SALTED

Situation-Aware Linked
heTerogeneous Enriched Data

D3.2: Report on publication of linked and enriched data streams to the European Data Portal

Work package	WP 3
Task	Task 3.2
Due date	30/09/2023
Submission date	02/11/2023
Deliverable lead	UC
Version	1.0
Authors	Laura Martín (UC), Víctor González (UC), Pablo Sotres (UC), Juan Ramón Santana (UC), Luis Sánchez (UC)
Reviewers	Jorge Lanza (UC), Benjamin Hebgen (NEC)

Abstract	This document, developed by the SALTED project, represents the D3.2 deliverable on the publication of linked and enriched data streams to the European Data Portal. The focus of this document is the description of the last step of the Scorpio Context Broker – CKAN connector regarding the publication of datasets into the EDP. Furthermore, D3.2 includes information on updates to component

This project is co-financed by the Connecting Europe Facility of the European Union under the Action Number 2020-EU-IA-0274.

	development since D3.1.
Keywords	CKAN, EDP, Open Data, DCAT-AP

Table of Contents

Table of Figures

Table of Textboxes

1 INTRODUCTION

1.1 Scope of Document

One of the most relevant objectives of SALTED is the publication of high-value data generated throughout the project in the European Data Portal (EDP). This document D3.2 outlines the final connector providing the interconnection of the SALTED architecture with the EDP by means of the supported harvesting process.

This document discusses the implementation process of such connector and the prior assurance processes necessary for the verification of its successful performance. Furthermore, as a result of these implementations, the quality metrics obtained after the publication of the data in the EDP are included. This document is focused on obtaining the expected outcome from Work Package 3, particularly Task 3.2.

1.2 Target Audience

This document is mainly intended for internal use, although it is publicly available in order to raise awareness of the SALTED Project and its contribution to the Open Data community through the EDP. The target audience is the SALTED technical team including all partners involved in the delivery of Work Packages 1, 2 and 4, and especially those involved in the activities of Work Package 3. The document provides a thorough definition of the infrastructure connecting the private SALTED development platform with the public endpoints that will allow users to access the data provided by the project.

1.3 Structure of the Document

First, an overview of the overall connector architecture presented in the previous deliverable [1] is given, with emphasis on the updates and progress arising from the development of the project in Section 2. Section 3 focuses on the last connector in the chain, which is in charge of linking the SALTED platform to the EDP, including the external validation processes of the subcomponents. Section 4 shows the results obtained after the publication of the high-value data in the EDP. Section 5 includes essential conclusions reached with the development and implementation of the components. Finally, an appendix comprises the Textboxes (examples of data and entity information models and representations) referenced throughout the document, in order to facilitate the reading due to their length.

2 SALTED PLATFORM TO EDP CONNECTOR IMPLEMENTATION

In this section a summary of the architecture proposed in Deliverable 3.1 [1] for the connector between the SALTED Platform and the EDP (European Data Portal) will be provided as a brief reminder. In addition, new updates of the components to cover more general use cases, which arise from the further development of the modules and adapt them to possible scenarios outside the SALTED Project, will be included.

2.1 Architecture Summary

As previously described in [1], the architecture corresponding to the communication between the SALTED Platform and the EDP consists of a series of connectors between the components involved. This architecture allows the publication of the enriched data stored in each of the so-called Satellite Scorpio Context Brokers through the Federation setup explained in Deliverable 2.2 [2].

As can be seen in Figure 1, which depicts the high-level architecture, the principal connectors are identified by the arrows between components. From left to right, the first arrow corresponds to the connector between the Dataset Registry and the Federator Context Broker (User to Context Broker connector), allowing direct user interaction with the generation and personalisation of datasets. The second arrow interconnects the Federator Context Broker with the CKAN portal (Context Broker to CKAN connector), transforming NGSI-LD entities to CKAN format and injecting them into the CKAN. And finally, the last arrow refers to the CKAN to EDP connector, enabling the necessary endpoints for the EDP to harvest the data collected in the CKAN. Details of the specific functionalities of all three connectors will be presented in the following sections, with an emphasis on the new features with respect to what was presented in [1].

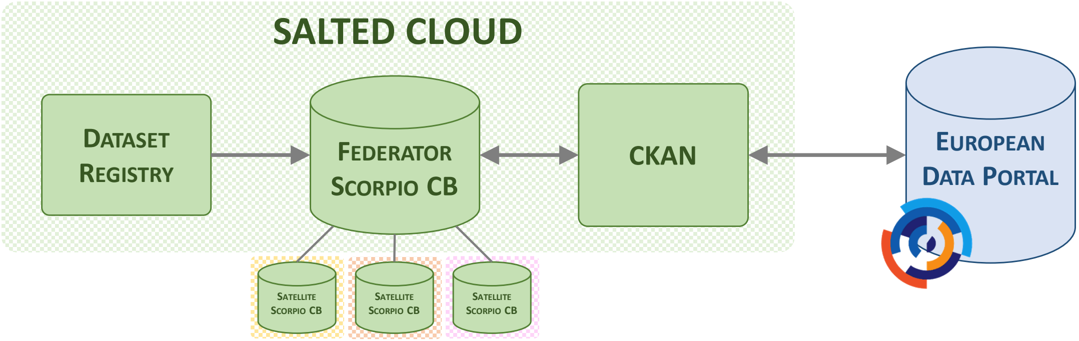


Figure 1. SALTED Platform to EDP connector architecture

2.2 Connector Updates

This section outlines the connectors involved in the architecture and presents the new developments compared to their first versions.

2.2.1 User to Context Broker Connector

The first connector is the User to Context Broker connector. In Figure 1 it is represented as the link between the Dataset Registry component and the Federator Context Broker. Figure 2 shows its architecture in greater detail, including the form available to the user and the NGSI-LD entities that are generated and injected into the local database of the Federator Context Broker.

The main purpose of this connector is the registration of the desired datasets, together with their necessary data representations (also known as distributions in the DCAT-AP terminology and resources in the CKAN terminology) as well as the context source registration of the Satellite Context Broker containing such data in the Federator Context Broker.

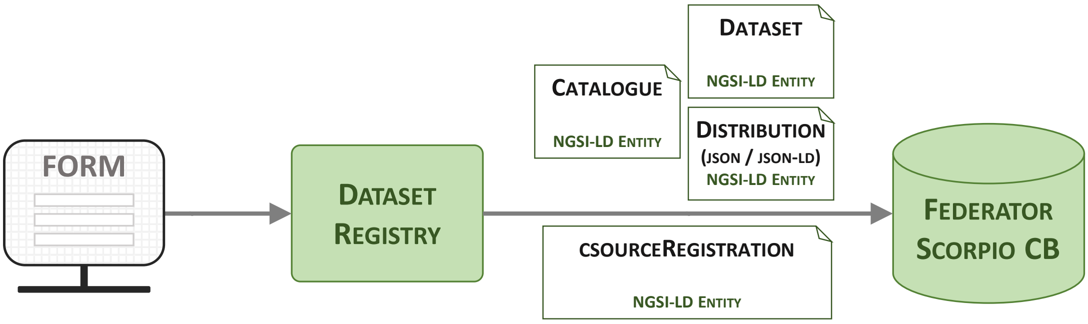


Figure 2. User to Context Broker connector architecture

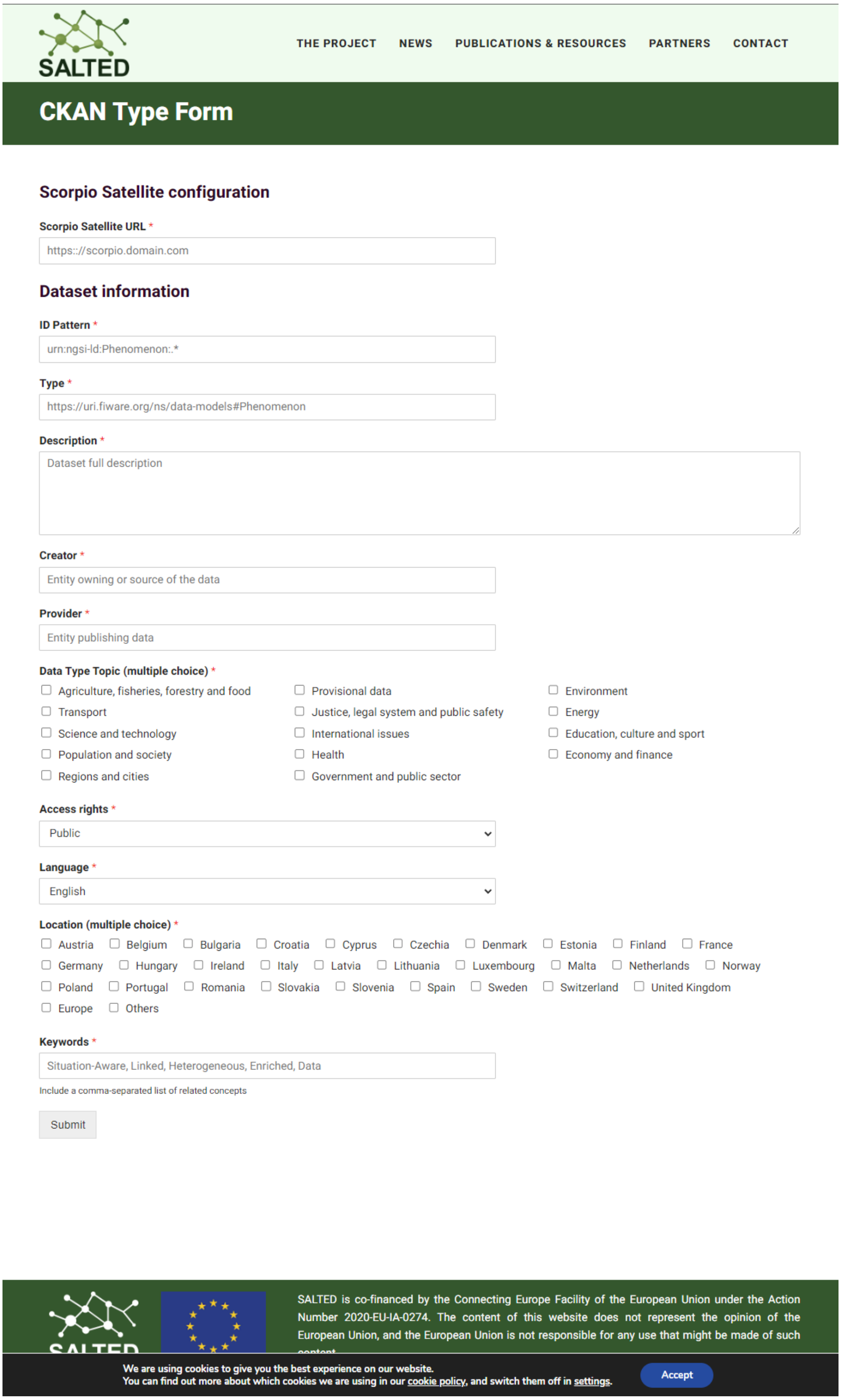
The workflow described in [1] remains in place, with some modifications and adjustments towards a more comprehensive and user customisable environment. The whole process starts with the web form, shown in Figure 3. As can be seen in the figure, the current fields are divided into two parts: Scorpio Satellite configuration and Dataset information.

The first part consists of the URL of the Satellite Context Broker where the data to be federated and, therefore, the data to be used to create a dataset are stored[1]. The second part, Dataset information, contains the necessary fields for the metadata description of the dataset, such as the type of data it contains (Type), the description itself or the chosen keywords. As can be seen, the end user is given a great deal of freedom of choice and customisation when defining the datasets.

Once the form has been filled in and submitted, the Dataset Registry module receives the data and generates the entities shown in Textbox 1, Textbox 2, Textbox 3 and Textbox 4 based on the information obtained. Regarding the first version described in [1], the progress made in the automation of these entities' creation and their use in the workflow can be appreciated.

Firstly, a `csourceRegistration`, in the current version of the platform, is dynamically generated using the information received by the form, as shown in Textbox 1. Thanks to this registration, data of the registered type `<type>` stored in the corresponding Satellite Context Broker will be available through the Federator Context Broker.

[1] It is worth noting that the approach followed for the generation of the datasets is based on the types of data (or the types of Smart Data Models used), i.e. an Environment/AirQualityObserved dataset, a Battery/BatteryStatus dataset...



CKAN Type Form

Scorpio Satellite configuration

Scorpio Satellite URL *

https://scorpio.domain.com

Dataset information

ID Pattern *

urn:ngsi-ld:Phenomenon:.*

Type *

https://uri.fiware.org/ns/data-models#Phenomenon

Description *

Dataset full description

Creator *

Entity owning or source of the data

Provider *

Entity publishing data

Data Type Topic (multiple choice) *

☐ Agriculture, fisheries, forestry and food
☐ Transport
☐ Science and technology
☐ Population and society
☐ Regions and cities
☐ Provisional data
☐ Justice, legal system and public safety
☐ International issues
☐ Health
☐ Government and public sector
☐ Environment
☐ Energy
☐ Education, culture and sport
☐ Economy and finance

Access rights *

Public

Language *

English

Location (multiple choice) *

☐ Austria ☐ Belgium ☐ Bulgaria ☐ Croatia ☐ Cyprus ☐ Czechia ☐ Denmark ☐ Estonia ☐ Finland ☐ France
☐ Germany ☐ Hungary ☐ Ireland ☐ Italy ☐ Latvia ☐ Lithuania ☐ Luxembourg ☐ Malta ☐ Netherlands ☐ Norway
☐ Poland ☐ Portugal ☐ Romania ☐ Slovakia ☐ Slovenia ☐ Spain ☐ Sweden ☐ Switzerland ☐ United Kingdom
☐ Europe ☐ Others

Keywords *

Situation-Aware, Linked, Heterogeneous, Enriched, Data

Include a comma-separated list of related concepts

SALTED is co-financed by the Connecting Europe Facility of the European Union under the Action Number 2020-EU-IA-0274. The content of this website does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

We are using cookies to give you the best experience on our website.
You can find out more about which cookies we are using in our cookie policy, and switch them off in settings.

Figure 3. SALTED CKAN webform[2]

[2] https://salted-project.eu/ckan-type-form/

Moving on to the other entities generated by the Dataset Registry module, comes `Catalogue`. `Catalogue` follows the data model proposed by the Smart Data Models initiative[3], and defines a catalogue of datasets following DCAT-AP [3]. As shown in Textbox 2, most of the fields are already set because the catalogue corresponds to the SALTED Project entity and the characteristics imposed by the project in terms of data access rights, licenses and access URLs. An interesting property is `dataset`, which creates a relation with all those datasets that are included in the catalogue. In the example, both `Environment:AirQualityObserved` and `Battery:BatteryStatus` datasets are included (`urn:ngsi-ld:Dataset:SALTED_Project:Environment:AirQualityObserved`, `urn:ngsi-ld:Dataset:SALTED_Project:Battery:BatteryStatus`, respectively).

Next is `Dataset`, which is the most user dependent entity due to the information submitted in the form. This data model type is also defined by the Smart Data Models initiative[4] and in Textbox 3 an example of its use is shown. As can be seen, this instance represents the `Environment:AirQualityObserved` dataset, which is one of those linked in the `Catalogue` entity. On the other hand, relations between the form fields and the properties of the data model are relatively straightforward. Analogously to the catalogue, a `distribution` property in relation to the two data representations currently offered by the development is included. These two representations are JSON and JSON-LD (`urn:ngsi-ld:Distribution:SALTED_Project:Environment:AirQualityObserved:json`, `urn:ngsi-ld:Distribution:SALTED_Project:Environment:AirQualityObserved:jsonld`, respectively). Nevertheless, it is foreseen for the near future to include other data formats as RDF, Turtle, N-Triples, etc.

Finally, one of the aforementioned data representations (JSON) is included in Textbox 4 as an example. The data model used is `Distribution` and is, like the previous two, defined by the Smart Data Models initiative[5]. In this case, the Dataset Registry module generates a `Distribution` entity for each of the representation formats available, providing information on properties such as `mediaType`, `format` and different URLs for accessing and downloading the data being described. These URLs correspond to the endpoints offered by the Retriever module, which will be explained further below.

By generating these entities, the dataset (and their distributions) definitions and descriptions linked to the Satellite Context Broker that offers them is ultimately registered in the Federator Context Broker and made available through the SALTED catalogue. This injection of entities into the Federator Context Broker will trigger a series of operations in subsequent modules (CKAN extensions) for the actual inclusion of these descriptions in this CKAN portal.

2.2.2 Context Broker to CKAN Connector

The second connector of the architecture is the Context Broker to CKAN connector, which is shown in Figure 4. The components involved are the Federator Context Broker, the CKAN portal with its NGSI-LD harvesting extension (`ckanext-ngsild-harvester`) and the Retriever module. Broadly speaking, the workflow between the components does not differ significantly from what is described in [1]. The CKAN extension module subscribes to `Dataset` entities in the Federator

[3] https://github.com/smart-data-models/dataModel.DCAT-AP/tree/master/Catalogue

[4] https://github.com/smart-data-models/dataModel.DCAT-AP/tree/master/Dataset

[5] https://github.com/smart-data-models/dataModel.DCAT-AP/tree/master/Distribution

Context Broker, so that each time a new entity is injected through the Dataset Registry module, described in Section 2.2.1, a notification containing that entity is received by the CKAN NGSI-LD harvesting module. By means of this information, this module transforms the catalogue, the dataset and the associated distributions from a Smart Data Model into corresponding CKAN format (organisation, package and resources) and injects them into the CKAN portal. As discussed in the distribution description (Section 2.2.1), some of its properties (`accessURL` and `downloadURL`) point to the endpoints made accessible by the Retriever module, which is intended to provide the data stored in the Satellite Context Brokers and available through the Federator Context Broker in different formats. Therefore, not only does the Retriever module act as a reverse proxy but it also transforms the data into multiple representations.

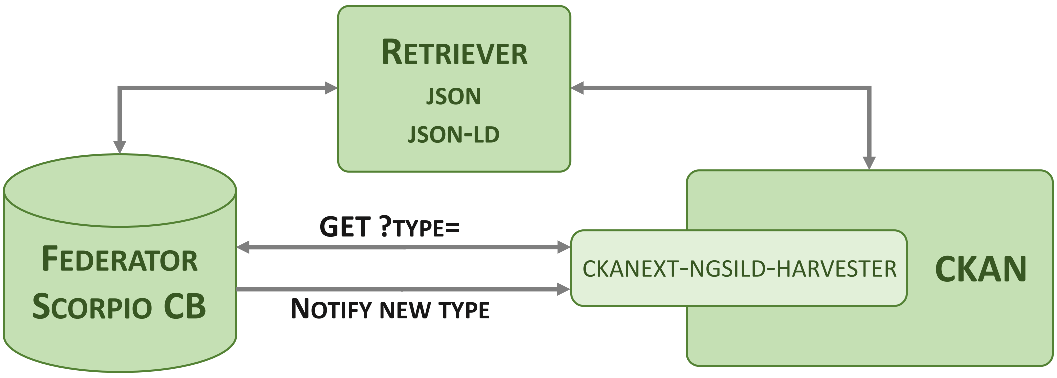


Figure 4. Context Broker to CKAN connector architecture

Going into more detail, there are additional steps following the progress in the development and automation of the modules compared to their first versions.

The first major point to note is the availability of a public IP address for accessing the CKAN portal: https://ckan.salted-project.eu/.

Secondly, there is the NGSI-LD CKAN harvester module deployed as a CKAN extension, `ckanext-ngsild-harvester`. This new version of this module currently adds three endpoints to the CKAN portal URL. It is noteworthy that requests to these endpoints are restricted to the system administrator.

- `/ngsi-ld/subscribe`: In order to make a request to this endpoint, certain access data must be included, shown in Textbox 5. Once this request is made, two operations are triggered: first, an initialisation in the CKAN portal of the catalogue, dataset and distribution information stored in the indicated Context Broker is performed. Next, the subscription shown in Textbox 6 is generated, through which the module will receive the notifications regarding the injection of new datasets. As you can see, the endpoint property is filled with the notification endpoint available by the module itself, which is explained below.
- `/nsgi-ld/unsubscribe`: Analogous to the previous endpoint, a request to this endpoint requires exactly the same parameters shown in Textbox 5, and is responsible for unsubscribing from the indicated Context Broker, stopping the reception of notifications due to the registration of new entities of type Dataset.
- `/ngsi-ld/notifications`: This last endpoint corresponds to the URL resource that receives the notifications from the Context Broker to which the module is subscribed.

When a notification arrives, it triggers the transformation to CKAN format and the creation (or update) of the package (dataset) together with its resources (distributions) in the CKAN portal.

As it has already been mentioned, in order to make certain requests to the CKAN NGSI-LD Harvester module, it is necessary to include specific parameters as well as these requests being restricted to the system administrator. Nevertheless, the CKAN NGSI-LD Harvester module has been designed to register the operations carried out on the CKAN portal (creation and update of organisations, packages and resources) under the username chosen by the information provider, i.e. the user who has filled in the form. For this purpose, the user is asked to contact the system administrator (SALTED Project) and provide the username wanted within the CKAN domain (`friendlyName`) and the name of the organisation to which the user wants to contribute (`organisation`). Based on these data, the administrator creates a CKAN user for this username (`friendlyName`) and generates the API Token (`ckan_token`) necessary to complete the request parameters required for the module.

Lastly, there is the Retriever module, which acts as a reverse proxy. Updates to this module have been focused on the available endpoints, including the `<subject>` parameter for full identification of the data type (Smart Data Model type) being requested.

- `/retriever/realtime/<subject>/<type>.<format>`
- `/retriever/temporal/<subject>/<type>.<format>?<temporal_unit>=<value>`
    - ○ `temporal_unit = [“year”, “months”, “weeks”, “days”, “hours”]`

Likewise, the `wrapped_contexts` used in the Link header of the Context Broker requests have been updated, and are available in the project's GitHub repository[6]. Rather than being split by `entityType`, they are currently divided by `subject`. Thus, the relationship between the request and the `@context` file to be used is done through this new parameter included in the available endpoints. Here is an example of how to use the Retriever module in the current version:

- Request sent to Retriever:

```
GET /retriever/realtime/<subject>/<type>.<format>
```

- Request internally generated sent to the Federated Context Broker:

```
GET /entities?type=<type>
Accept: application/<format>
Link:  <https://raw.githubusercontent.com/SALTED-
       Project/contexts/main/wrapped_contexts/<subject>-
       context.jsonld>;rel="http://www.w3.org/ns/json-
       ld#context";type="application/ld+json"
```

2.2.3 CKAN to EDP Connector

The last connector of the architecture, named CKAN to EDP connector and shown in Figure 5, links the CKAN portal within the SALTED domain to the EDP, allowing the publication of the SALTED data (descriptions of the existing datasets in CKAN) in the European Open Data Portal.

[6] https://github.com/SALTED-Project/contexts/tree/main/wrapped_contexts

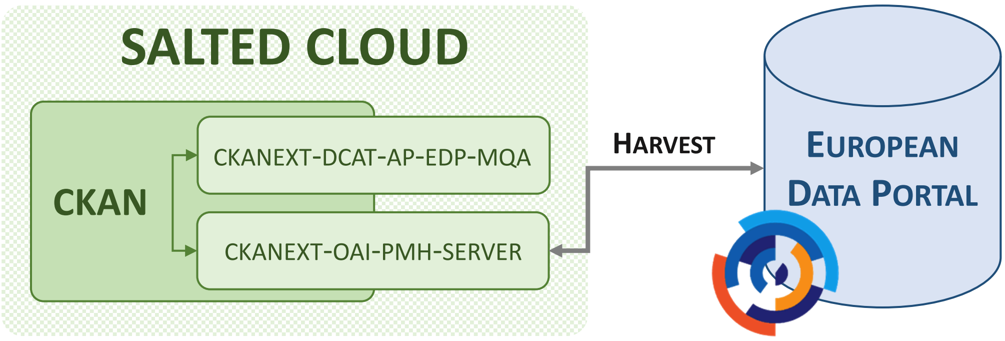


Figure 5. CKAN to EDP connector architecture

As already explained in [1], the implementation of this last connector consists of two new CKAN extensions, enabling the transformation of dataset descriptions in CKAN format to a DCAT-AP compliant format.

- `ckanext-dcat-ap-edp-mqa`: carrying out this first adaptation towards compatibility.
- `ckanext-oai-pmh-server`: offering an endpoint supporting the harvesting of these descriptions using the Open Archives Initiative Protocol for Metadata Harvesting[7] (OAI-PMH) in DCAT-AP format to be performed by the EDP.

Since the main objective of this deliverable is based on this final part of the workflow, the performance of these two extensions is further explained in the following section.

[7] https://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm

3 CKAN TO EDP CONNECTOR INSIGHTS

The CKAN to EDP connector enables these two final components of the workflow presented in Figure 1 to be bridged. Essentially, this connector is responsible for facilitating the import, publication and dissemination of SALTED data into the EDP. To this end, as shown in Figure 5, two extensions inherent to the CKAN have been developed, both of which enable the harvesting of the information stored in the CKAN portal in a DCAT-AP compatible format and through an endpoint that supports communication with the EDP.

Although these two extensions have already been described in [1], the following section summarises the operation of each of them and the results obtained with their use.

3.1 EXTENSIONS

3.1.1 ckanext-dcat-ap-edp-mqa

The purpose of this plugin is to transform and translate the descriptions of organisations, packages and resources stored in the CKAN portal from CKAN format to a DCAT-AP compliant format. Through this extension, full compatibility with DCAT-AP v2.1.0 [3] is achieved.

As discussed in the previous deliverable, this plugin is based on the well-known `ckanext-dcat`[8] extension, which allows the serialisation of data in different formats (RDF/XML, Turtle, Notation3 and JSON-LD). In addition, it offers two endpoints through which the description at organisation or package level is accessible. These endpoints are:

- `/dataset/<dataset_id>.<format>?profiles=<profile>`
- `/catalog.<format>?profiles=<profile>`

Regarding the profile parameter, it is reminded that a usage profile (`dcat-ap-edp-mqa`) has been developed to facilitate this desired full compatibility with the DCAT-AP v2.1.0 specification.

An example of a DCAT-AP compliant RDF representation of the `Environment:AirQualityObserved` dataset available in the CKAN is shown in Textbox 7. The request made to achieve this result is:
`https://ckan.salted-project.eu/dataset/environment_airqualityobserved.rdf?profiles=dcat_ap_edp_mqa`

This RDF body corresponds to the description of the `Environment:AirQualityObserved` dataset. This plugin, besides providing this type of export and representation of the datasets, has been developed (as the rest of the previous modules) with the aim of achieving the best quality score in the EDP through the Metadata Quality Assessment process[9]. As a means of checking the quality of the data prior to its import into the EDP, the MQA Scoring module[10] developed under the scope of the YODA Project[11], has been used. This module simulates the EDP MQA process, obtaining the score for each of the listed properties.

[8] https://github.com/ckan/ckanext-dcat
[9] https://dataeuropa.gitlab.io/data-provider-manual/metadata-quality-dashboard/#the-metadata-quality-assessment-process
[10] https://github.com/YourOpenDAta/mqa-scoring
[11] https://yoda.dit.upm.es/

Textbox 9 contains the result obtained by running this module on the RDF body exported from the `Environment:AirQualityObserved` dataset. As can be seen, a total score of 400 points is obtained out of an overall total of 405 points[12]. These 5 remaining points are due to the non-existence of the `ByteSize` property, since the resource's data is provided in real time and therefore it is not possible to estimate the size of the Context Broker's response in advance. Regardless of these 5 points, we are within the Excellent ranking (351-405 points).

3.1.2 ckanext-oai-pmh-server

This plugin aims to provide an accessible and compatible endpoint for the EDP harvesting process. It is based on OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting), which allows data providers to expose their metadata and consumers to use this metadata to build value-added services.

Hence, this plugin allows the descriptions (metadata) of the packages (datasets) stored in the CKAN portal to be available and accessible for the harvesting process and subsequent publication in the EDP.

In [1] all relevant plugin endpoints were described. However, the EDP will only make use of one of them, `ListRecords`, which allows the listing of package descriptions:

`https://ckan.salted-project.eu/oai?verb=ListRecords&metadataPrefix=dcat`

(Successive requests with `resumptionToken`: `https://ckan.salted-project.eu/oai?verb=ListRecords&metadataPrefix=dcat&resumptionToken=<resumptionToken>`)

Assuming for the sake of simplicity that only the `Environment:AirQualityObserved` dataset is available in the CKAN portal, the response to launching a request against this URL returns the RDF body shown in Textbox 8. This is the information needed by the EDP for the publication of these datasets.

As a means to check if the plugin is properly developed and provides the metadata as expected, the data.europa.eu team facilitates the `importing-oaipmh` module[13], which simulates the harvesting process. Deploying this module on localhost, the request described in Textbox 10 is performed, obtaining the logs shown in Textbox 11.

The most important parameters in the request (Textbox 10) are those contained under the `config` property: `address`, `metadata`, `catalogue` and `outputFormat`. These parameters specify the base URL of our plugin, the catalogue/organisation to be harvested and the metadata format. Regarding `outputFormat`, the plugin code internally links `metadataPrefix=dcat` and the previously mentioned `dcat-ap-edp-mqa` profile. This link enables a coherent combination of both plugins, so that the descriptions consumed by the EDP are fully compatible with DCAT-AP v2.1.0.

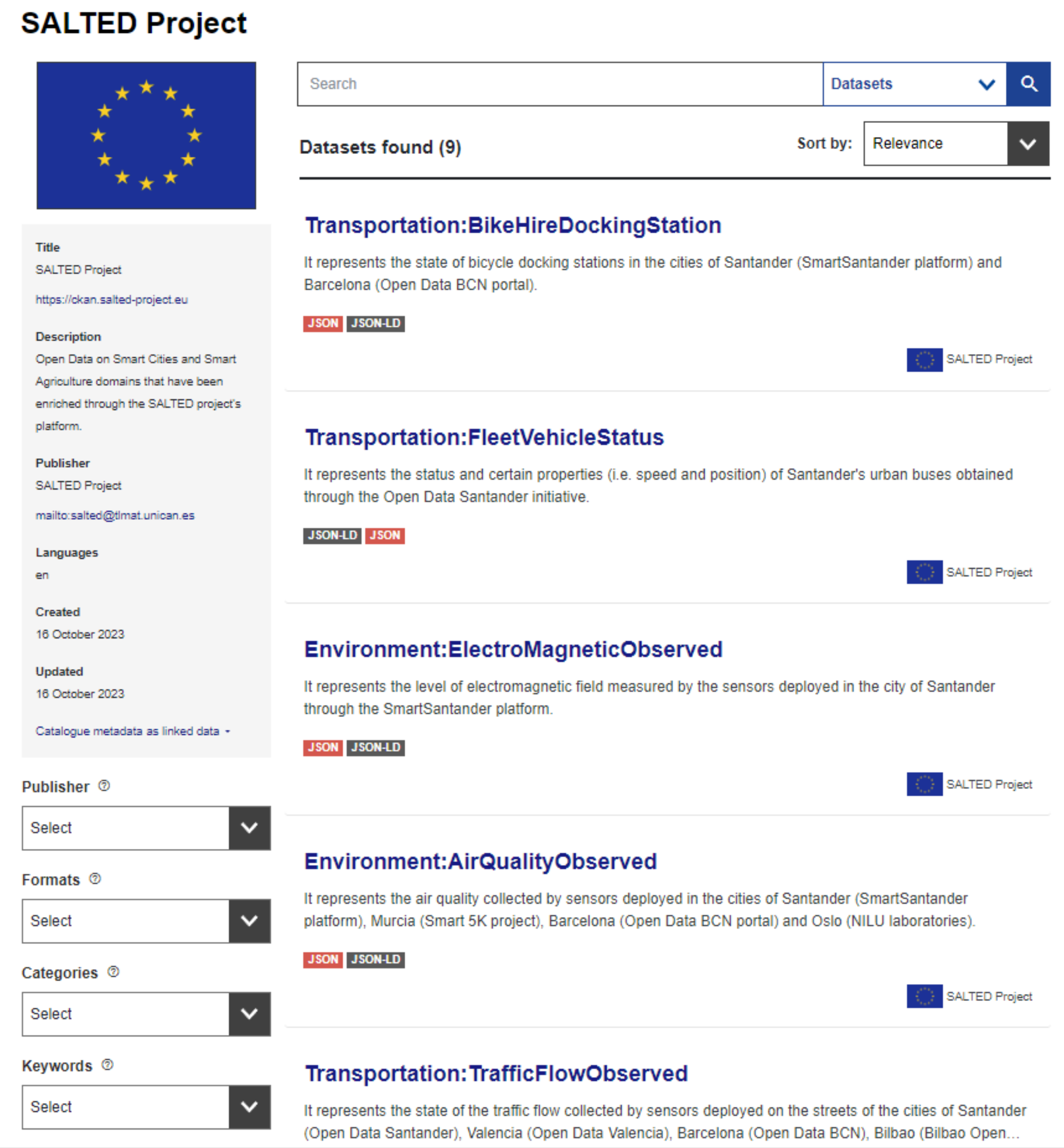
Moreover, the module logs (Textbox 11) obtained with the triggering of the request show that the harvesting or metadata importing process has been successfully completed.

[12] https://dataeuropa.gitlab.io/data-provider-manual/metadata-quality-dashboard/#rating

[13] https://gitlab.com/dataeuropa/harvester/importing-oaipmh

4 PUBLICATION TO EDP

This section presents the Salted catalogue published in the EDP. Through the whole chain of connectors described in the previous sections, the publication of the high-value data generated in the enrichment pipeline deployed in the project has been achieved. The main page of the published SALTED catalogue is available (as shown in Figure 6), with a complete migration of the information and descriptions stored in the CKAN instance of the project.



SALTED Project

Search | Datasets

Datasets found (9) Sort by: Relevance

Title
SALTED Project
https://ckan.salted-project.eu

Description
Open Data on Smart Cities and Smart Agriculture domains that have been enriched through the SALTED project's platform.

Publisher
SALTED Project
mailto:salted@tlmat.unican.es

Languages
en

Created
16 October 2023

Updated
16 October 2023

Catalogue metadata as linked data

Publisher: Select

Formats: Select

Categories: Select

Keywords: Select

Transportation:BikeHireDockingStation
It represents the state of bicycle docking stations in the cities of Santander (SmartSantander platform) and Barcelona (Open Data BCN portal).
JSON JSON-LD
SALTED Project

Transportation:FleetVehicleStatus
It represents the status and certain properties (i.e. speed and position) of Santander's urban buses obtained through the Open Data Santander initiative.
JSON-LD JSON
SALTED Project

Environment:ElectroMagneticObserved
It represents the level of electromagnetic field measured by the sensors deployed in the city of Santander through the SmartSantander platform.
JSON JSON-LD
SALTED Project

Environment:AirQualityObserved
It represents the air quality collected by sensors deployed in the cities of Santander (SmartSantander platform), Murcia (Smart 5K project), Barcelona (Open Data BCN portal) and Oslo (NILU laboratories).
JSON JSON-LD
SALTED Project

Transportation:TrafficFlowObserved
It represents the state of the traffic flow collected by sensors deployed on the streets of the cities of Santander (Open Data Santander), Valencia (Open Data Valencia), Barcelona (Open Data BCN), Bilbao (Bilbao Open...

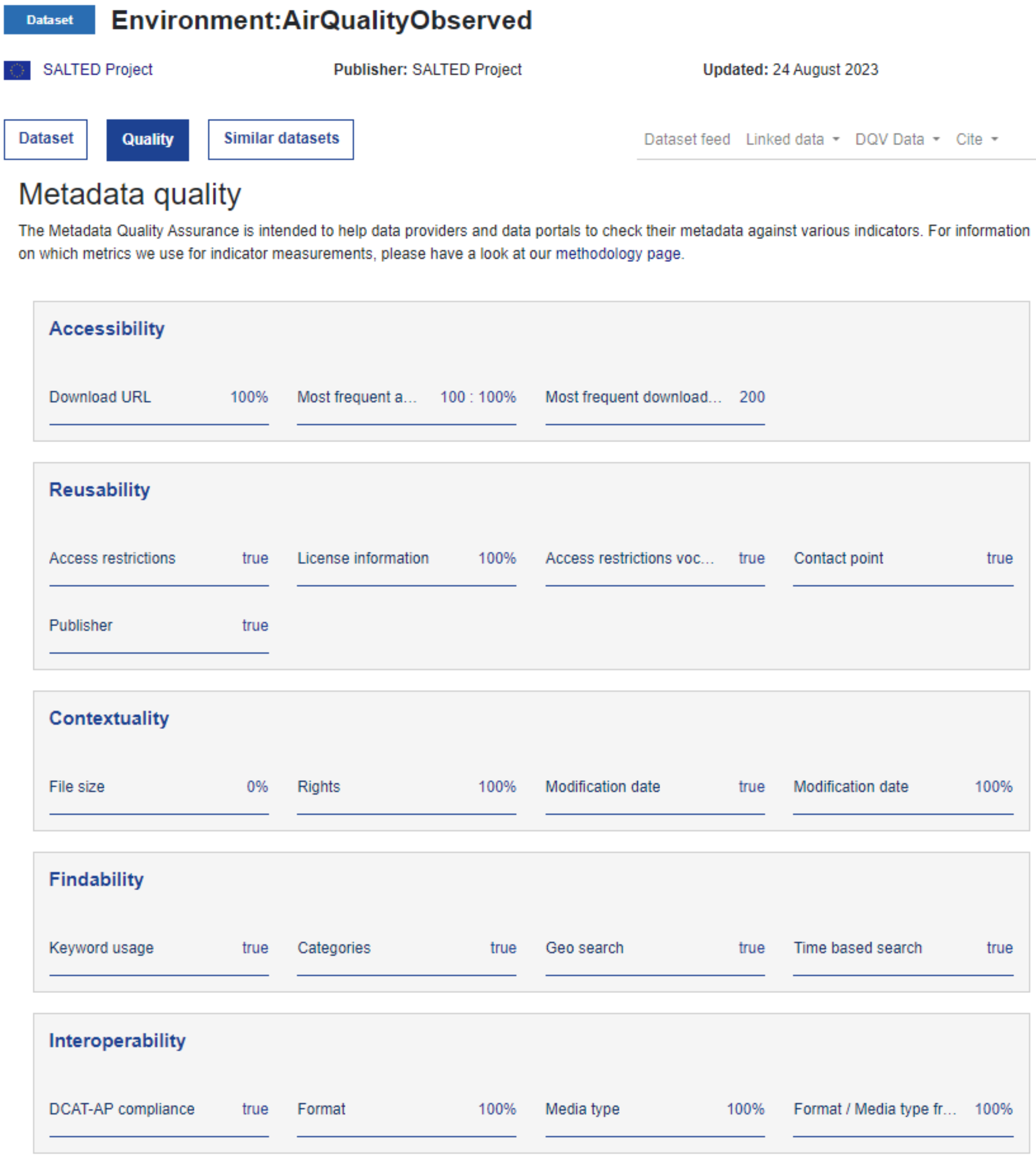
Figure 6. SALTED Project catalogue in the EDP[14]

Currently, there are 9 datasets, with more to come in the upcoming days. Every one of these datasets is available in JSON and JSON-LD formats, as outlined in previous sections.

[14] https://data.europa.eu/data/catalogues/salted?locale=en

One of the crucial features offered by the EDP is the metadata quality assessment of the datasets included in the catalogues. Figure 7 shows an example of the results obtained in one of our datasets (Environment:AirQualityObserved). As can be seen, 100% is achieved in all metrics (or *true* value otherwise) except File Size (*ByteSize*). This is the expected behaviour as discussed at the end of Section 3.1.1.

Another relevant characteristic of the EDP allows the visualisation of the metadata quality at catalogue level. Currently, this feature is not available for our SALTED catalogue, as the computation of such metadata by the European Data Portal is not done very frequently, but it is expected to be introduced in the upcoming weeks.



Dataset **Environment:AirQualityObserved**

SALTED Project **Publisher:** SALTED Project **Updated:** 24 August 2023

Metadata quality

The Metadata Quality Assurance is intended to help data providers and data portals to check their metadata against various indicators. For information on which metrics we use for indicator measurements, please have a look at our methodology page.

Accessibility

Download URL 100%
Most frequent a... 100 : 100%
Most frequent download... 200

Reusability

Access restrictions true
License information 100%
Access restrictions voc... true
Contact point true
Publisher true

Contextuality

File size 0%
Rights 100%
Modification date true
Modification date 100%

Findability

Keyword usage true
Categories true
Geo search true
Time based search true

Interoperability

DCAT-AP compliance true
Format 100%
Media type 100%
Format / Media type fr... 100%

Figure 7. Metadata quality results of Environment:AirQualityObserved dataset[15]

[15] https://data.europa.eu/data/datasets/f836d31d-045d-4182-bc31-9242278ef8b2/quality?locale=en

5 CONCLUSIONS

In this document we provide the D3.2 deliverable on the publication of SALTED data to the European Data Portal, focusing on the CKAN to EDP connector within the architecture presented previously. To this end, changes and updates to each of the components of the overall architecture have first been described. One remarkable aspect is that the generation and creation process of the final datasets has been significantly automated, allowing the user greater decision-making power over the attributes that describe them.

Moving on to the core of the document, the CKAN to EDP connector, we describe further details of the two extensions developed for the CKAN instance and which integrate the bridge with the EDP. Due to these extensions, the descriptions of the organisations, packages and distributions are represented in formats fully compatible with DCAT-AP v2.1.0 and can be accessed through an endpoint that supports the harvesting of such metadata, finally achieving a successful publication.

As a result of this fruitful work, we have included a set of samples capturing the publication of our datasets in the EDP, as well as the quality metrics obtained after the MQA process. Thanks to this assessment procedure, as well as prior quality checks, it is proven that the automation of the components together with the thorough development of the `ckanext-dcat-ap-edp-mqa` extension provide a complete RDF corpus as regards the descriptive metadata of the datasets.

This deliverable provides the culmination of the progress of Work Package 3. The compilation of this text with the previous deliverable (D3.1) composes an extensive and readable documentation of the proposed architecture developed from the requirements imposed. Furthermore, considering the current development and implementation of the architecture, it cannot be denied that the key milestone of the Work Package 3 is more than achieved.

6 BIBLIOGRAPHY

[1] SALTED, “D3.1: Report on development of Scorpio Context Broker - CKAN connector,” 2023.

[2] SALTED, “D2.2: Report on data modelling and linking,” 2023.

[3] European Commission, “DCAT Application Profile for data portals in Europe. Version 2.1.0.,” 2021.

7 APPENDIX

```
{
  "id": "urn:ngsi-ld:ContextSourceRegistration:<provider>",
  "type": "ContextSourceRegistration",
  "endpoint": <scorpiosatellite_endpoint>,
  "information": [
    {
      "entities": [
        {
          "type": <type>,
          "idPattern": <idPattern>
        }
       ]
     }
   ],
   "@context": [
     "https://raw.githubusercontent.com/SALTED-
Project/contexts/main/wrapped_contexts/environment-context.jsonld",
     "https://uri.etsi.org/ngsi-ld/v1/ngsi-ld-core-context-v1.6.jsonld"
   ]
}
```

Textbox 1. csourceRegistration entity example

```
{
  "id": "urn:ngsi-ld:Catalogue:SALTED_Project",
  "type": "Catalogue",
  "description": {
    "type": "Property",
    "value": "This is the SALTED Project Catalogue"
  },
  "title": {
    "type": "Property",
    "value": "SALTED Project"
  },
  "dataset": {
    "type": "Relationship",
    "object": [
      "urn:ngsi-ld:Dataset:SALTED_Project:Battery:BatteryStatus",
      "urn:ngsi-ld:Dataset:SALTED_Project:Environment:AirQualityObserved"
    ]
  },
  "homepage": {
    "type": "Property",
    "value": "https://salted-project.eu"
  },
  "licence": {
    "type": "Property",
    "value": "https://creativecommons.org/licenses/by/4.0/"
  },
  "publisher": {
    "type": "Property",
    "value": "salted-project"
  },
  "rights": {
    "type": "Property",
    "value": "PUBLIC"
  },
  "@context": [
    "https://raw.githubusercontent.com/SALTED-
Project/contexts/main/wrapped_contexts/dcat-ap-context.jsonld",
    "https://uri.etsi.org/ngsi-ld/v1/ngsi-ld-core-context-v1.6.jsonld"
  ]
}
```

Textbox 2. Catalogue entity example

```
{
  "id": "urn:ngsi-ld:Dataset:SALTED_Project:Environment:AirQualityObserved",
  "type": "Dataset",
  "description": {
    "type": "Property",
    "value": "It represents the air quality collected by sensors deployed in the ci
ties of Santander (SmartSantander platform), Murcia (Smart 5K project), Barcelona (
Open Data BCN portal) and Oslo (NILU laboratories)."
  },
  "title": {
    "type": "Property",
    "value": "Environment:AirQualityObserved"
  },
  "Type": {
    "type": "Property",
    "value": "Environment/AirQualityObserved"
  },
  "accessRights": {
    "type": "Property",
    "value": "PUBLIC"
  },
  "creator": {
    "type": "Property",
    "value": ["SmartSantander", "Smart 5K", "Open Data BCN", "Open Data", "NILU"]
  },
  "distribution": {
    "type": "Relationship",
    "object": [
      "urn:ngsi-
ld:Distribution:SALTED_Project:Environment:AirQualityObserved:json",
      "urn:ngsi-
ld:Distribution:SALTED_Project:Environment:AirQualityObserved:jsonld"
    ]
  },
  "keyword": {
    "type": "Property",
    "value": ["AirQualityObserved", "Air Quality", "SmartSantander", "Smart 5K",
"Open Data BCN", "Open Data", "NILU"]
  },
  "landingPage": {
    "type": "Property",
    "value": "https://salted-project.eu/"
  },
  "language": {
    "type": "Property",
    "value": "ENG"
  },
  "license": {
    "type": "Property",
    "value": "https://creativecommons.org/licenses/by/4.0/"
  },
  "publisher": {
    "type": "Property",
    "value": "salted-project"
  },
  "spatial": {
    "type": "Property",
    "value": ["NOR", "ESP"]
  },
  "temporal": {
    "type": "Property",
    "value": "2023-08-24T11:57:50.888311"
  },
  "theme": {
    "type": "Property",
    "value": "ENVI"
  },
  "version": {
    "type": "Property",
    "value": "1.0"
  },
  "dataProvider": {
    "type": "Property",
    "value": "UC"
  },
  "dateCreated": {
    "type": "Property",
    "value": "2023-08-24T11:57:51.121703"
  },
  "dateModified": {
    "type": "Property",
    "value": "2023-08-24T11:57:51.121739"
  },
  "@context": [
    "https://raw.githubusercontent.com/SALTED-
Project/contexts/main/wrapped_contexts/dcat-ap-context.jsonld",
    "https://uri.etsi.org/ngsi-ld/v1/ngsi-ld-core-context-v1.6.jsonld"
    ]
}
```

Textbox 3. Dataset entity example

```
{
  "id": "urn:ngsi-
ld:Distribution:SALTED_Project:Environment:AirQualityObserved:json",
  "type": "Distribution",
  "description": {
    "type": "Property",
    "value": "Realtime information in JSON format"
  },
  "title": {
    "type": "Property",
    "value": "Environment:AirQualityObserved realtime data in JSON format"
  },
  "accessUrl": {
    "type": "Property",
    "value": "https://ckan.salted-
project.eu/retriever/realtime/Environment/AirQualityObserved.json"
  },
  "availability": {
    "type": "Property",
    "value": "FOREVER"
  },
  "mediaType": {
    "type": "Property",
    "value": "application/json"
  },
  "rights": {
    "type": "Property",
    "value": "PUBLIC"
  },
  "dateCreated": {
    "type": "Property",
    "value": "2023-08-24T11:57:51.121703"
  },
  "dateModified": {
    "type": "Property",
    "value": "2023-08-24T11:57:51.121703"
  },
  "downloadUrl": {
    "type": "Property",
    "value": "https://ckan.salted-
project.eu/retriever/realtime/Environment/AirQualityObserved.json"
  },
  "format": {
    "type": "Property",
    "value": "JSON"
  },
  "@context": [
    "https://raw.githubusercontent.com/SALTED-
Project/contexts/main/wrapped_contexts/dcat-ap-context.jsonld",
    "https://uri.etsi.org/ngsi-ld/v1/ngsi-ld-core-context-v1.6.jsonld"
  ]
}
```

Textbox 4. Distribution entity example

```
{
  "hostname": <hostname or IP of the Federator Context Broker>,
  "port": <port of the Federator Context Broker>,
  "friendlyName": <CKAN username>,
  "ckan_token": <CKAN API Token>,
  "organization": <organization name>
}
```

Textbox 5. ckanext-ngsild-haverter enpoints parameters

```
{
  "id": "urn:ngsi-ld:Subscription:CKAN:salted_project:<friendlyName>",
  "type": "Subscription",
  "description": "Notify me on new datasets",
  "entities": [
    {
      "type": "https://smartdatamodels.org/dataModel.DCAT-AP/Dataset"
    }
  ],
  "isActive": true,
  "notification": {
    "endpoint": {
      "accept": "application/ld+json",
      "receiverInfo": [
        {"Authorization": <CKAN API Token>},
        {"X-CKAN-Organization": <organization>},
        {"X-NGSILD-Broker-Host": <hostname>},
        {"X-NGSILD-Broker-Port": <port>}
      ],
      "uri": "https://ckan.salted-project.eu/ngsi-ld/notifications"
    },
    "format": "normalized"
  },
  "subscriptionName": "CKAN subscription for <friendlyName>
and organization SALTED_Project",
  "@context": ["https://uri.etsi.org/ngsi-ld/v1/ngsi-ld-core-context-v1.6.jsonld"]
}
```

Textbox 6. Subscription example

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:dct="http://purl.org/dc/terms/"
  xmlns:vcard="http://www.w3.org/2006/vcard/ns#"
  xmlns:dcat="http://www.w3.org/ns/dcat#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:schema1="http://schema.org/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
>
  <dcat:Dataset rdf:about="https://ckan.salted-project.eu/dataset/f836d31d-045d-
4182-bc31-9242278ef8b2">
    <dct:title>Environment:AirQualityObserved</dct:title>
    <dct:description>It represents the air quality collected by sensors deployed in
 the cities of Santander (SmartSantander platform), Murcia (Smart 5K project), Barc
elona (Open Data BCN portal) and Oslo (NILU laboratories).</dct:description>
    <dcat:landingPage rdf:resource="https://salted-project.eu/"/>
    <dct:identifier>f836d31d-045d-4182-bc31-9242278ef8b2</dct:identifier>
    <owl:versionInfo>1.0</owl:versionInfo>
    <dct:accessRights rdf:resource="http://publications.europa.eu/resource/authorit
y/access-right/PUBLIC"/>
    <dcat:keyword>Air Quality</dcat:keyword>
    <dcat:keyword>AirQualityObserved</dcat:keyword>
    <dcat:keyword>NILU</dcat:keyword>
    <dcat:keyword>Open Data</dcat:keyword>
    <dcat:keyword>Open Data BCN</dcat:keyword>
    <dcat:keyword>Smart 5K</dcat:keyword>
    <dcat:keyword>SmartSantander</dcat:keyword>
    <dct:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2023-08-
24T11:57:51.121703</dct:issued>
    <dct:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2023-08-
24T11:57:51.121739</dct:modified>
    <dct:language rdf:resource="http://publications.europa.eu/resource/authority/
language/ENG"/>
    <dcat:theme rdf:resource="http://publications.europa.eu/resource/authority/data
-theme/ENVI"/>
    <dcat:contactPoint>
      <vcard:Organization rdf:nodeID="N8680108ca4ce4f869e805031d853bc0d">
        <vcard:fn>UC</vcard:fn>
      </vcard:Organization>
    </dcat:contactPoint>
    <dct:publisher>
      <foaf:Organization rdf:about="https://ckan.salted-
project.eu/organization/urn:ngsi-ld:Catalogue:SALTED_Project">
        <foaf:name>SALTED Project</foaf:name>
      </foaf:Organization>
    </dct:publisher>
    <dct:temporal>
      <dct:PeriodOfTime rdf:nodeID="N27ce577366764c21b77efe6564021f4d">
        <schema1:startDate rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime"
>2023-08-24T11:57:50.888311</schema1:startDate>
      </dct:PeriodOfTime>
    </dct:temporal>
    <dct:temporal>
      <dct:PeriodOfTime rdf:nodeID="N0f275bb542fd4312b33e174e84bdb4d4">
        <dcat:startDate rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">20
23-08-24T11:57:50.888311</dcat:startDate>
      </dct:PeriodOfTime>
    </dct:temporal>
    <dct:spatial rdf:resource="http://publications.europa.eu/resource/authority/
country/NOR"/>
    <dct:spatial rdf:resource="http://publications.europa.eu/resource/authority/
country/ESP"/>
    <dcat:distribution>
      <dcat:Distribution rdf:about="https://ckan.salted-
project.eu/dataset/f836d31d-045d-4182-bc31-9242278ef8b2/resource/urn_ngsi-
ld_distribution_salted_project_environment_airqualityobserved_json">
        <dct:title>
          environment:airqualityobserved_realtime_data_in_json_format
        </dct:title>
        <dct:description>Realtime information in JSON format</dct:description>
        <dct:rights>PUBLIC</dct:rights>
        <dcat:accessURL rdf:resource="https://ckan.salted-
project.eu/retriever/realtime/Environment/AirQualityObserved.json"/>
        <dcat:downloadURL rdf:resource="https://ckan.salted-
project.eu/retriever/realtime/Environment/AirQualityObserved.json"/>
        <dct:license rdf:resource="https://creativecommons.org/licenses/by/4.0/"/>
        <dcat:mediaType rdf:resource="https://www.iana.org/assignments/media-
types/application/json"/>
        <dct:format rdf:resource="http://publications.europa.eu/resource/authority/
file-type/JSON"/>
        <dct:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2023-
08-24T11:57:51.121703</dct:issued>
        <dct:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2023
-08-24T11:57:52.916649</dct:modified>
      </dcat:Distribution>
    </dcat:distribution>
    <dcat:distribution>
```

```
      <dcat:Distribution rdf:about="https://ckan.salted-
project.eu/dataset/f836d31d-045d-4182-bc31-9242278ef8b2/resource/urn_ngsi-
ld_distribution_salted_project_environment_airqualityobserved_jsonld">
        <dct:title>
          environment:airqualityobserved_realtime_data_in_json_ld_format
        </dct:title>
        <dct:description>Realtime information in JSON_LD format</dct:description>
        <dct:rights>PUBLIC</dct:rights>
        <dcat:accessURL rdf:resource="https://ckan.salted-
project.eu/retriever/realtime/Environment/AirQualityObserved.jsonld"/>
        <dcat:downloadURL rdf:resource="https://ckan.salted-
project.eu/retriever/realtime/Environment/AirQualityObserved.jsonld"/>
        <dct:license rdf:resource="https://creativecommons.org/licenses/by/4.0/"/>
        <dcat:mediaType rdf:resource="https://www.iana.org/assignments/media-
types/application/ld+json"/>
        <dct:format rdf:resource="http://publications.europa.eu/resource/authority/
file-type/JSON_LD"/>
        <dct:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2023-
08-24T11:57:51.121703</dct:issued>
        <dct:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2023
-08-24T11:57:52.917815</dct:modified>
      </dcat:Distribution>
    </dcat:distribution>
  </dcat:Dataset>
</rdf:RDF>
```

Textbox 7. RDF response when exporting the Environment:AirQualityObserved dataset with dcat_ap_edp_mqa profile

```
<OAI-PMH
  xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2023-09-05T14:28:12Z</responseDate>
  <request verb="ListRecords" metadataPrefix="dcat">
    https://ckan.salted-project.eu/oai
  </request>
  <ListRecords>
    <record>
      <header>
        <identifier>f836d31d-045d-4182-bc31-9242278ef8b2</identifier>
        <datestamp>2023-08-24T11:57:52Z</datestamp>
        <setSpec>salted_project</setSpec>
      </header>
      <metadata>
        <rdf:RDF xmlns:dct="http://purl.org/dc/terms/"
          xmlns:dcat="http://www.w3.org/ns/dcat#"
          xmlns:vcard="http://www.w3.org/2006/vcard/ns#"
          xmlns:foaf="http://xmlns.com/foaf/0.1/"
          xmlns:locn="http://www.w3.org/ns/locn#"
          xmlns:owl="http://www.w3.org/2002/07/owl#"
          xmlns:schema1="http://schema.org/"
          xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
          <dcat:Dataset rdf:about="https://ckan.salted-project.eu/dataset/f836d31d-
045d-4182-bc31-9242278ef8b2">
            <dct:title>Environment:AirQualityObserved</dct:title>
            <dct:description>
              It represents the air quality collected by sensors deployed in the
              cities of Santander (SmartSantander platform), Murcia (Smart 5K
              project), Barcelona (Open Data BCN portal) and Oslo (NILU
              laboratories).
            </dct:description>
            <dcat:landingPage rdf:resource="https://salted-project.eu/"/>
            <dct:identifier>f836d31d-045d-4182-bc31-9242278ef8b2</dct:identifier>
            <owl:versionInfo>1.0</owl:versionInfo>
            <dct:accessRights rdf:resource="http://publications.europa.eu/resource/
authority/access-right/PUBLIC"/>
            <dcat:keyword>Air Quality</dcat:keyword>
            <dcat:keyword>AirQualityObserved</dcat:keyword>
            <dcat:keyword>NILU</dcat:keyword>
            <dcat:keyword>Open Data</dcat:keyword>
            <dcat:keyword>Open Data BCN</dcat:keyword>
            <dcat:keyword>Smart 5K</dcat:keyword>
            <dcat:keyword>SmartSantander</dcat:keyword>
            <dct:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">
              2023-08-24T11:57:51.121703
            </dct:issued>
            <dct:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">
              2023-08-24T11:57:51.121739
            </dct:modified>
            <dct:language rdf:resource="http://publications.europa.eu/resource/
authority/language/ENG"/>
            <dcat:theme rdf:resource="http://publications.europa.eu/resource/author
ity/data-theme/ENVI"/>
            <dcat:contactPoint>
              <vcard:Organization rdf:nodeID="N96dc70b771f14693852498076547c2a8">
                <vcard:fn>UC</vcard:fn>
              </vcard:Organization>
            </dcat:contactPoint>
            <dct:publisher>
              <foaf:Organization rdf:about="https://ckan.salted-
project.eu/organization/urn:ngsi-ld:Catalogue:SALTED_Project">
                <foaf:name>SALTED Project</foaf:name>
              </foaf:Organization>
            </dct:publisher>
            <dct:temporal>
              <dct:PeriodOfTime rdf:nodeID="Nfd4e65ce17cd463ebe6aac4961445e06">
                <schema1:startDate rdf:datatype="http://www.w3.org/2001/XMLSchema#d
ateTime">2023-08-24T11:57:50.888311</schema1:startDate>
              </dct:PeriodOfTime>
            </dct:temporal>
            <dct:temporal>
              <dct:PeriodOfTime rdf:nodeID="Nc9e40b5c26b24385bd20cfd81b7d5396">
                <dcat:startDate rdf:datatype="http://www.w3.org/2001/XMLSchema#date
Time">2023-08-24T11:57:50.888311</dcat:startDate>
              </dct:PeriodOfTime>
            </dct:temporal>
            <dct:spatial rdf:resource="http://publications.europa.eu/resource/
authority/country/NOR"/>
            <dct:spatial rdf:resource="http://publications.europa.eu/resource/
authority/country/ESP"/>
            <dcat:distribution>
              <dcat:Distribution rdf:about="https://ckan.salted-
project.eu/dataset/f836d31d-045d-4182-bc31-9242278ef8b2/resource/urn_ngsi-
ld_distribution_salted_project_environment_airqualityobserved_json">
                <dct:title>
                  environment:airqualityobserved_realtime_data_in_json_format
                </dct:title>
                <dct:description>
                  Realtime information in JSON format
                </dct:description>
                <dct:rights>PUBLIC</dct:rights>
                <dcat:accessURL rdf:resource="https://ckan.salted-
project.eu/retriever/realtime/Environment/AirQualityObserved.json"/>
                <dcat:downloadURL rdf:resource="https://ckan.salted-
project.eu/retriever/realtime/Environment/AirQualityObserved.json"/>
                <dct:license rdf:resource="https://creativecommons.org/licenses/by/
4.0/"/>
                <dcat:mediaType rdf:resource="https://www.iana.org/assignments/medi
a-types/application/json"/>
                <dct:format rdf:resource="http://publications.europa.eu/resource/au
thority/file-type/JSON"/>
                <dct:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime
">2023-08-24T11:57:51.121703</dct:issued>
                <dct:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTi
me">2023-08-24T11:57:52.916649</dct:modified>
              </dcat:Distribution>
            </dcat:distribution>
            <dcat:distribution>
```

```
              <dcat:Distribution rdf:about="https://ckan.salted-
project.eu/dataset/f836d31d-045d-4182-bc31-9242278ef8b2/resource/urn_ngsi-
ld_distribution_salted_project_environment_airqualityobserved_jsonld">
                <dct:title>
                  environment:airqualityobserved_realtime_data_in_json_ld_format
                </dct:title>
                <dct:description>
                  Realtime information in JSON_LD format
                </dct:description>
                <dct:rights>PUBLIC</dct:rights>
                <dcat:accessURL rdf:resource="https://ckan.salted-
project.eu/retriever/realtime/Environment/AirQualityObserved.jsonld"/>
                <dcat:downloadURL rdf:resource="https://ckan.salted-
project.eu/retriever/realtime/Environment/AirQualityObserved.jsonld"/>
                <dct:license rdf:resource="https://creativecommons.org/licenses/by/
4.0/"/>
                <dcat:mediaType rdf:resource="https://www.iana.org/assignments/medi
a-types/application/ld+json"/>
                <dct:format rdf:resource="http://publications.europa.eu/resource/au
thority/file-type/JSON_LD"/>
                <dct:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime
">2023-08-24T11:57:51.121703</dct:issued>
                <dct:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTi
me">2023-08-24T11:57:52.917815</dct:modified>
              </dcat:Distribution>
            </dcat:distribution>
          </dcat:Dataset>
        </rdf:RDF>
      </metadata>
    </record>
  </ListRecords>
</OAI-PMH>
```

Textbox 8. XML/RDF response to request perform to /oai endpoint

```
  * SHACL validation
     Result:  OK.  The  metadata  has  successfully  passed  the  EDP  validator.  Weight
assigned 30
     Current weight = 30
  * dcat:landingPage
     Result: WARN. Not included in MQA - dcat:landingPage
     Current weight = 30
  * dct:title
     Result: WARN. Not included in MQA - dct:title
     Result: WARN. Not included in MQA - dct:title
     Result: WARN. Not included in MQA - dct:title
     Current weight = 30
  * dct:description
     Result: WARN. Not included in MQA - dct:description
     Result: WARN. Not included in MQA - dct:description
     Result: WARN. Not included in MQA - dct:description
     Current weight = 30
  * dcat:keyword
     Result: OK. The property is set. Weight assigned 30
     Current weight = 60
  * dct:license
     Result: OK. The property is set. Weight assigned 20
     Result:  OK.  The  property  provides  the  correct  license  information.  Weight
assigned 10
     Current weight = 90
  * dct:issued
     Result: OK. The property is set. Weight assigned 5
     Current weight = 95
  * dcat:distribution
     Result: WARN. Not included in MQA - dcat:distribution
     Result: WARN. Not included in MQA - dcat:distribution
     Current weight = 95
  * dct:format
     Result: OK. The property is set. Weight assigned 20
     Result: OK. The property is machine-readable. Weight assigned 20
     Result: OK. The property is non-propietary. Weight assigned 20
     Current weight = 155
  * rdf:type
     Result: WARN. Not included in MQA - dct:PeriodOfTime
     Result: WARN. Not included in MQA - foaf:Organization
     Result: WARN. Not included in MQA - dcat:Dataset
     Result: WARN. Not included in MQA - vcard:Organization
     Result: WARN. Not included in MQA - dct:PeriodOfTime
     Result: WARN. Not included in MQA - dcat:Distribution
     Result: WARN. Not included in MQA - dcat:Distribution
     Current weight = 155
  * dcat:downloadURL
     Result: OK. The property is set. Weight assigned 20
     Result: OK. Weight assigned 30
     Current weight = 205
  * dct:publisher
     Result: OK. The property is set. Weight assigned 10
     Current weight = 215
  * schema1:startDate
     Result: WARN. Not included in MQA - schema1:startDate
     Current weight = 215
  * dcat:accessURL
     Result: OK. Weight assigned 50
     Current weight = 265
  * dct:spatial
     Result: OK. The property is set. Weight assigned 20
     Current weight = 285
  * dct:temporal
     Result: OK. The property is set. Weight assigned 20
     Current weight = 305
  * foaf:name
     Result: WARN. Not included in MQA - foaf:name
     Current weight = 305
  * dct:rights
     Result: OK. The property is set. Weight assigned 5
     Current weight = 310
  * dct:modified
     Result: OK. The property is set. Weight assigned 5
     Current weight = 315
  * dct:identifier
     Result: WARN. Not included in MQA - dct:identifier
     Current weight = 315
  * dcat:mediaType
     Result: OK. The property is set. Weight assigned 10
     Current weight = 325
  * dct:accessRights
     Result: OK. The property is set. Weight assigned 10
     Result: OK. The property uses a controlled vocabulary. Weight assigned 5
     Current weight = 340
  * vcard:fn
     Result: WARN. Not included in MQA - vcard:fn
     Current weight = 340
  * dcat:theme
     Result: OK. The property is set. Weight assigned 30
     Current weight = 370
  * dcat:startDate
     Result: WARN. Not included in MQA - dcat:startDate
     Current weight = 370
  * dct:language
     Result: WARN. Not included in MQA - dct:language
     Current weight = 370
  * owl:versionInfo
     Result: WARN. Not included in MQA - owl:versionInfo
     Current weight = 370
  * dcat:contactPoint
     Result: OK. The property is set. Weight assigned 20
     Current weight = 390
  * dct:format & dcat:mediaType
     Result: OK. The properties belong to a controlled vocabulary. Weight assigned
10
     Current weight= 400


  Overall MQA scoring: 400
```

Textbox 9. MQA scoring result

```
curl --location 'http://localhost:8080/pipe' \
--header 'Cache-Control: no-cache' \
--header 'Content-Type: application/json' \
--data '{
    "header": {
      "id": "497f6eca-6276-4993-bfeb-53cbbbba6f08",
      "name": "SALTED-testing",
      "title": "SALTED testing",
      "context": "SALTED tests",
      "transport": "payload",
      "version": "1.0"
    },
    "body": {
      "segments": [{
        "header": {
          "id": "497f6eca-6276-4993-bfeb-53cbbbba6f08",
          "name": "SALTED-testing_importing-oaipmh",
          "segmentNumber": 1,
          "title": "SALTED testing --importing oaipmh--",
          "processed": false
        },
        "body": {
          "config": {
            "address": "https://ckan.salted-project.eu/oai",
            "metadata": "dcat",
            "catalogue": "salted_project",
            "outputFormat": "application/rdf+xml"
          }
```

```
                }
            }]
        }
}'
```

Textbox 10. cURL request to importing-oaipmh module deployed in localhost

```
12:43:47.486 [vert.x-worker-thread-6] INFO  p.497f6eca-6276-4993-bfeb-53cbbbba6f08
- d40661d5-34ed-4bf9-b3ca-553b40f7dac2 [SALTED-testing (SALTED tests)] [SALTED-
testing_importing-oaipmh] Import started.

12:43:49.823 [vert.x-worker-thread-11] INFO  p.497f6eca-6276-4993-bfeb-53cbbbba6f08
- d40661d5-34ed-4bf9-b3ca-553b40f7dac2 [SALTED-testing (SALTED tests)] [SALTED-
testing_importing-oaipmh] Data imported:
{"total":9,"counter":1,"identifier":"2514d4af-9e21-4588-aaff-
f189995edc3b","catalogue":"salted_project"}

12:43:49.837 [vert.x-worker-thread-11] INFO  p.497f6eca-6276-4993-bfeb-53cbbbba6f08
- d40661d5-34ed-4bf9-b3ca-553b40f7dac2 [SALTED-testing (SALTED tests)] [SALTED-
testing_importing-oaipmh] Data imported:
{"total":9,"counter":2,"identifier":"f836d31d-045d-4182-bc31-
9242278ef8b2","catalogue":"salted_project"}

12:43:49.851 [vert.x-worker-thread-11] INFO  p.497f6eca-6276-4993-bfeb-53cbbbba6f08
- d40661d5-34ed-4bf9-b3ca-553b40f7dac2 [SALTED-testing (SALTED tests)] [SALTED-
testing_importing-oaipmh] Data imported:
{"total":9,"counter":3,"identifier":"1b1ff072-7583-4fe8-8cf1-
64cdae3a050b","catalogue":"salted_project"}

12:43:49.862 [vert.x-worker-thread-11] INFO  p.497f6eca-6276-4993-bfeb-53cbbbba6f08
- d40661d5-34ed-4bf9-b3ca-553b40f7dac2 [SALTED-testing (SALTED tests)] [SALTED-
testing_importing-oaipmh] Data imported:
{"total":9,"counter":4,"identifier":"7b1d4785-f23e-47f2-a49e-
14685be7fbd1","catalogue":"salted_project"}

12:43:49.870 [vert.x-worker-thread-11] INFO  p.497f6eca-6276-4993-bfeb-53cbbbba6f08
- d40661d5-34ed-4bf9-b3ca-553b40f7dac2 [SALTED-testing (SALTED tests)] [SALTED-
testing_importing-oaipmh] Data imported:
{"total":9,"counter":5,"identifier":"2d025904-0147-41aa-bad3-
823388dcdf28","catalogue":"salted_project"}

12:43:49.882 [vert.x-worker-thread-11] INFO  p.497f6eca-6276-4993-bfeb-53cbbbba6f08
- d40661d5-34ed-4bf9-b3ca-553b40f7dac2 [SALTED-testing (SALTED tests)] [SALTED-
testing_importing-oaipmh] Data imported:
{"total":9,"counter":6,"identifier":"2407c1df-7b85-4345-810e-
ce50cc9800df","catalogue":"salted_project"}

12:43:49.889 [vert.x-worker-thread-11] INFO  p.497f6eca-6276-4993-bfeb-53cbbbba6f08
- d40661d5-34ed-4bf9-b3ca-553b40f7dac2 [SALTED-testing (SALTED tests)] [SALTED-
testing_importing-oaipmh] Data imported:
{"total":9,"counter":7,"identifier":"babcc4a4-062d-48d1-a145-
e2d9ba7d6596","catalogue":"salted_project"}

12:43:49.917 [vert.x-worker-thread-11] INFO  p.497f6eca-6276-4993-bfeb-53cbbbba6f08
- d40661d5-34ed-4bf9-b3ca-553b40f7dac2 [SALTED-testing (SALTED tests)] [SALTED-
testing_importing-oaipmh] Data imported:
{"total":9,"counter":8,"identifier":"ff4f3194-f413-4e34-b924-
0cb01ff15df1","catalogue":"salted_project"}

12:43:49.924 [vert.x-worker-thread-11] INFO  p.497f6eca-6276-4993-bfeb-53cbbbba6f08
- d40661d5-34ed-4bf9-b3ca-553b40f7dac2 [SALTED-testing (SALTED tests)] [SALTED-
testing_importing-oaipmh] Data imported:
{"total":9,"counter":9,"identifier":"8e7aff2d-39b1-412d-96c6-
18960a7ff41b","catalogue":"salted_project"}

12:43:49.925 [vert.x-worker-thread-11] INFO  p.497f6eca-6276-4993-bfeb-53cbbbba6f08
- d40661d5-34ed-4bf9-b3ca-553b40f7dac2 [SALTED-testing (SALTED tests)] [SALTED-
testing_importing-oaipmh] Import metadata finished
```

Textbox 11. Output log of importing-oaipmh module