

Fraunhofer Institute for Open Communication Systems

---

## Registries – Technical Considerations and Showcases

Prof. Dr. Sonja Schimmler

Fraunhofer FOKUS & TU Berlin



## Introduction

### Registries – Technical Considerations and Showcases



#### Main Questions:

- What is the state of the art in setting up registries (for semantic models)?
- What special requirements do we have when it comes to registries for semantic models?

## piveau: Overview

---

### piveau

- Data management ecosystem for the public sector
  - Used in a variety of different projects



## piveau: Components

---

### piveau

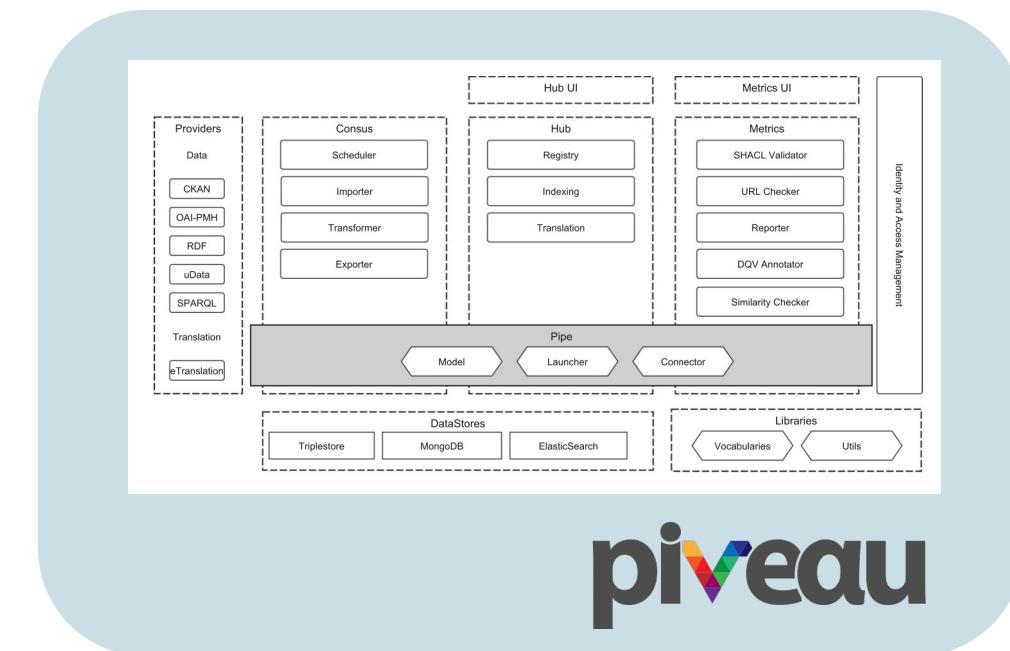
- Data management ecosystem for the public sector
  - Provides components and tools to support the entire data processing chain
  - piveau HUB, piveau CONSUS and piveau METRICS, together with piveau UI as core components



## piveau: Architecture

### piveau

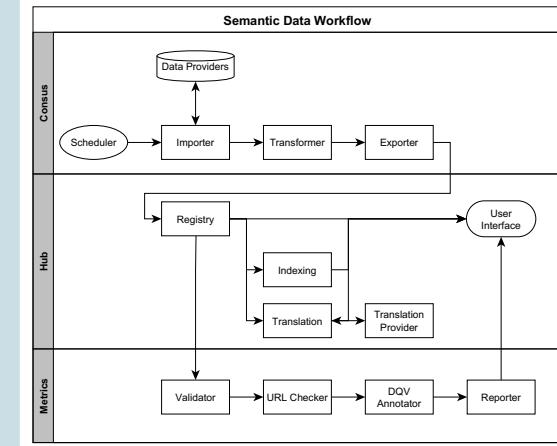
- Based on a microservice architecture and a custom pipeline system
- The services can be connected and orchestrated in a generic fashion to implement specific data processing chains
- There is no central instance, which is responsible for orchestrating the services



## piveau: Semantic Data Workflow

### piveau

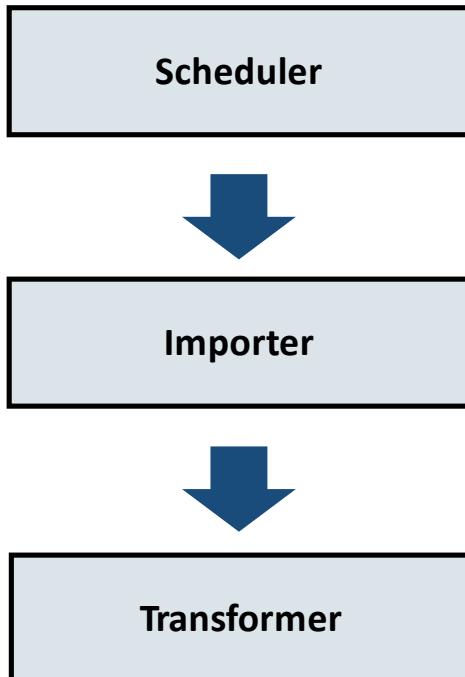
- Semantic workflow consists of
  - Data Acquisition
  - Data Processing and Data Storing
  - Data Quality Evaluation



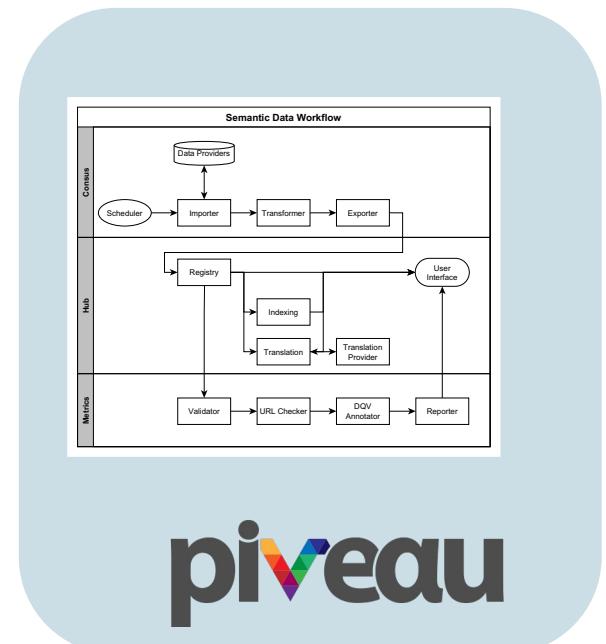
**piveau**

## piveau: Semantic Data Workflow

### Data Acquisition



- Main entry point for the service orchestration.
- Periodically triggers the harvesting process, defined as a **pipeline descriptor**.
- Retrieves the metadata from the source portal(s).
- Supports a variety of interfaces and data formats: **OAI-PMH, RDF, SPARQL, ...**
- Generates a complete list of identifiers of all datasets.
- Applies lightweight scripting-based transformation rules.
- The final output is DCAT-AP-compliant RDF.

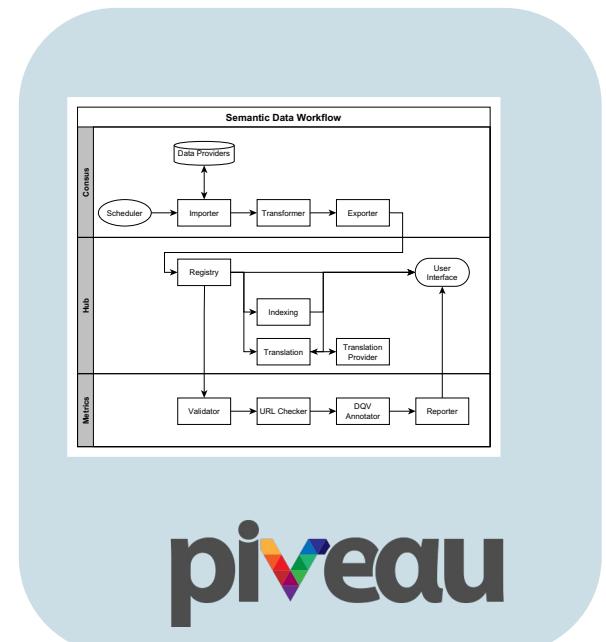


## piveau: Semantic Data Workflow

### Data Processing and Data Storing

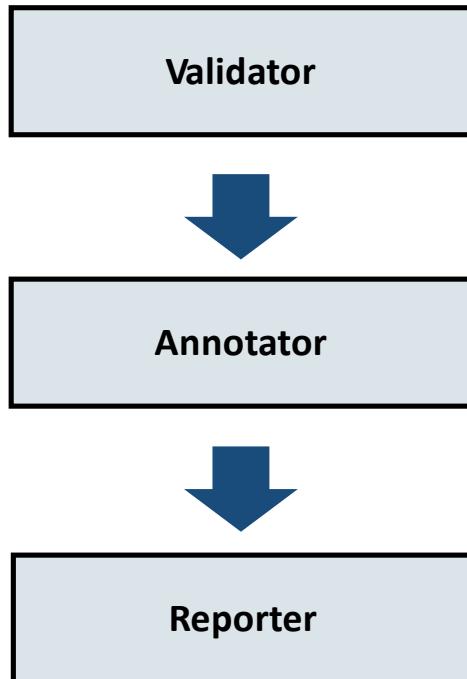


- Middleware and abstraction layer to interact with the triplestore (**Virtuoso**).
- RESTful interface for RDF (**Turtle**, **JSON-LD**, **RDF/XML**, ...).
- Responsible for managing the high-performance search index (**Elasticsearch**).
- ‘Flattening’ of the DCAT RDF to simple JSON.
- Middleware to third-party translation services.
- Returns the translation by applying the native **multi-language features of RDF**.

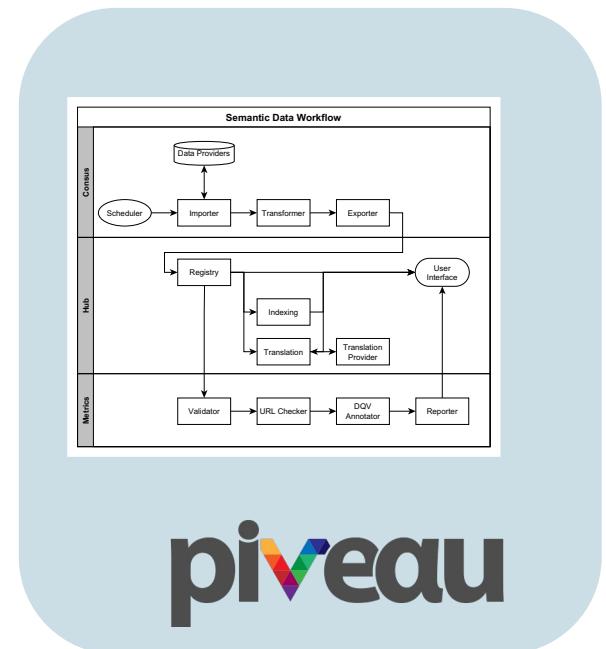


## piveau: Semantic Data Workflow

### Data Quality Evaluation



- Application of the **W3C SHACL**.
- Applied rules can also be extended or replaced (built-in DCAT-AP).
  
- Quality assessment for each dataset with a custom metrics scheme.
- Inspired by the **FAIR principles** and the **5\* principles** for Linked Open Data.
  
- Applies W3C Data Quality Vocabulary (**DQV**) for creating quality reports.
- Based on the results of the Validator and Annotator.
- Attached as RDF to the concerned dataset.



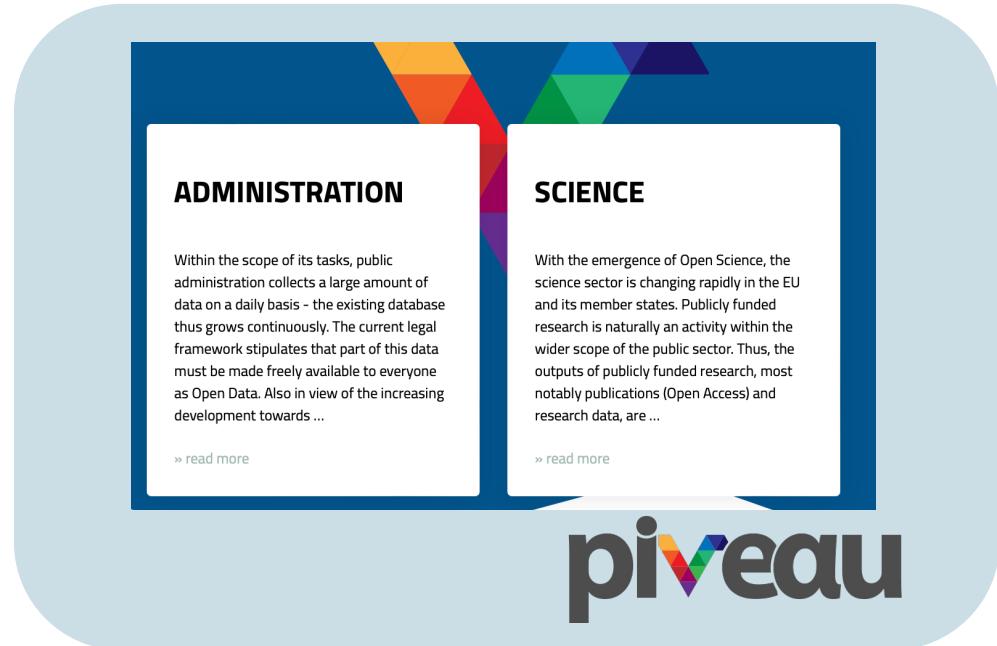
**piveau**

## piveau: Use Cases

---

### piveau

- piveau is used within different projects and domains



The image shows a light blue rounded rectangle containing two white rectangular cards. The card on the left is titled "ADMINISTRATION" and the card on the right is titled "SCIENCE". Both cards contain text and a "» read more" link at the bottom. Below the cards is a large, bold "piveau" logo.

**ADMINISTRATION**

Within the scope of its tasks, public administration collects a large amount of data on a daily basis - the existing database thus grows continuously. The current legal framework stipulates that part of this data must be made freely available to everyone as Open Data. Also in view of the increasing development towards ...

[» read more](#)

**SCIENCE**

With the emergence of Open Science, the science sector is changing rapidly in the EU and its member states. Publicly funded research is naturally an activity within the wider scope of the public sector. Thus, the outputs of publicly funded research, most notably publications (Open Access) and research data, are ...

[» read more](#)

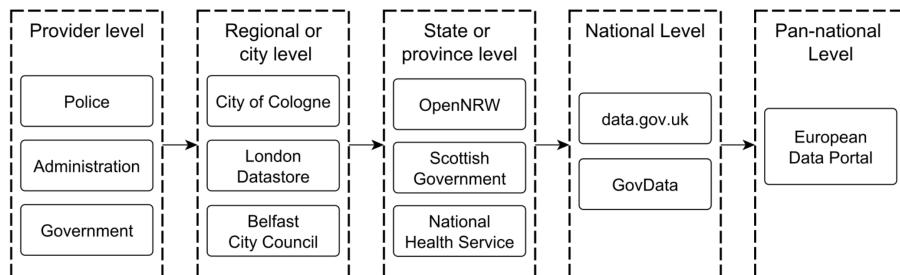
**piveau**

Use Case – data.europa.eu

**data.europa.eu**

## The Open Data Portal for Europe (data.europa.eu)

- More than 1.6 million datasets, gathered from more than 170 regional, national, and pan-national data catalogs



The screenshot shows the data.europa.eu portal homepage. The header includes the European Commission logo, a search bar, and navigation links for Home, Data, Academy, Community, Publications, and Documentation. The main content area is titled 'Datasets' and displays a search results page for 'Workforce Management Information - Defra'. The results show 1,541,430 datasets. Other visible datasets include 'Tariffs to trade with the UK from 1 January 2021' and links for CSV, Excel XL, HTML, UNKNOWN, and ODS formats.

## NFDI4DataScience

### German National Research Data Infrastructure for Data Science and Artificial Intelligence

- Makes available all digital research artifacts, including publications, data, models, workflows, and code/scripts
- Provides a registry for research knowledge graphs
- Internal representation as a research knowledge graph, connecting all digital research artifacts



A screenshot of the NFDI4 DataScience website's dataset search interface. The top navigation bar includes links for Datasets, Catalogues, SPARQL Editor, and Login. The main search area has a red header titled "Datasets Feed". Below it is a "Datasets" section with a map filter for "Leaflet | Eurostat - GISCO", a search bar, and buttons for Datasets, Catalogues, Editorial content, Last modified, Relevance, and More. It shows 49 datasets found, with one entry for "Integrating Long-Term Access into DataPLANT Data Management Workflows" by Zenodo Community CoRDI 2023. The entry includes a description, metadata (Created: 13.09.2023), and a download link. Another entry for "Integrating Data Literacy into University Curricula" is also visible.

## Registries for Semantic Models

---

### Takeaways and questions

- Registry should be based on Semantic Web technologies itself
- Registry should be simple, and should provide some additional features

#### Main Questions:

What is the state of the art in setting up registries (for semantic models)?

What special requirements do we have when it comes to registries for semantic models?

## Registries for Semantic Models

---

### Takeaways and questions

- What systems and tools can we learn from?
- Which systems already exist? Which systems should be incorporated/harvested?
- Which kind of semantic models should be supported?  
Who provides example semantic models?
- Who are the users? What functionality do they need?

#### Main Questions:

What is the state of the art in setting up registries (for semantic models)?

What special requirements do we have when it comes to registries for semantic models?

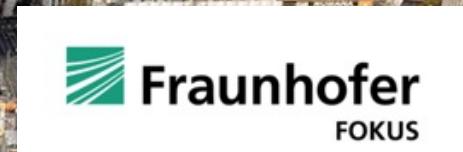
## Introduction

### Registries – Technical Considerations and Showcases



#### Main Questions:

- What is the state of the art in setting up registries?
- What special requirements do we have when it comes to registries for semantic models?



Fraunhofer Institute for Open Communication Systems

---

## Registries – Technical Considerations and Showcases

Prof. Dr. Sonja Schimmler

Fraunhofer FOKUS & TU Berlin

