2020-12-16 VIVO-DataConnect ORCID Demo

Summary

Reference:

- GItHub Source code: https://github.com/vivo-community/vivo-data-connect/tree/POC-extract-orcid
- VIVO-conf. 2020: VIVO-DATACONNECT: TOWARDS AN ARCHITECTURAL MODEL FOR INTERCONNECTING HETEROGENEOUS DATA SOURCES TO POPULATE THE VIVO TRIPLESTORE
- ORCID useful information

Use case

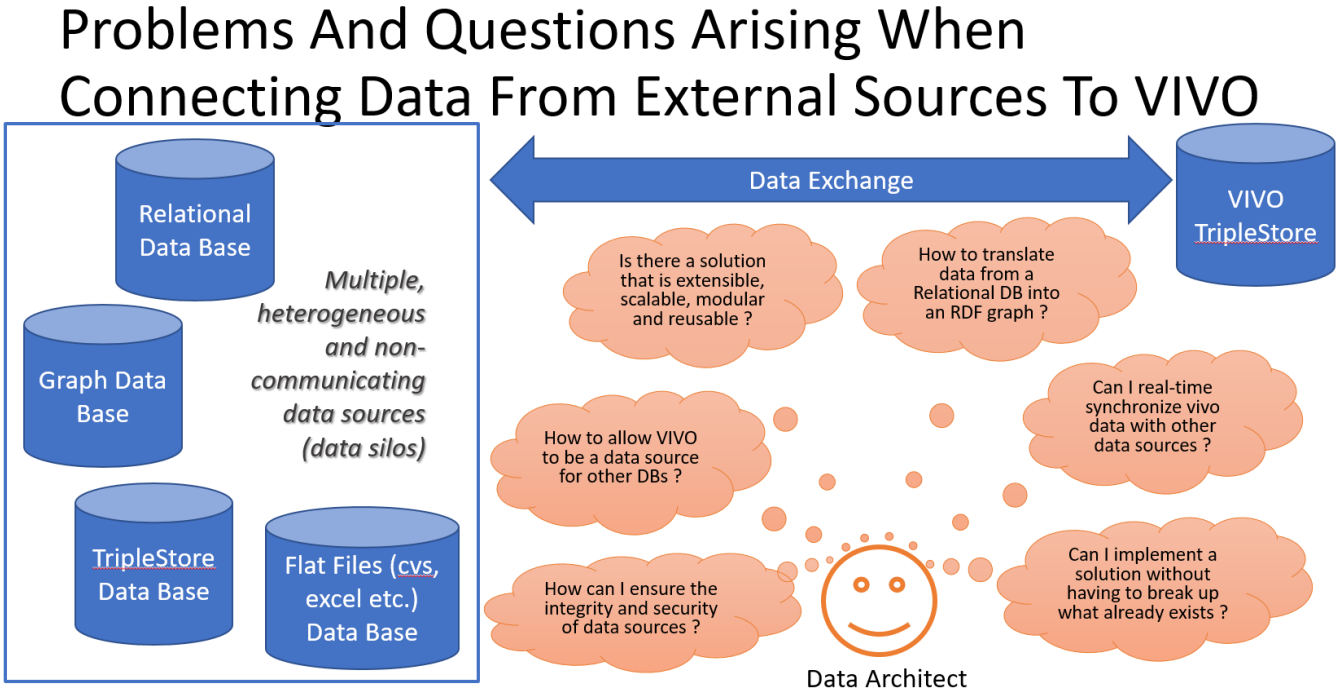
Epic

A professor wishes to add the reference to a scientific article, Irrespective of whether he chooses ORCID or VIVO, the information he will enter in either of these platforms will be mutually updated,

User Story

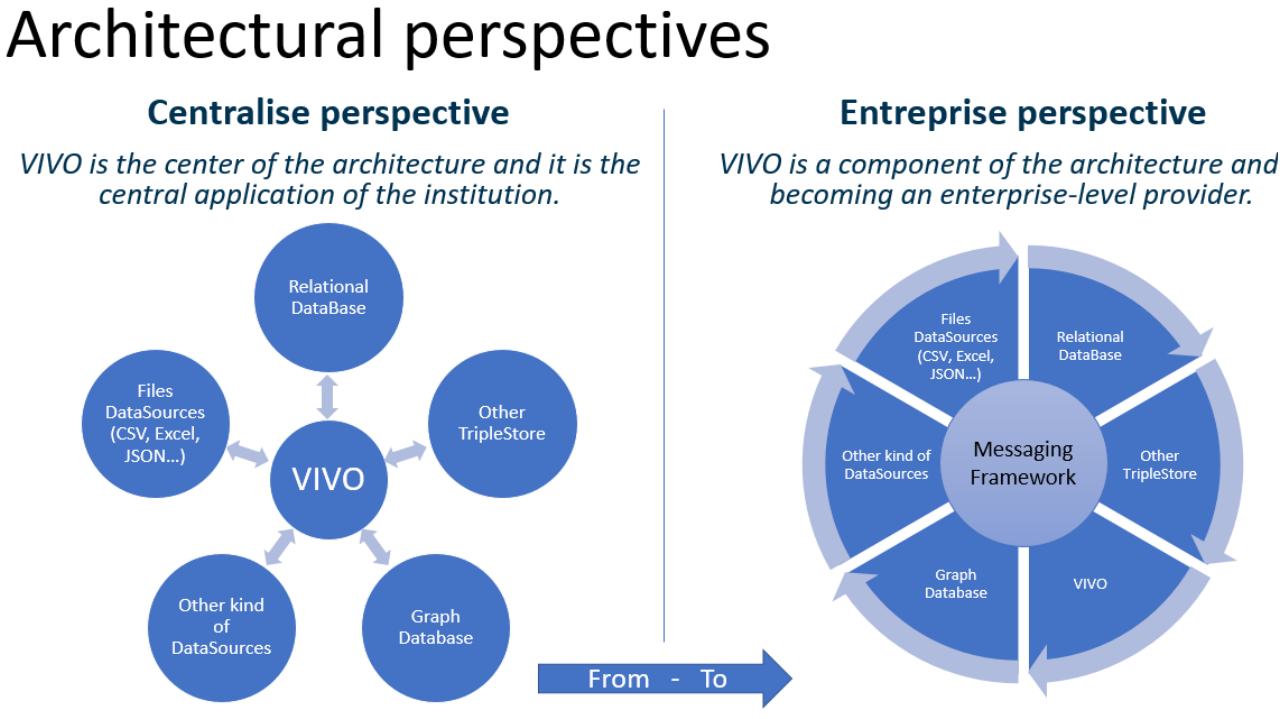
Migrating ORCID data to a VIVO instance

Issue



About Kafka

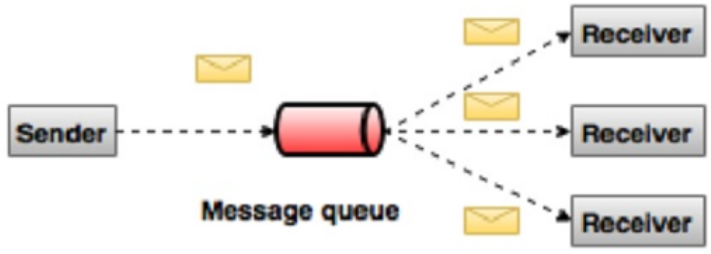
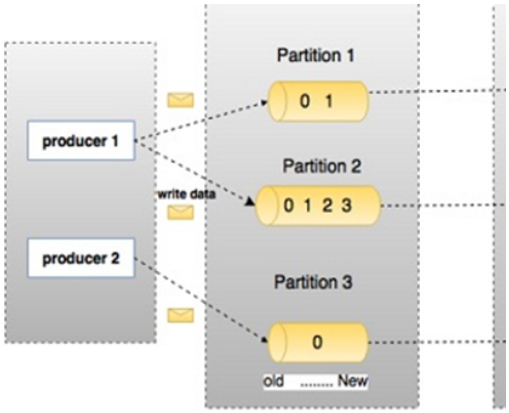
Goal of using Kafka



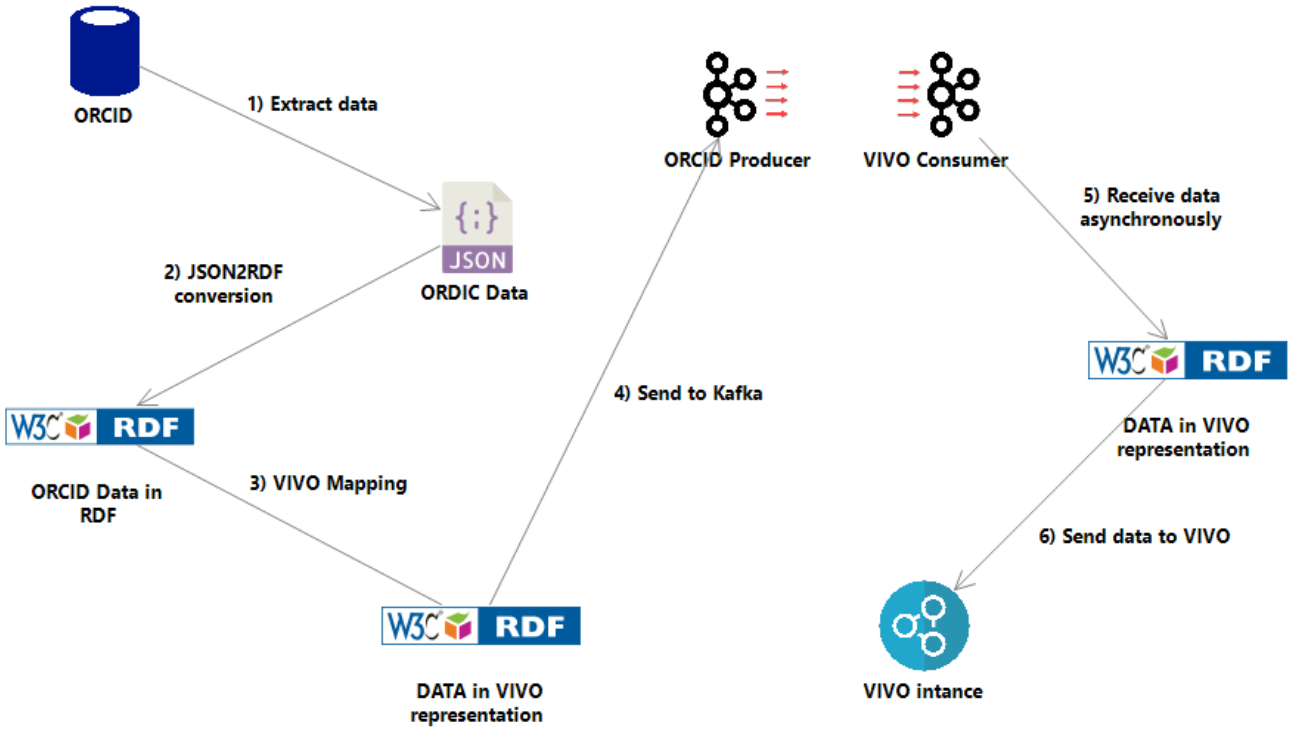
What is Kafka?

see also https://kafka.apache.org/intro

Messaging system	Event streaming

Event streaming thus ensures a continuous flow and interpretation of data so that the right information is at the right place, at the right time.	1. To **publish** (write) and **subscribe to** (read) streams of e your data from other systems. 2. To **store** streams of events durably and reliably for as lo 3. To **process** streams of events as they occur or retrospe
	

ORCID to VIVO Dataflow through Kafka



Dataflow Implementation

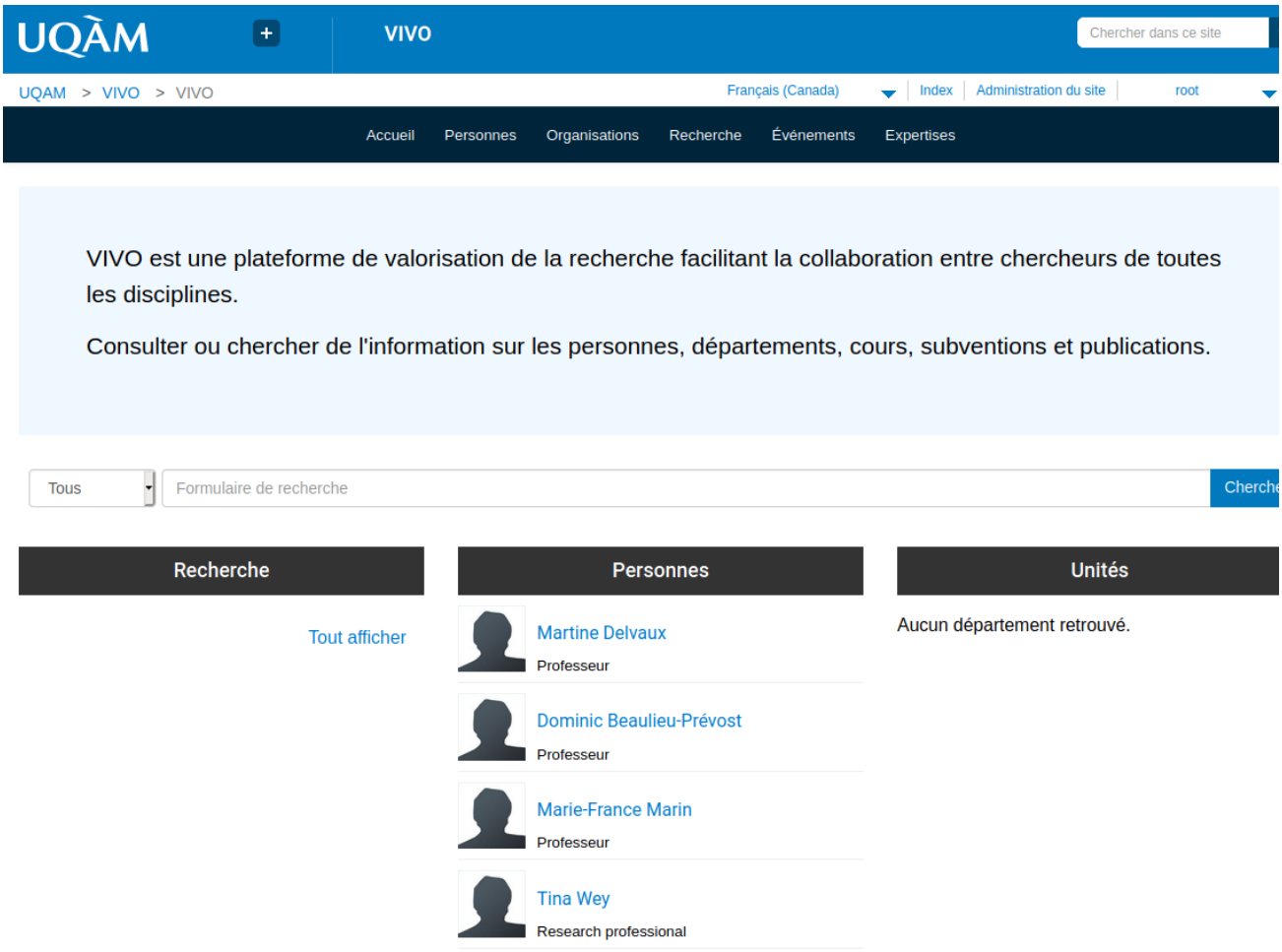
Prerequisite

- Installing and first run of Kafka https://kafka.apache.org/quickstart and https://computingforgeeks.com/configure-apache-kafka-on-ubuntu/
- Installation de Jena
- Installation de json2rdf : https://github.com/AtomGraph/JSON2RDF

Dataflow execution

- 2020-12-16 VIVO-DataConnect ORCID Demo - 1) Extract data from ORCID and 2) Convert to RDF
- 2020-12-16 VIVO-DataConnect ORCID Demo - 3) VIVO Mapping
- 2020-12-16 VIVO-DataConnect ORCID Demo - 4) Send to Kafka - 5) Receive data - 6) Send to VIVO

Results



In summary

It has been shown that it is possible to use Kafka to populate VIVO from ORCID

Several points require special attention

- The ORCID ontology needs to be refined and clarified.
- The mapping between ORCID and VIVO also needs to be worked on
- The structure of the Kafka message has to be designed to respect the add/delete/modify record actions
- Several minor bugs need to be fixed in the scripts.

Plan for future

- Building a POC VIVO Kafka ORCID
- Proving the architecture to operate in event-driven and real-time mode
- Getting POCs to Java
- Redesigning the mapping process, ORCID ontology structure and message structure