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, …)

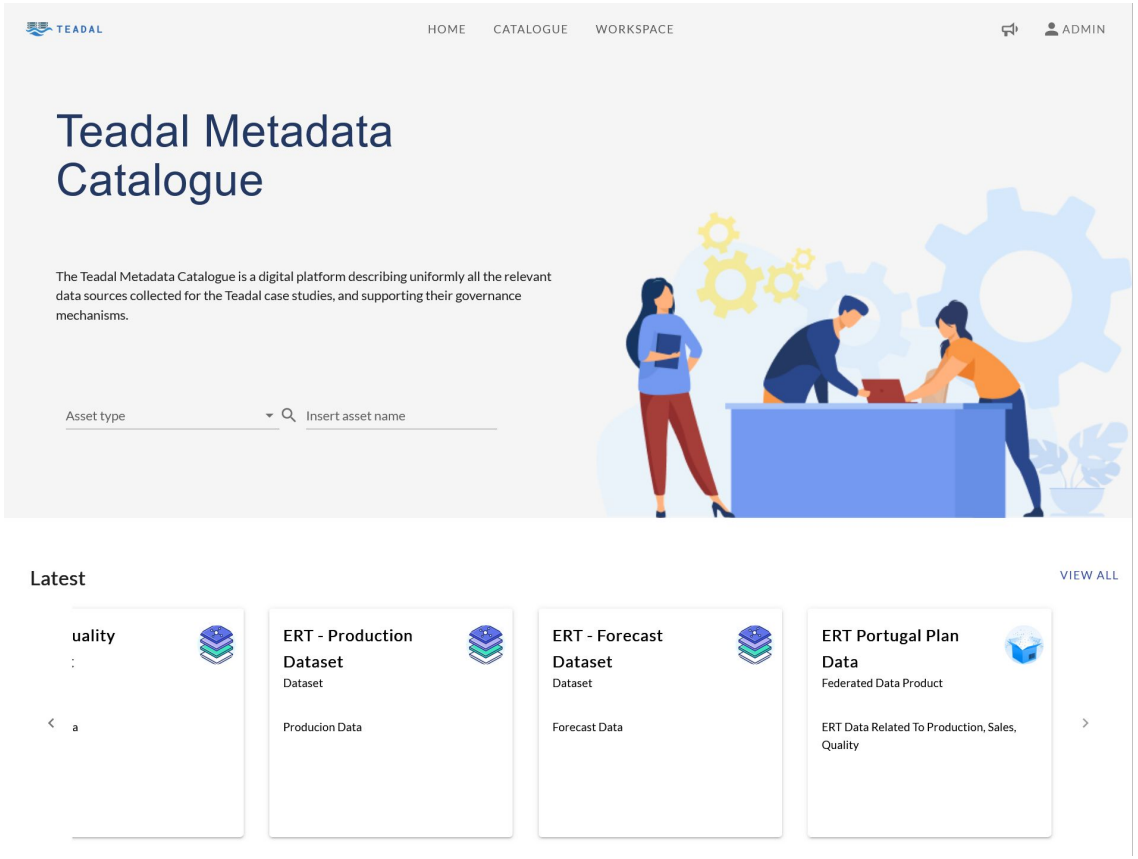
KCONG allows to

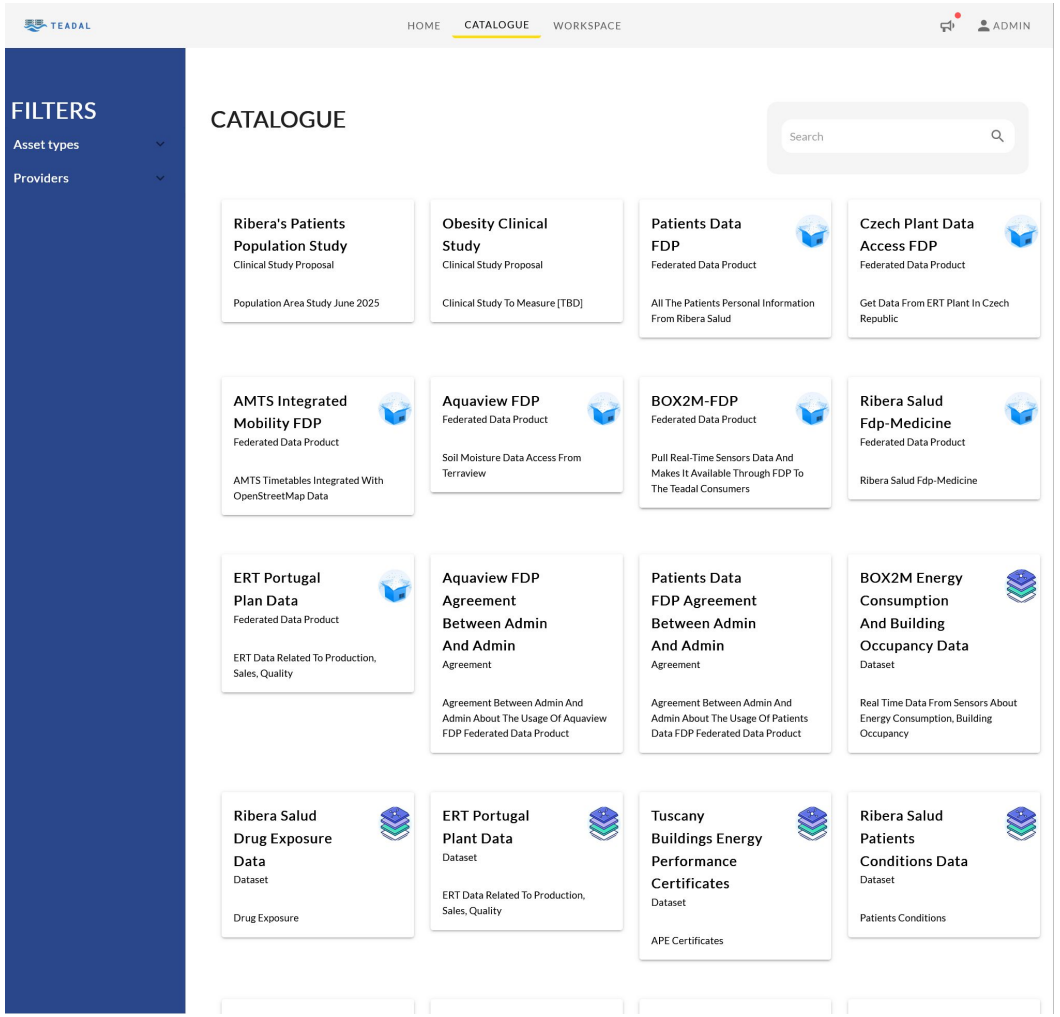
- SHARE the knowledge about your digital assets within your organization and with your partners
- CONTROL/REGULATE asset publication and access through customizable governance rules
- AUTOMATE the asset lifecycle, including the creation of software artifacts and their deployment

Emphasis on flexibility

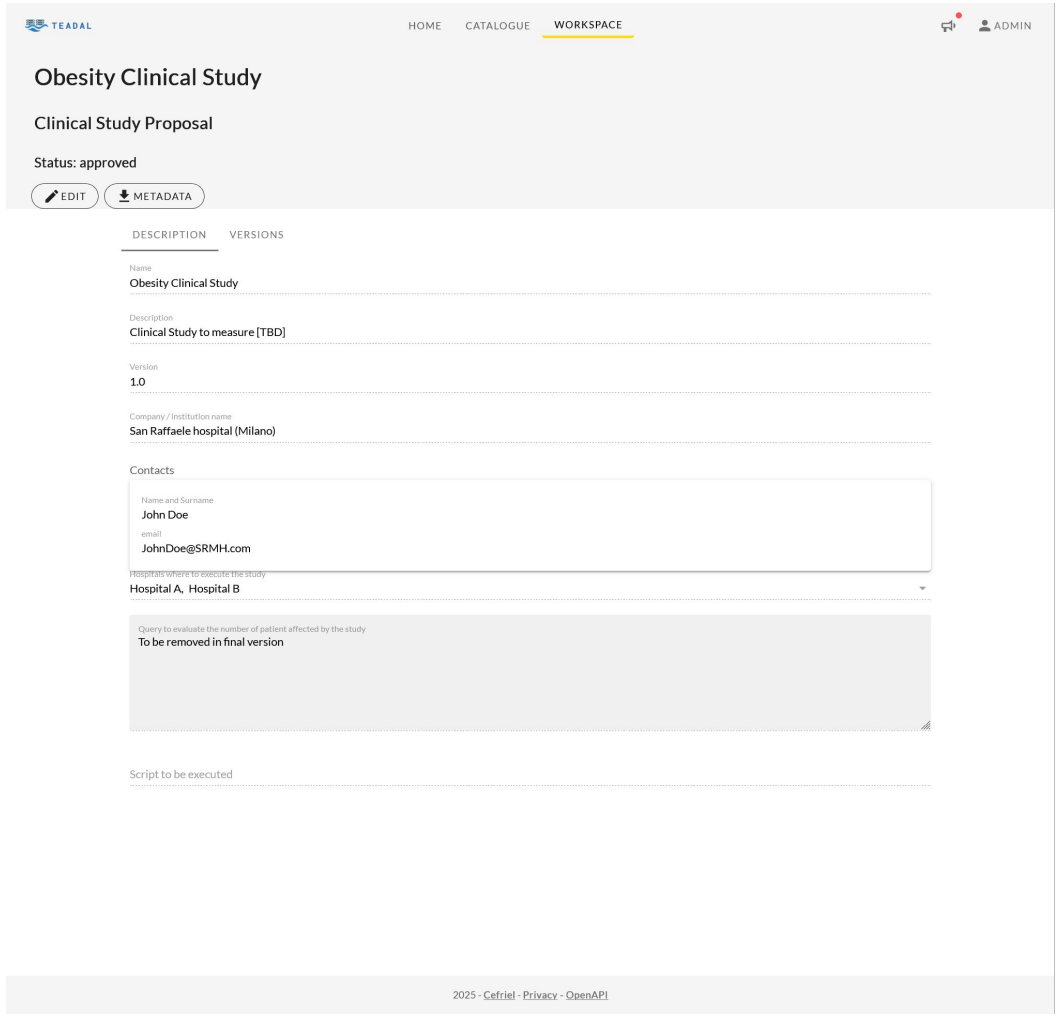
- Multiple asset types
- Complex governance processes
- Automation

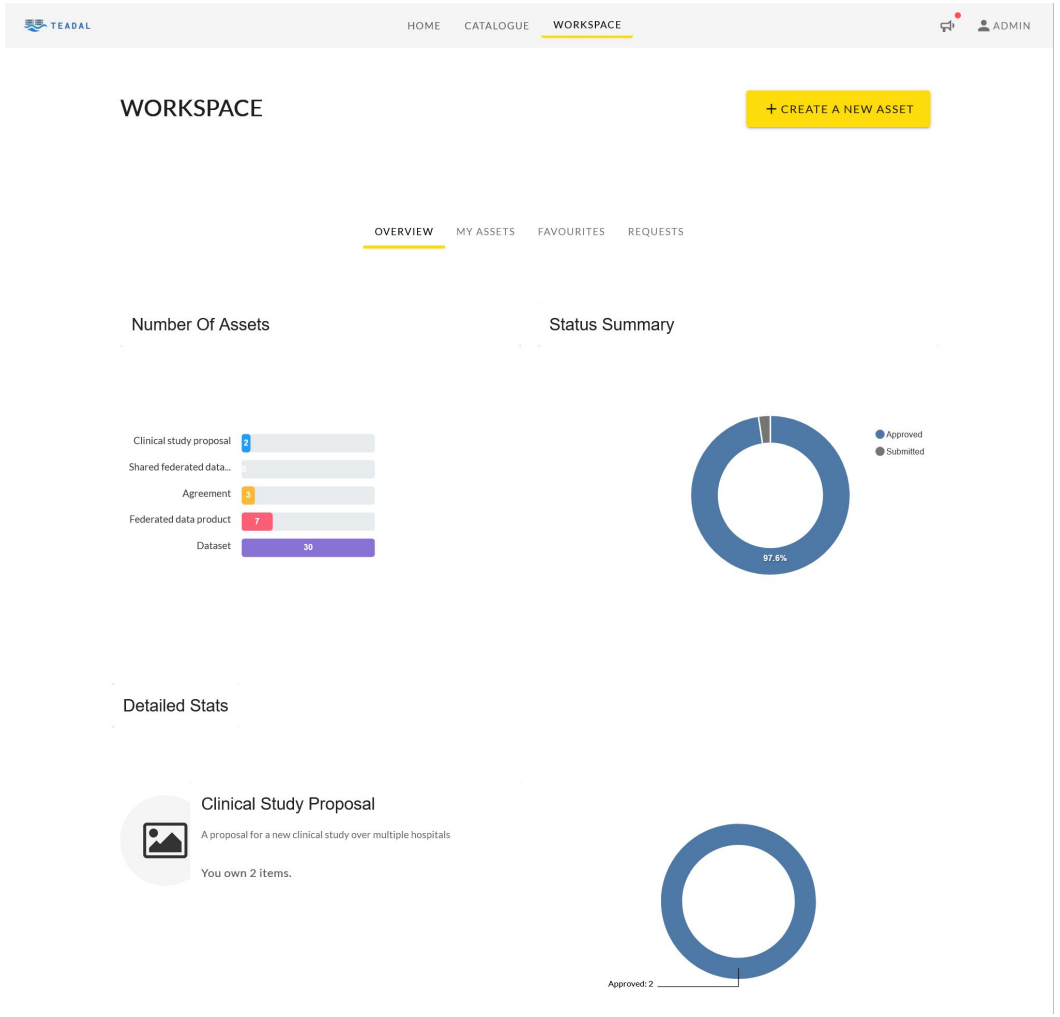
TEADAL Federated Catalogue

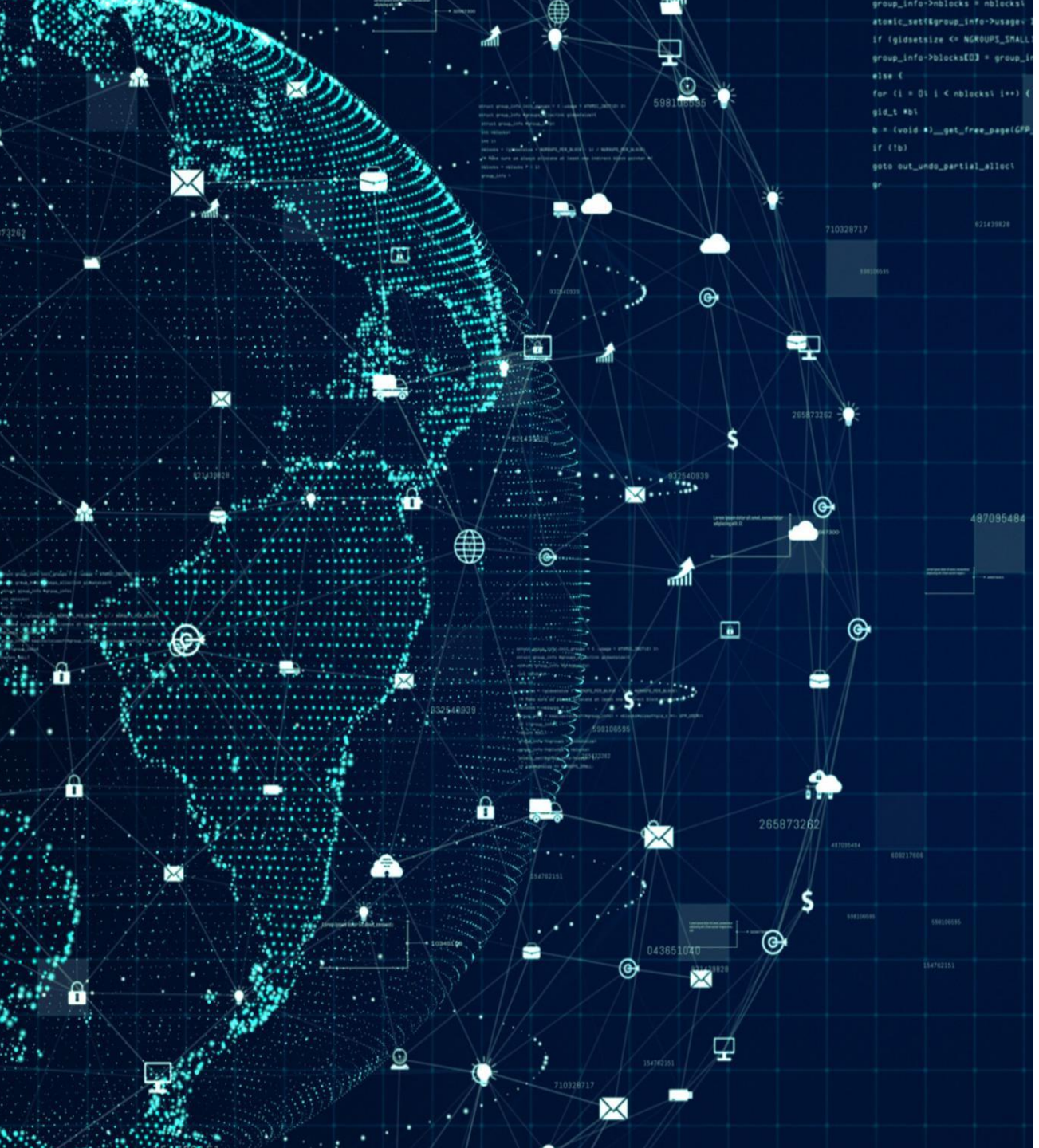




TEADAL Federated Catalogue



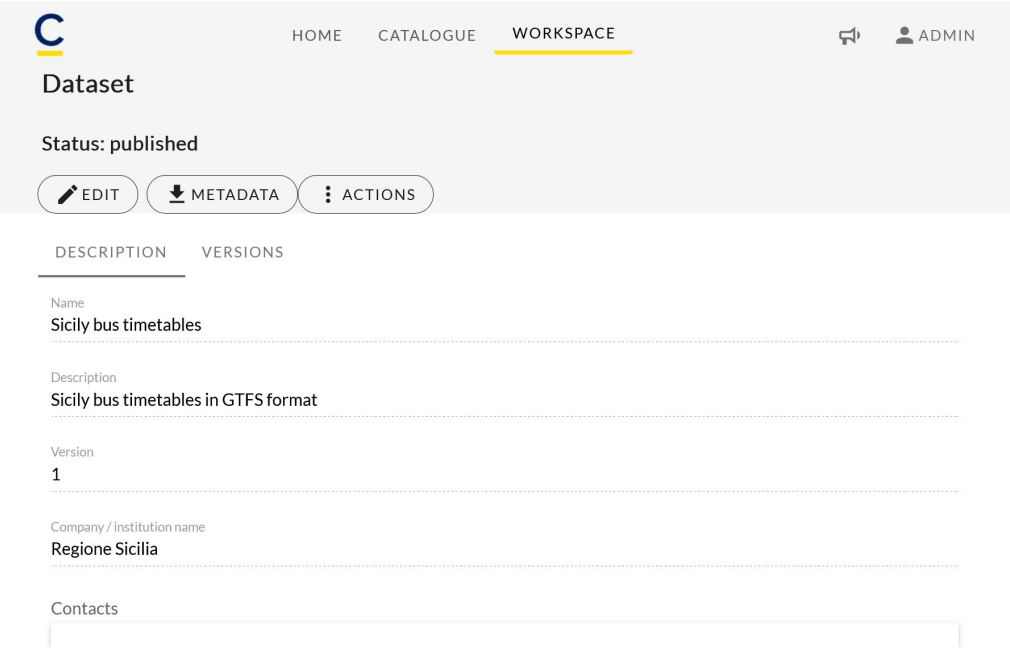




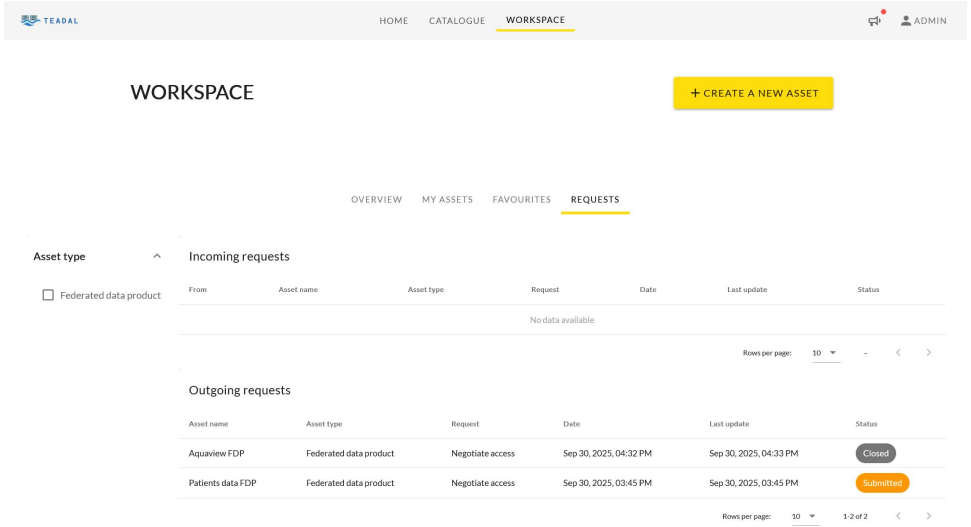
How do I selectively advertise my Data Products?

Cross-publishing via governance processes

"Actions" can be attached to asset types

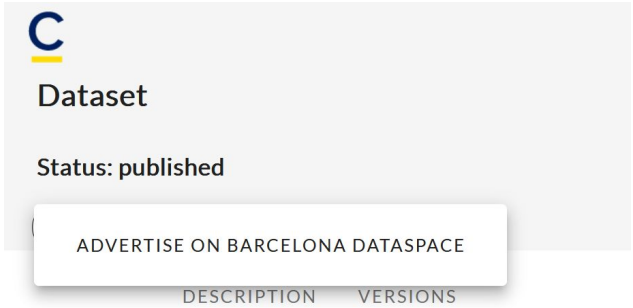


A ticketing system keeps track of pending requests and past decisions

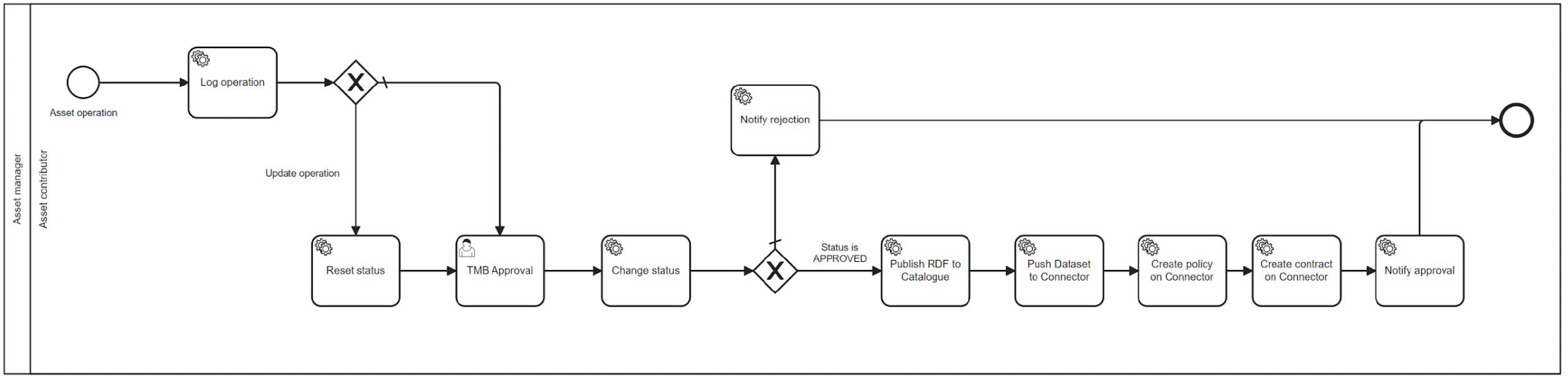


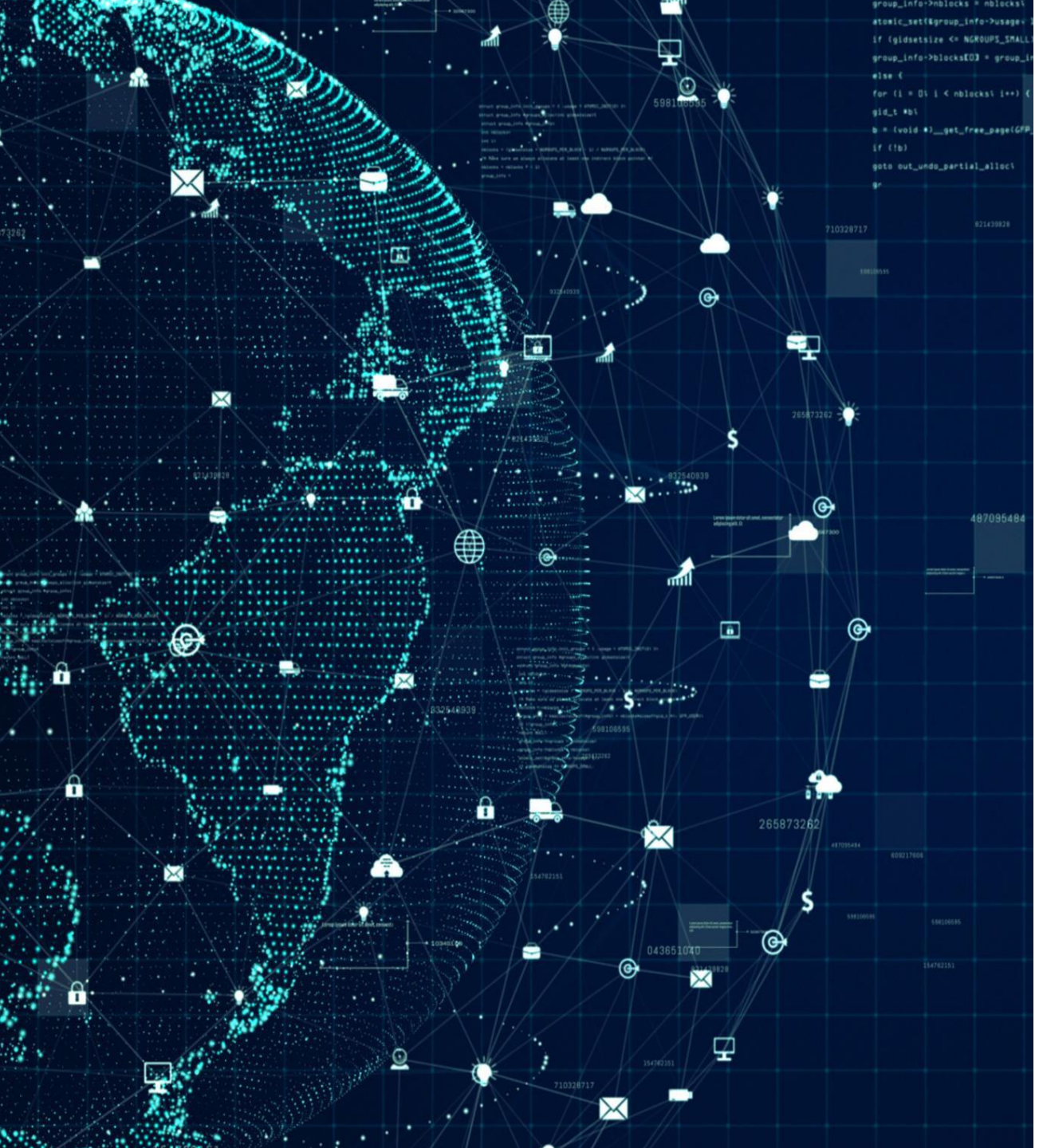
Hiding governance complexities

What the user sees



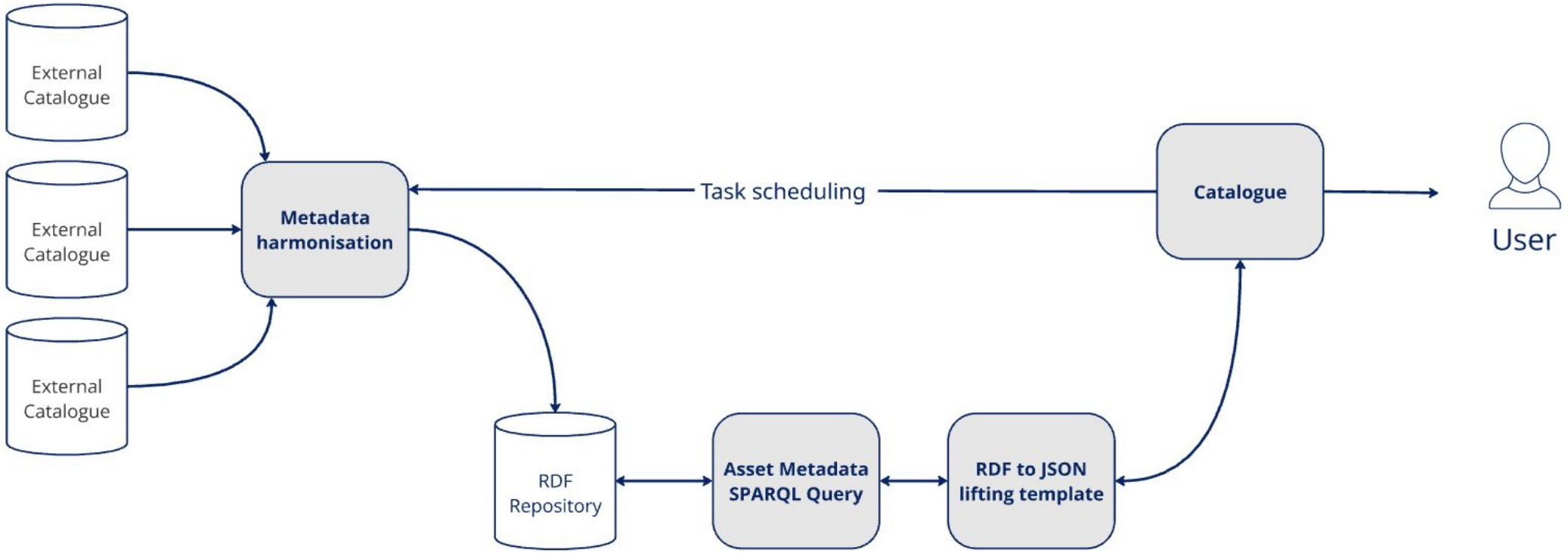
What is executed





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

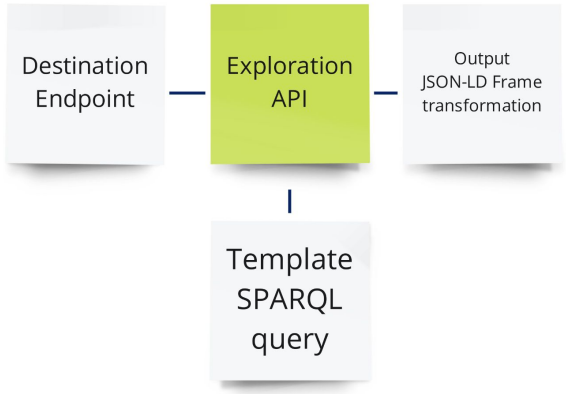
Feeding the UI: Exploration API in practice

API call (with parameter) is received

Exploration API object is retrieved with all the details

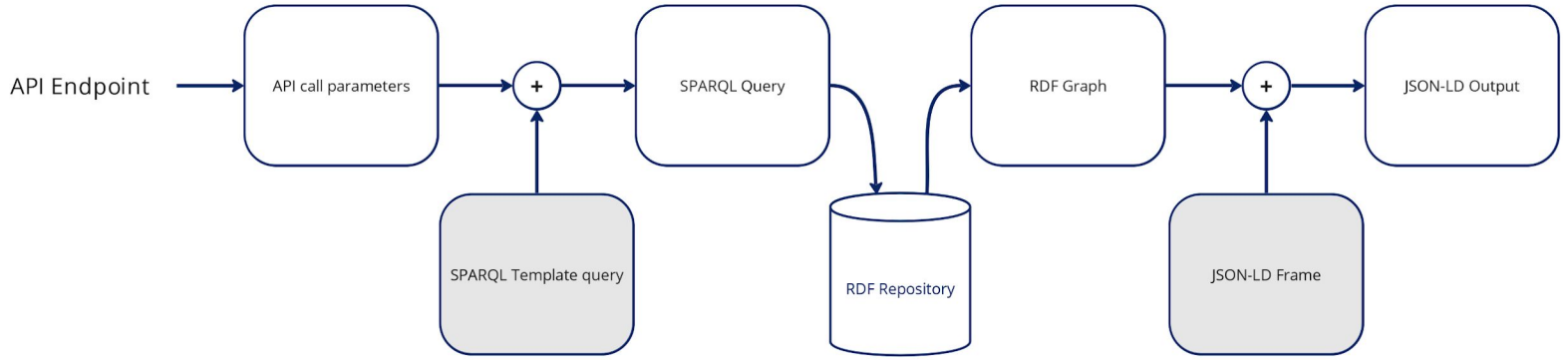
Parameters + SPARQL template = SPARQL query

SPARQL query is executed on the desidered RDF graph



If it is a CONSTRUCT

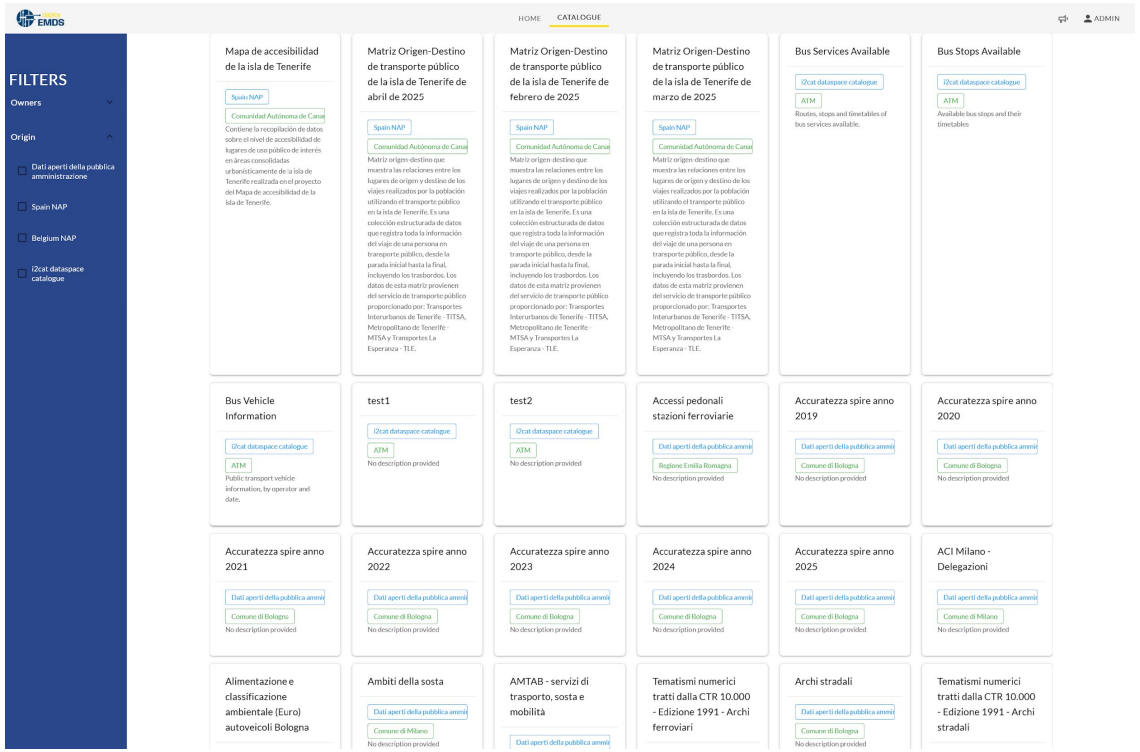
- Apply JSON-LD Frame to the resulting RDF graph



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



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

@TEADAL_eu

@TEADAL

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