ODCleanStore

ETL Tool for RDF Data

Tomas Knap

Charles University in Prague,
Department of Software Engineering,
XML and Web Engineering Research Group

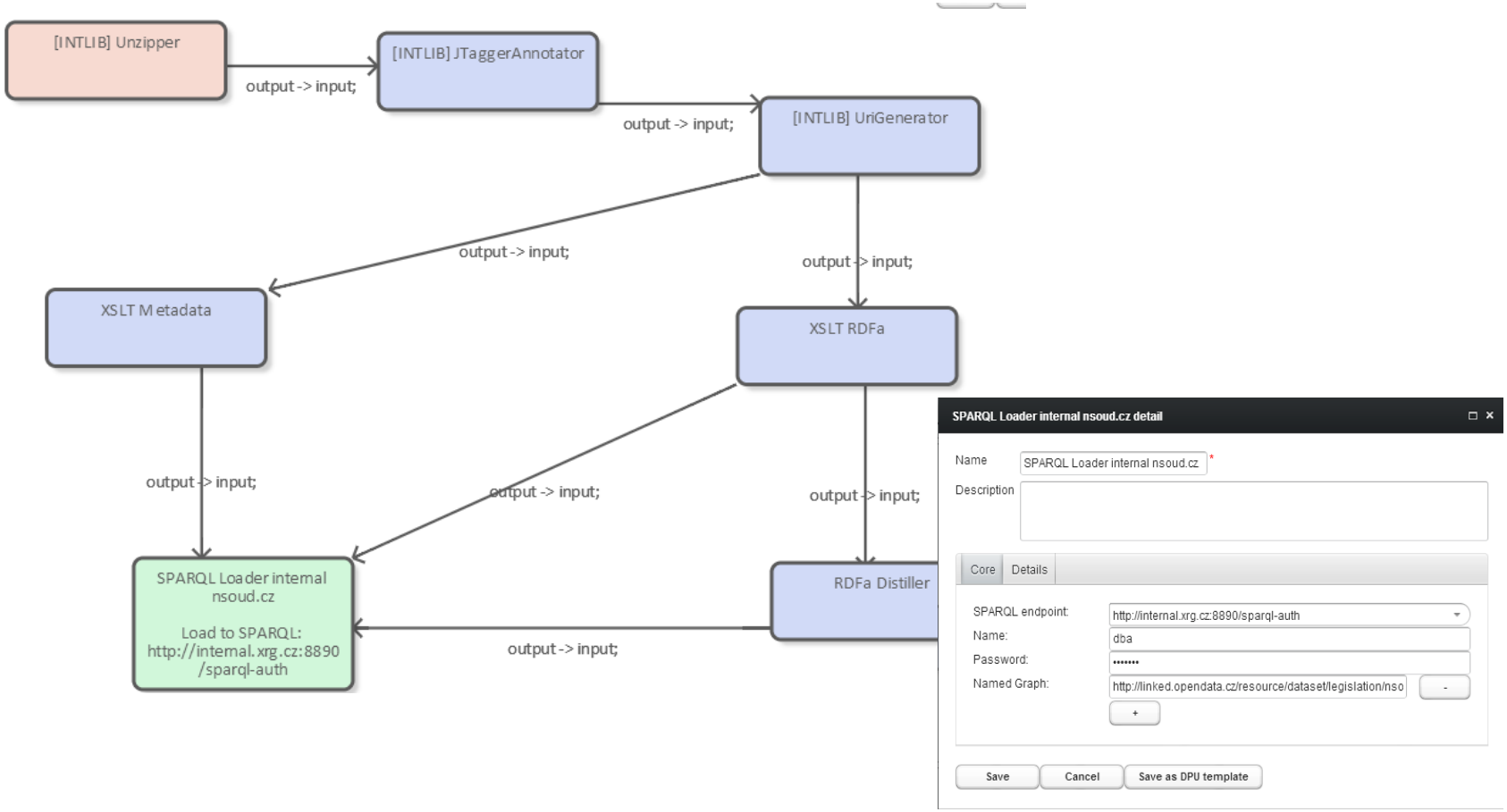
The COMSODE project has received funding from the Seventh Framework Programme of the European Union in the grant agreement number 611358.

Motivation

- Tools available for RDF data extraction, enrichment, linking, transforming, ...
- No tool for flexible management of RDF data extraction/transformation **tasks**
  - task = progression of data processing units (DPUs)
    - Extract data from SPARQL Endpoint A
    - Extract data from CSV files B
    - Refine data with SPARQL queries X,Y, Z
    - Deduplicate data using Linker L
    - Publish data to SPARQL Endpoint B

COMSODE

ETL Task



COMSODE

ODCleanStore

- Framework for managing ETL tasks and data processing units (DPUs)
- Developed in the EU FP7 Project COMSODE. Based on the previous development effort at Charles University in Prague (with support from Semantic Web Company, Austria)

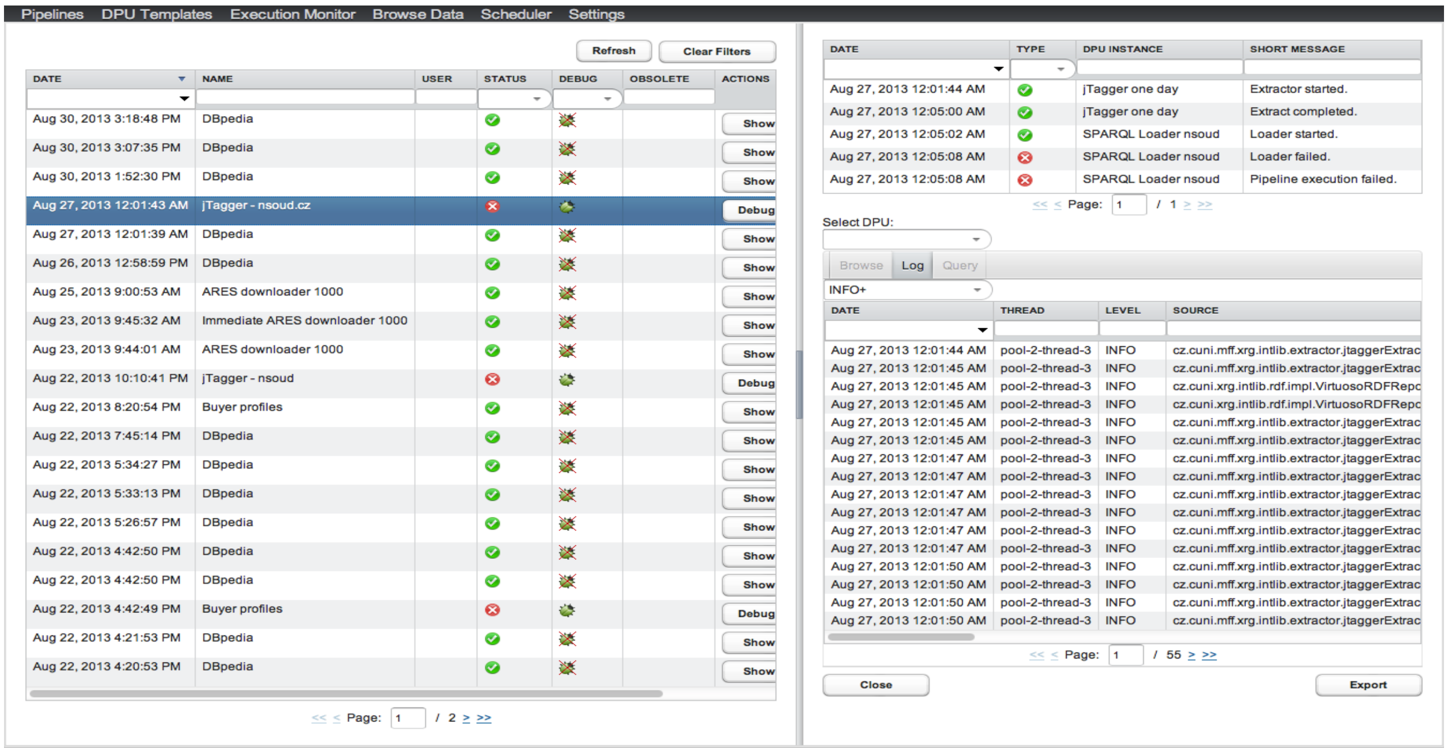
COMSODE

Key Features

- Administration interface to create/configure/manage tasks, DPUs
  - Possibility to schedule tasks
    - Time based, chaining of tasks
  - Possibility to debug tasks, browse logs/events produced by DPUs
  - Multi-user environment, possibility to share pipelines, DPUs
- Robust engine running the tasks
- Core DPUs available
  - Easy way how to add your own DPUs

Demo

- http://odcs.xrg.cz:8080/odcleanstore/



COMSODE Project

- EU FP7 project
- Goals: to provide publication platform for publishing (Linked) Open Data, enable searching on top of the published data
  - It will use ODCleanStore for preparing the data
  - Proofed on datasets from various institutions
  - We are just starting, building the user group!
    - Please contact me if your institution would like to participate
  - To follow the news: FP7 COMSODE linkedIn Group
  -

COMSODE

Conclusions

- ODCleanStore
  - Framework for managing ETL tasks and data processing units (DPUs)
  - Developed in COMSODE project
- Used:
  - in COMSODE Project
  - in Open Data activities in Czech Republic
  - by Semantic Web Company and their customers (planned)
  - LOD2 stack (in preparation)

COMSODE

Thank You!

- Would you like to try ODCleanStore?
  - http://www.ksi.mff.cuni.cz/~knap/odcs/
- Would you like to know more about the tool? Would you like to participate in the user group of COMSODE?
  - knap@xrg.cz

www.comsode.eu

COMSODE