues.

The solution to both problems is to make use of a hierarchical fragmentation. At the highest level, we employ a coarse-grained fragmentation, using monthly or weekly time intervals. The fragment's hypermedia controls include a link to the next or previous fragment if, respectively, a timerel=after or timerel=before query returns any data. If a fragment contains too much data, relations to a more fine-grained level of fragmentation are returned instead of the data itself. Intuitively, if a Basic LDES is a linked list, this hierarchical LDES forms a B-tree. This has the following important consequence: generating any fragment requires only a limited amount of simple queries to the NGSI-LD system.

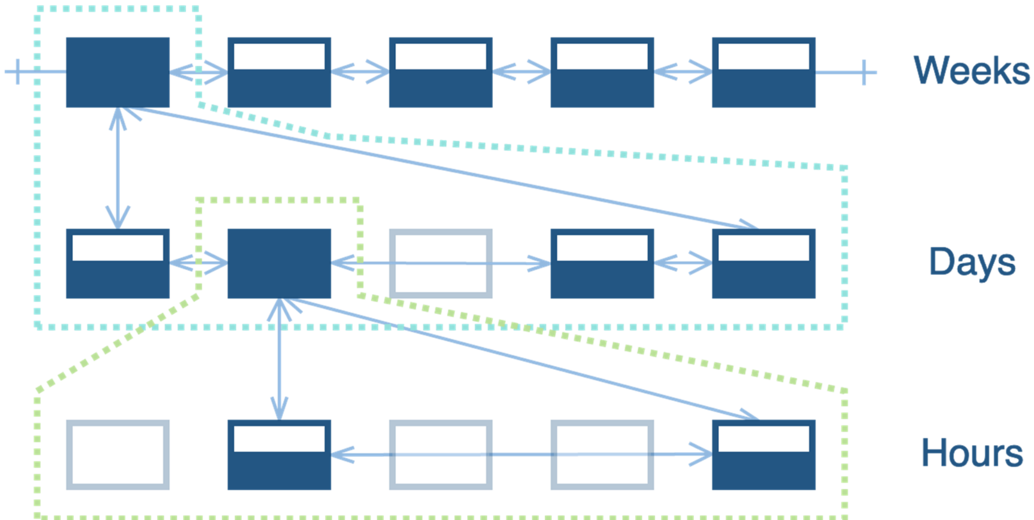



***Figure 27:*** A schematic overview of an LDES collection with a hierarchical fragmentation. At the highest level, there's a link to a previous or next fragment if, respectively, a before or after temporal query indicates there is data in that direction. If a fragment would return too much data, an empty fragment with links to fragments at a lower level is returned instead. Finally, empty fragments are skipped by the hypermedia controls.

## How to deploy NGSI-LDES in practice?

NGSI-LDES[26] publishes data from an NGSI-LD system with LDES using a hierarchical fragmentation (see Fig. X above). At the beginning of the ODALA project, no broker was able to paginate temporal queries and return a count of the number of results, which are both requirements for the NGSI-LDES. However, at the end of 2021 Scorpio broker supported both requirements (issue #234) allowing the development of NGSI-LDES. Before, we implemented a proxy server in Node.js and Typescript, branded as NGSI-LDF, that can be configured to:

- republish specific entity types;
- use a given base URI for the fragment and version URIs;
- only republish data from a specific time interval (e.g., only the last 7 days);
- limit the fragments to contain a specific amount of data objects.

In NGSI-LDES[27], data is fragmented per day at the highest level. The NGSI-LD counting mechanism is used in several scenarios. First, to create a relation towards the previous or next day if necessary. Second, whether to return the entities within the fragment's time interval. When a certain number of entities per fragment *m* is met, relations are created towards more fine-grained fragments. This is done by splitting the time interval into *m* parts. For every fragment, cache controls are added. When the fragment's time interval is already past, then the fragment is cached for one day (Cache-Control: public, max-age=86400). Otherwise, the fragment is cached for 5 seconds.

[26] https://github.com/TREEcg/ngsi-ldes/tree/feat/componentsjs
[27] https://github.com/TREEcg/ngsi-ldes/tree/feat/componentsjs

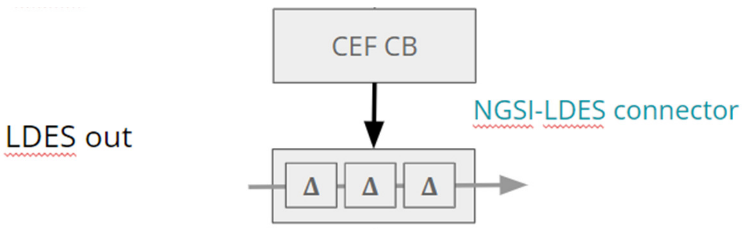



*Figure 28:* The NGSI-LDES connector runs on top of an NGSI-LD Context Broker and exposes its entities as a collection of versioned objects.

Finally, it is important to mention that the data from the broker is not copied into the NGSI-LDES. When a fragment is fetched, multiple NGSI-LD requests are executed towards the temporal interface binding. The second time a client fetches this fragment, a caching layer can provide the fragment increasing the scalability of the NGSI-LD system. To make our data API as easy to deploy as NGSI-LD systems themselves, we bundled the server code with a Docker Compose file alongside the Scorpio broker image. In the Greenmov project, NGSI-LDES has been succesfully deployed with Orion-LD using Mintaka and with IDLab's custom broker Obelisk.

## Task 5.3.: Tooling for NDSI-LDES - Extending Comunica with time series and geospatial queries supported by NGSI-LD

In this task, we developed tools to work with published Linked Data Event Streams, such as NGSI-LDES. Three tools have been created described below.

An LDES Client has been implemented for the replication and synchronization of an LDES. The client is written with the knowledge graph querying framework Comunica[28] and handles all the parsing of the hypermedia controls, interpretation of caching headers, and periodic polling of mutable data fragments. It provides a Javascript/Typescript API, and additionally provides a CLI for usage with other programming languages. This tool is key for third-parties as this allows them to build new indexes and query services that are in sync with an LDES.

A preliminary geospatial extension of the Comunica SPARQL query engine has been created[29]. This extension

[28] https://github.com/TREEcg/event-stream-client/
[29] https://github.com/dreeki/masterproef-industrial_engineer-poc

allows the client to filter a query response using the GeoSPARQL query language. Concretely, a Comunica actor is created that understands GeoSPARQL functionalities (topological and non-topological). This actor uses the “sparqlalgebrajs” package to transform the SPARQL query into SPARQL algebra and an extended "sparqlee” package as expression evaluator to have a correct execution of the SPARQL algebra with an understanding of GeoSPARQL functionalities. This client has been tested with heterogeneous interfaces that are supported by the Comunica SPARQL query engine: data dump, Triple Pattern Fragments interface and SPARQL endpoint. To conclude, a consumer of an NGSI-LD context broker first needs to replicate the data from the NGSI-LDES with the LDES client into a local system, such as GraphDB, to support one of the heterogeneous interfaces and then graph querying with geospatial functionalities can be achieved.

A package[30] is developed to replicate an event stream into a back-end system of choice (MongoDB, GraphQL, NGSI-LD) using connectors. We created such a connector[31] to load LDES data into an NGSI-LD system, using the standardized HTTP bindings (Fig. X below). This can be an interesting approach for cities that want to store event data from other data publishers, such as vendors that maintain air quality sensors.

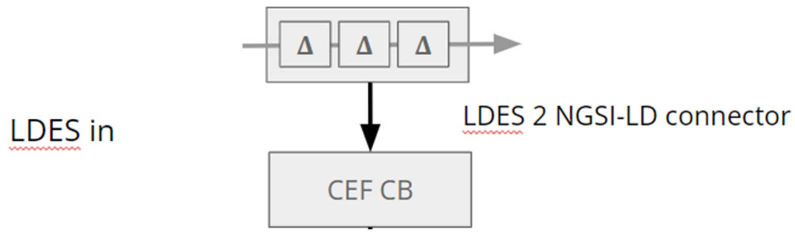



***Figure 29**:* Data from an existing LDES can be ingested into an NGSI-LD Context Broker

[30] https://github.com/Informatievlaanderen/ldes2service

[31] https://github.com/TREEcg/connectors/tree/main/packages/connector-write-ngsi-ld

# Future Work

The presented steps form a roadmap for any project that aims to provide a catalogue in a data space. The first steps concern the creation of a semantical agreement (an application profile) about the catalogued data. The following steps concern the activities that contributors to the catalogue have to do, while the last step is about the reuse of the catalogue by having it harvested by others or by providing a portal for it.

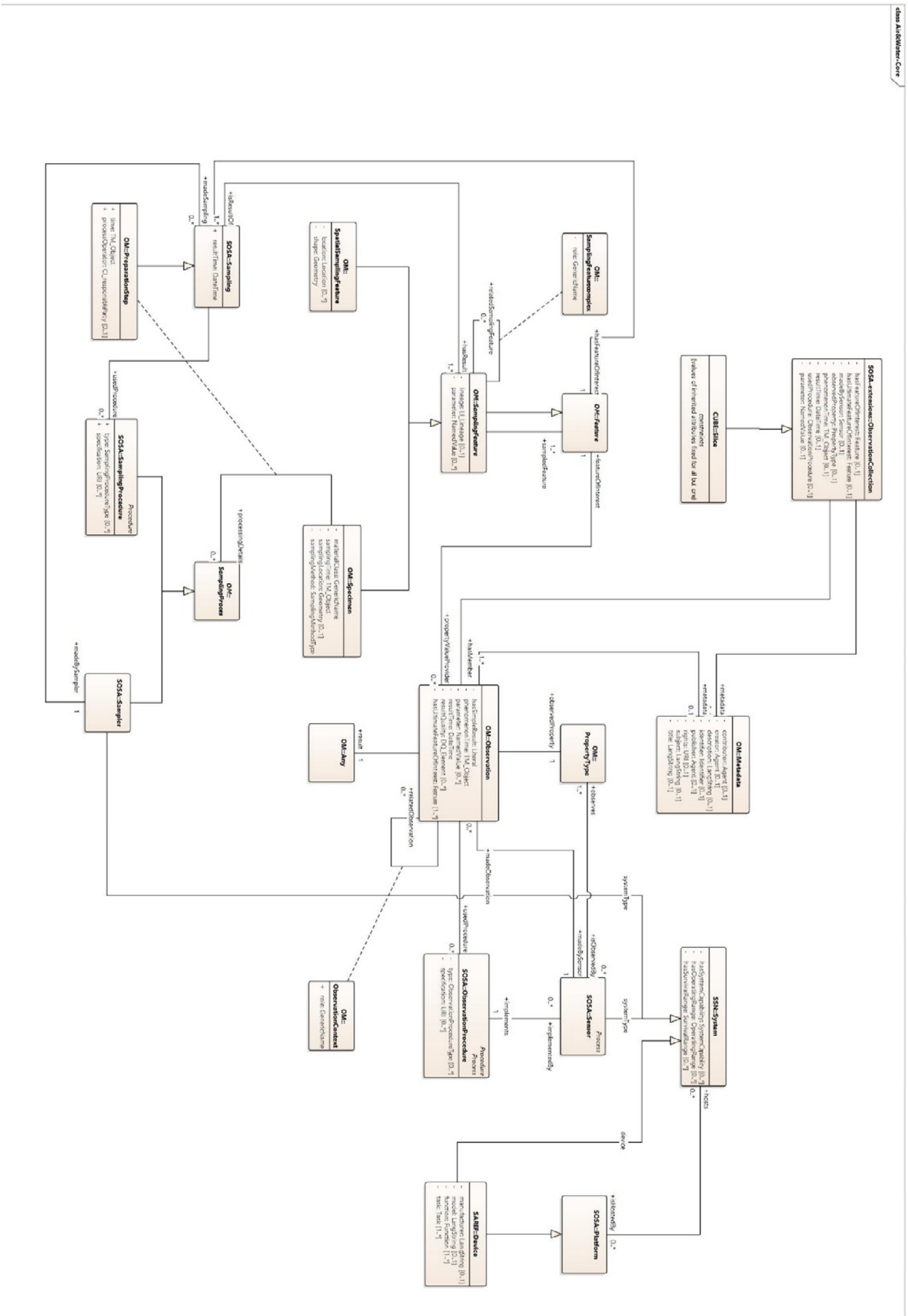
The groundwork in each of the steps has been done. To reach a production ready setup at all steps, iterative improvements are done. In this process the LDES case described here in this document is taken as one of the to-be-supported use cases.

When having a closer look at the current metadata agreements and the open discussions that exists in this area (beyond ODALA), the following topics must be discussed for the LDES case:

- **dataset series and dataset versioning**. Currently W3C DCAT is preparing guidelines[32] for this.
- exploring the information needs to facilitate better **user driven search** for LDES data services. E.g. would it be possible to find the LDES air quality data service based on the question "Give me data about O2 particles in air"?
- **self-descriptive LDES dataservices**. When LDES dataservices describe their data in a coherent way so that the metadata description can be derived from the LDES endpoint. That would avoid that a LDES publisher would create in a separate system a description about its LDES endpoint, and that this information gets out of date because the metadata is not directly coupled with the actual dataservice.
- **self-descriptive NGSI-LDES.** NGSI-LDES already publishes a DCAT catalogue of available LDESs. In the Greenmov project, the LDES description can be extended with SHACL shapes based on the types/ NGSI-LD resource.
- **dynamic coverage index.** In order to query over a federation of context brokers, applications need to understand what data is covered within the brokers' collections. In the Greenmov project, research is performed on how using NGSI-LDES as a basis an index can be created to increase the discoverability of a context broker inside a federation.
- **derived LDES.** Further work is also planned to map (or project) one LDES into another. Such a mapping could for example be used to support multiple data models. This is expected to become an important benefit of working with Event Streams, as cities may want to use the Smart Data Models to be interoperable with other cities - but they may also have to publish the data in local standards such as the OSLO application profiles.

[32] https://w3c.github.io/dxwg/dcat/

# ANNEX I: UML MODEL OSLO AIR & WATER[33]



[33] https://github.com/Informatievlaanderen/OSLOthema-airAndWater/blob/master/site-skeleton/OSLO-airAndWater-Core-ap/overview.jpg