



**TEADAL**



# **Increasing data availability awareness via metadata federation**

**Alessio Carenini**

*Cefriel*

**TEADAL Final Event**

*02/10/2025*

***WWW.TEADAL.EU***

# TEADAL Context: Catalogue and Governance

How to prevent a datalake to become a **data swamp**?

“A data swamp is a data repository that has become unusable due to a lack of proper data governance, leading to disorganized, undocumented, and poor-quality data that is difficult to access and leverage for business insights.”

A growing need for **control**

- **Governance:** authorization processes, organizational processes, compliance processes
- **Automation:** complex procedures enabled by structured metadata

A metadata catalogue can become the tool to exert such control



# Opening to a wider scope

The world doesn't end at the borders of your datalake

Many data sharing initiatives are already in place:

- Open data
- Domain-specific and purpose-specific custom platforms
- Dataspaces

Main questions involved:

- How do I selectively advertise my Data Products?
- How can I exert control over external digital assets and bring them under my governance rules?

# Implementing TEADAL Federated Catalogue via KCONG



KCONG (Knowledge Catalogue and Governance) is a shared catalogue containing structured descriptions (metadata) of «assets»

Digital assets (datasets, API, software artifacts, ML models, ...)

Physical assets (digital twin, desk services, ...)

KCONG allows to

- ▶ SHARE the knowledge about your digital assets within your organization and with your partners
- ▶ CONTROL/REGULATE asset publication and access through customizable governance rules
- ▶ AUTOMATE the asset lifecycle, including the creation of software artifacts and their deployment

Emphasis on flexibility

- ▶ Multiple asset types
- ▶ Complex governance processes
- ▶ Automation

# TEADAL Federated Catalogue



TEADAL HOME CATALOGUE WORKSPACE ADMIN

## Teadal Metadata Catalogue

The Teadal Metadata Catalogue is a digital platform describing uniformly all the relevant data sources collected for the Teadal case studies, and supporting their governance mechanisms.

Asset type  Insert asset name

Latest VIEW ALL

uality

< a

**ERT - Production Dataset**  
Dataset

Production Data

**ERT - Forecast Dataset**  
Dataset

Forecast Data

**ERT Portugal Plan Data**  
Federated Data Product

ERT Data Related To Production, Sales, Quality

TEADAL HOME CATALOGUE WORKSPACE ADMIN

## CATALOGUE

Search

**FILTERS**

Asset types

Providers

**Ribera's Patients Population Study**  
Clinical Study Proposal

Population Area Study June 2025

**Obesity Clinical Study**  
Clinical Study Proposal

Clinical Study To Measure [TBD]

**Patients Data FDP**  
Federated Data Product

All The Patients Personal Information From Ribera Salud

**Czech Plant Data Access FDP**  
Federated Data Product

Get Data From ERT Plant In Czech Republic

**AMTS Integrated Mobility FDP**  
Federated Data Product

AMTS Timetables Integrated With OpenStreetMap Data

**Aquaview FDP**  
Federated Data Product

Soil Moisture Data Access From Terraviv

**BOX2M-FDP**  
Federated Data Product

Pull Real-Time Sensors Data And Makes It Available Through FDP To The Teadal Consumers

**Ribera Salud Fdp-Medicine**  
Federated Data Product

Ribera Salud Fdp-Medicine

**ERT Portugal Plan Data**  
Federated Data Product

ERT Data Related To Production, Sales, Quality

**Aquaview FDP Agreement Between Admin And Admin**  
Agreement

Agreement Between Admin And Admin About The Usage Of Aquaview FDP Federated Data Product

**Patients Data FDP Agreement Between Admin And Admin**  
Agreement

Agreement Between Admin And Admin About The Usage Of Patients Data FDP Federated Data Product

**BOX2M Energy Consumption And Building Occupancy Data**  
Dataset

Real Time Data From Sensors About Energy Consumption, Building Occupancy

**Ribera Salud Drug Exposure Data**  
Dataset

Drug Exposure

**ERT Portugal Plant Data**  
Dataset

ERT Data Related To Production, Sales, Quality

**Tuscany Buildings Energy Performance Certificates**  
Dataset

APE Certificates

**Ribera Salud Patients Conditions Data**  
Dataset

Patients Conditions

# TEADAL Federated Catalogue



TEADAL HOME CATALOGUE **WORKSPACE** ADMIN

## Obesity Clinical Study

Clinical Study Proposal

Status: approved

[EDIT](#) [METADATA](#)

DESCRIPTION VERSIONS

Name  
Obesity Clinical Study

Description  
Clinical Study to measure [TBD]

Version  
1.0

Company / Institution name  
San Raffaele hospital (Milano)

Contacts

Name and Surname  
John Doe

email  
JohnDoe@SRMH.com

Hospitals where to execute the study  
Hospital A, Hospital B

Query to evaluate the number of patient affected by the study  
To be removed in final version

Script to be executed

2025 - Cefriel - Privacy - OpenAPI

TEADAL HOME CATALOGUE **WORKSPACE** ADMIN

[+ CREATE A NEW ASSET](#)

OVERVIEW MY ASSETS FAVOURITES REQUESTS

## WORKSPACE

### Number Of Assets

- Clinical study proposal: 5
- Shared federated data...: 3
- Agreement: 3
- Federated data product: 7
- Dataset: 30

### Status Summary

Status	Count	Percentage
Approved	29	97.6%
Submitted	1	2.4%

### Detailed Stats

**Clinical Study Proposal**  
A proposal for a new clinical study over multiple hospitals  
You own 2 items.

Approved: 2



# ***How do I selectively advertise my Data Products?***



# Cross-publishing via governance processes



“Actions” can be attached to asset types

The screenshot shows the TEADAL workspace interface. At the top, there are navigation tabs: HOME, CATALOGUE, and WORKSPACE (which is active). A user profile icon and the name 'ADMIN' are visible in the top right. Below the navigation, the page title is 'Dataset'. The status is 'published'. There are three buttons: 'EDIT', 'METADATA', and 'ACTIONS'. Below this, there are two tabs: 'DESCRIPTION' and 'VERSIONS'. Under 'DESCRIPTION', the following fields are visible: Name: 'Sicily bus timetables', Description: 'Sicily bus timetables in GTFS format', Version: '1', and Company / institution name: 'Regione Sicilia'. There is also a 'Contacts' field at the bottom.

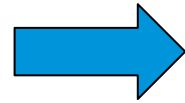
A ticketing system keeps track of pending requests and past decisions

The screenshot shows the TEADAL workspace interface with the 'REQUESTS' tab selected. At the top, there are navigation tabs: HOME, CATALOGUE, and WORKSPACE (which is active). A user profile icon and the name 'ADMIN' are visible in the top right. Below the navigation, the page title is 'WORKSPACE'. There is a yellow button labeled '+ CREATE A NEW ASSET'. Below this, there are four tabs: 'OVERVIEW', 'MY ASSETS', 'FAVOURITES', and 'REQUESTS' (which is active). Under 'REQUESTS', there is a section for 'Incoming requests' with a table that is currently empty, showing 'No data available'. Below this, there is a section for 'Outgoing requests' with a table containing two rows of data. The table has columns for Asset name, Asset type, Request, Date, Last update, and Status.

Asset name	Asset type	Request	Date	Last update	Status
Aquaview FDP	Federated data product	Negotiate access	Sep 30, 2025, 04:32 PM	Sep 30, 2025, 04:33 PM	Closed
Patients data FDP	Federated data product	Negotiate access	Sep 30, 2025, 03:45 PM	Sep 30, 2025, 03:45 PM	Submitted

# Hiding governance complexities

What the user sees

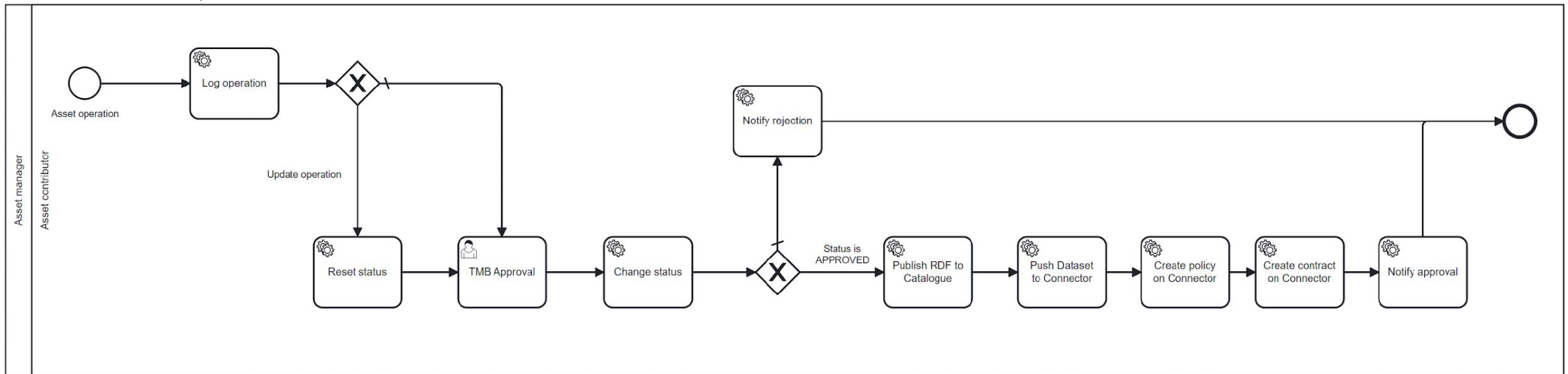


**C**  
Dataset  
Status: published

ADVERTISE ON BARCELONA DATASPACE

DESCRIPTION    VERSIONS

What is executed

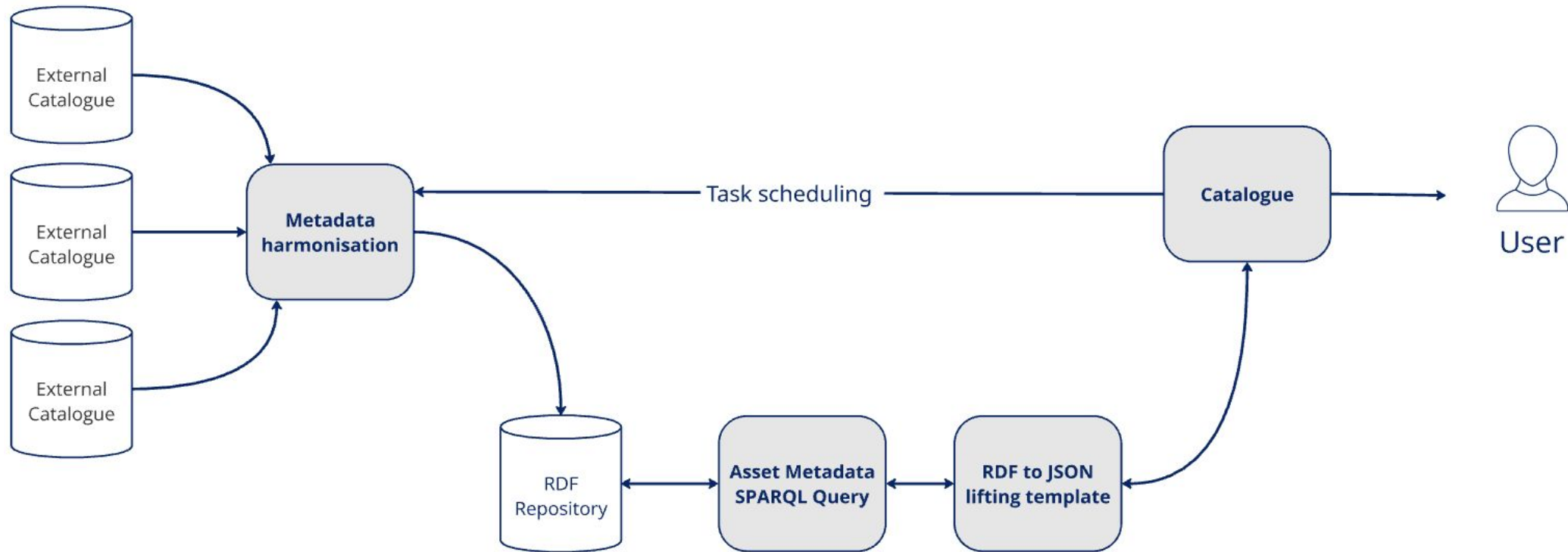




***How can I exert control over external digital assets and bring them under my governance rules?***



# Metadata gathering and harmonisation



# Integration of new metadata sources

Three main techniques to add a new source:

- API polling from Catalogue: API access to a shared catalogue in case it exists in the implementation site
- Connector polling from Catalogue: API access to a connector with a valid identity for the local dataspace
- External metadata push to Catalogue

For metadata harmonisation we need to know the metadata schema of the new source to create the mapping onto the metadata schema used in the Catalogue:

- If the metadata schema is different, we have to find mappings
- Well-known schemas (like CKAN) allow for nearly 100% reusability (as in the case of Spain and Italy Open Data Portals)

## Exploration API: Constraining access to queries over the KG

SPARQL endpoints have a poor/non existing way to be protected from unintended usage

Only authentication is considered

No authorisation (at least in open source products)

What if we want to constrain what the user can do?

Define a SPARQL query

Define a protected API endpoint with parameters

Run the query at every API call

Map the SPARQL results onto the desired results of the API

# Feeding the UI: Exploration API in practice

API call (with parameter) is received

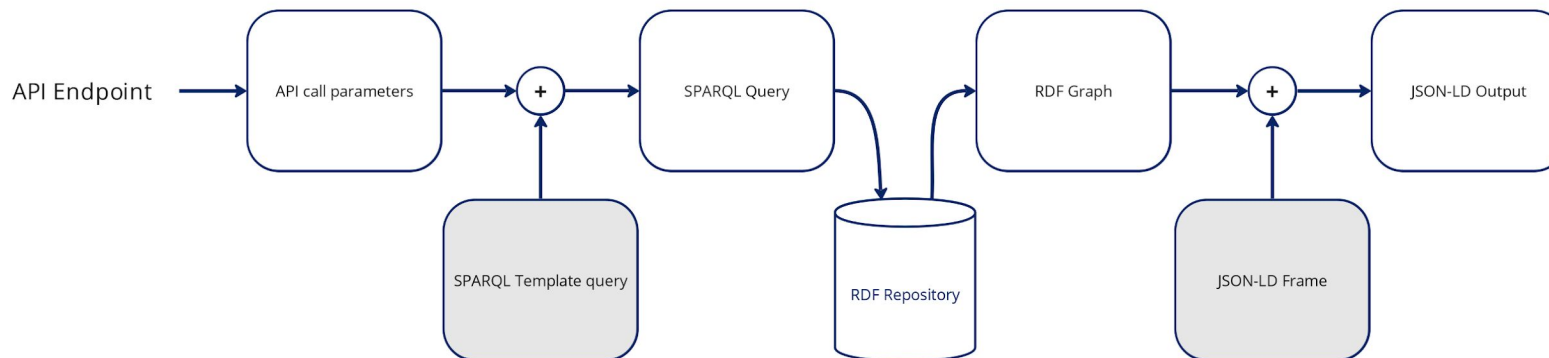
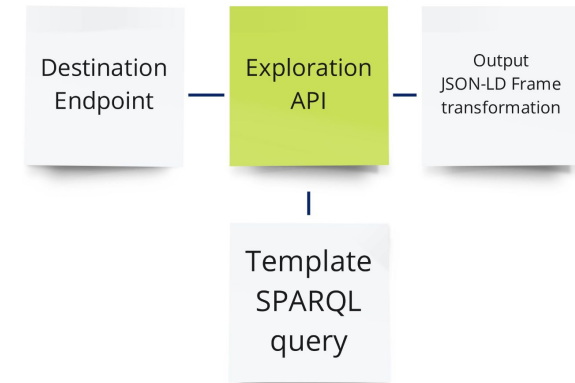
Exploration API object is retrieved with all the details

Parameters + SPARQL template = SPARQL query

SPARQL query is executed on the desired RDF graph

If it is a CONSTRUCT

- Apply JSON-LD Frame to the resulting RDF graph



# Mobility metadata aggregation



Four data sharing initiatives made available to the deployEMDS project via TEADAL technologies:

- Mobility Open Data Portal from Spain: CKAN-based, custom metadata model
- Mobility Open Data Portal from Italy: CKAN-based, custom metadata model
- Belgium National Access Point: CKAN-based, metadata in mobilityDCAT-AP format
- Barcelona mobility dataspace: access via dataspace connector, DCAT+ODRL metadata

The screenshot displays the EMDS Catalogue interface. On the left is a 'FILTERS' sidebar with sections for 'Owners' (Spain NAP, Belgium NAP, DCAT dataspace catalogue), 'Origin' (Dati aperti della pubblica amministrazione, Spain NAP, Belgium NAP, DCAT dataspace catalogue), and 'Destination' (Dati aperti della pubblica amministrazione, Spain NAP, Belgium NAP, DCAT dataspace catalogue). The main area shows a grid of data cards, each with a title, a 'Spain NAP' or 'Dati aperti della pubblica amministrazione' button, and a description. The cards include:

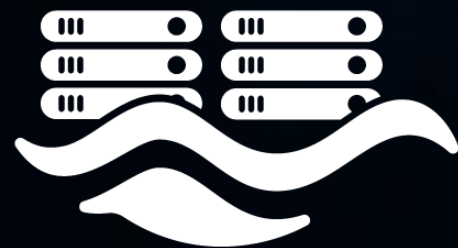
- Mapa de accesibilidad de la isla de Tenerife
- Matriz Origen-Destino de transporte público de la isla de Tenerife de abril de 2025
- Matriz Origen-Destino de transporte público de la isla de Tenerife de febrero de 2025
- Matriz Origen-Destino de transporte público de la isla de Tenerife de marzo de 2025
- Bus Services Available
- Bus Stops Available
- Bus Vehicle Information
- test1
- test2
- Accesi pedonali stazioni ferroviarie
- Accuratezza spire anno 2019
- Accuratezza spire anno 2020
- Accuratezza spire anno 2021
- Accuratezza spire anno 2022
- Accuratezza spire anno 2023
- Accuratezza spire anno 2024
- Accuratezza spire anno 2025
- ACI Milano - Delegazioni
- Alimentazione e classificazione ambientale (Euro) autoveicoli Bologna
- Ambiti della sosta
- AMTAB - servizi di trasporto, sosta e mobilità
- Tematismi numerici tratti dalla CTR 10.000 - Edizione 1991 - Archi ferroviari
- Archi stradali
- Tematismi numerici tratti dalla CTR 10.000 - Edizione 1991 - Archi stradali

# Conclusions



TEADAL technologies provided to be ready to connect datalakes to existing data sharing ecosystem

- Dataspace-ready
- Better governance for external data sources and data products
- Increased awareness and reuse of existing data products



TEADAL



THANKS



[TEADAL.EU](http://TEADAL.EU)



[@TEADAL\\_eu](https://twitter.com/TEADAL_eu)



[@TEADAL](https://www.linkedin.com/company/teadal)



TEADAL project is funded by the EU's Horizon Europe programme under Grant Agreement number 101070186