

Public Procurement Data Space



Client profile

The Commission's Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs is responsible for EU policy on the single market, industry, entrepreneurship and small businesses.

Why NTT DATA?

- We build innovative, industry-leading solutions that grow enterprises' revenue and keep them ahead the competition.
- We take advantage of the growing convergence of IT and connectivity services to connect people and things.
- We manage companies' applications, data and infrastructure to decrease costs and create greater efficiencies.



“Harness the Power of Public Procurement Data”

PPDS Slogan, EC - DG GROW

Business need

The client looks to create an **integrated space for public procurement data** above and below the EU thresholds, combined with an analytics toolset using:

- Emerging technologies that will benefit public buyers.
- Policy-makers at the EU and national levels.
- And other relevant stakeholders such as companies, researchers, and taxpayers.

Solution

From NTT DATA, we supported our client to:

- Define a target architecture leveraging open-source technologies to avoid vendor lock in and capitalize on the portability of software binary images to be deployed on AWS infrastructure.
- Build an MVP to test the viability of the solution.
- Evolve it to become a pilot-ready and robust scale-up project in terms of integrations, data exploitation capacities, security, and elasticity

Outcome

The long-term objective is for that PPDS becomes the 'de facto' data access point for all procurement data in the EU, by ensuring the following capabilities:

- **Enhanced Single Market Scoreboard (SMS):** PPDS provides an enhanced version of SMS indicators allowing user access to a set of sophisticatedly enhanced indicator results through its state-of-the-art analytics and visualization tools.
- **Prescriptive Data Quality and Analytics Dashboards:** PPDS yields a rich set of data quality and advanced analytics dashboards applied to all participants datasets to proactively highlight common data issues and provide non-technical users with intuitive tools to get access to valuable insights.

TECHNICAL SPECIFICATIONS

PPDS Software Architecture constitutes of four service layers, described as follows:

Presentation Layer: A service layer that implements a web portal using Drupal as main technology to provide a single-entry point to all PPDS participants (Public and Advanced).

Business Layer: A service layer composed mainly of five capability components:

- Data Source Management: A capability component that implements a semantic pipeline to manage existent and new PPDS data sources.
- User Management: A capability component that oversees user access through a federated login mechanism leveraging EU Login.
- Linked Data Access Point: A SPARQL endpoint shared with public and advanced users to fetch PPDS data programmatically.
- Data Quality and Analytics Application: A data quality and analytics visualization application implemented using Apache Superset, which allows PPDS end-users to interact with curated linked data points and gain valuable insights.
- Data Catalogue: A public metadata catalog summarizes key figures about PPDS, shared with all PPDS participants.

Integration Layer: A service layer comprises four capability components:

- Data Collection: A set of ETL pipelines developed in Apache NiFi and aggregated as a data collector component to ensure data collection from different PPDS data providers.
- Semantic Data Management: A mash-up of microservices leveraging an in-house accelerator named, Knowledge Factory to orchestrate a semantic transformation pipeline of the collected data
- Data Quality: A semantic data quality workload written in SHACL validations to ensure collected and transformed linked data adheres to a set of business and technical rules.
- Data Analytics: A semantic data analytics component, which encapsulates the SMS indicators logic in SPARQL queries executed in an orchestrated fashion using Apache Airflow over OpenLink Virtuoso triple store.

Persistence Layer: A service layer comprises a PPDS storage layer using AWS S3 for raw collected data, and a set of “Public” and “Ancillary” knowledge graphs stored in OpenLink Virtuoso instance running AWS cloud infrastructure.

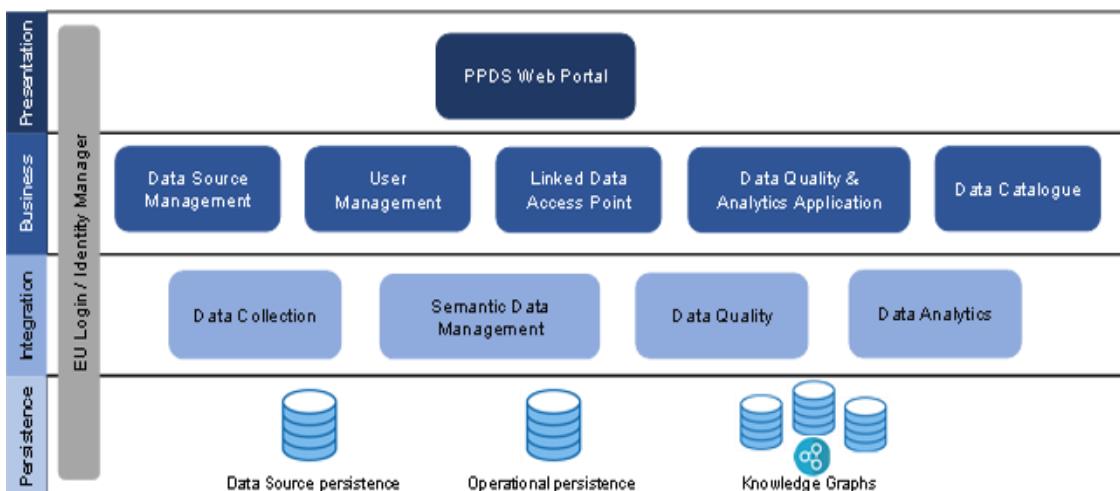


Fig. 1 – The four service layers of PPDS Software Architecture