

# 2020-12-16 VIVO-DataConnect ORCID Demo

## Summary

- [Reference:](#)
- [Use case](#)
  - [Epic](#)
  - [User Story](#)
- [Issue](#)
- [About Kafka](#)
  - [Goal of using Kafka](#)
  - [What is Kafka?](#)
- [ORCID to VIVO Dataflow through Kafka](#)
  - [Dataflow Implementation](#)
    - [Prerequisite](#)
    - [Dataflow execution](#)
    - [Results](#)
- [In summary](#)
  - [Several points require special attention](#)
  - [Plan for future](#)

## Reference:

- [GitHub Source code: https://github.com/vivo-community/vivo-data-connect/tree/POC-extract-orcid](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

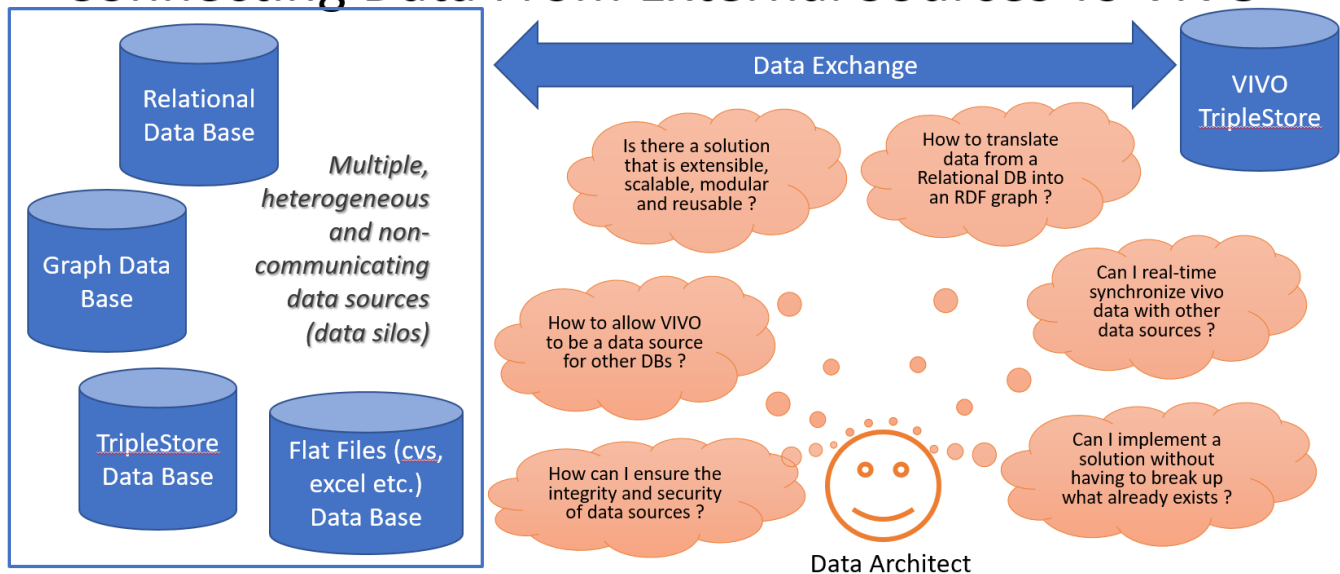
A professor wishes to add the reference to a scientific article, Irrespective of whether he chooses ORCID or VIVO, the information he will enter in either of these platforms will be mutually updated,

### User Story

Migrating ORCID data to a VIVO instance

## Issue

# Problems And Questions Arising When Connecting Data From External Sources To VIVO



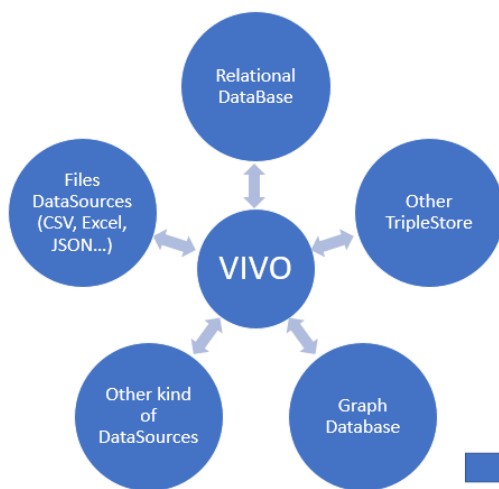
## About Kafka

Goal of using Kafka

## Architectural perspectives

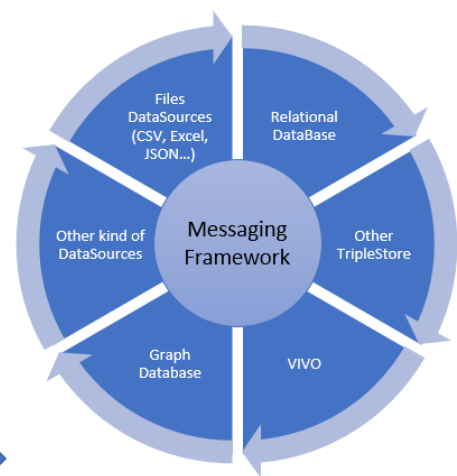
### Centralise perspective

*VIVO is the center of the architecture and it is the central application of the institution.*



### Enterprise perspective

*VIVO is a component of the architecture and becoming an enterprise-level provider.*



## What is Kafka?

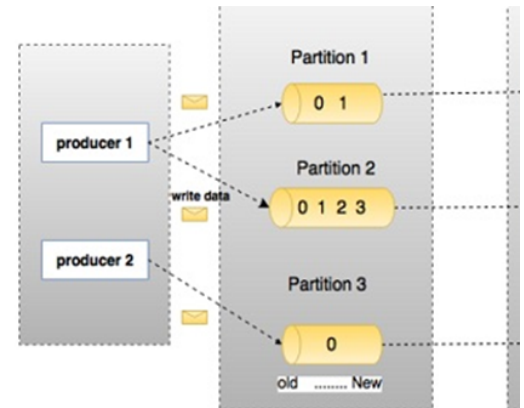
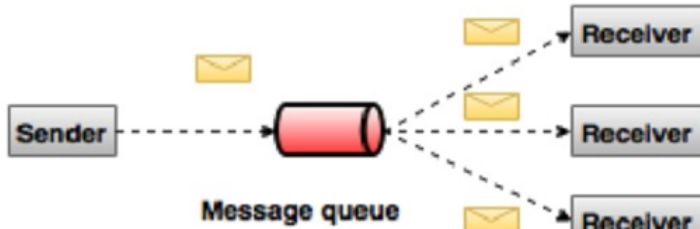
see also <https://kafka.apache.org/intro>

Messaging system

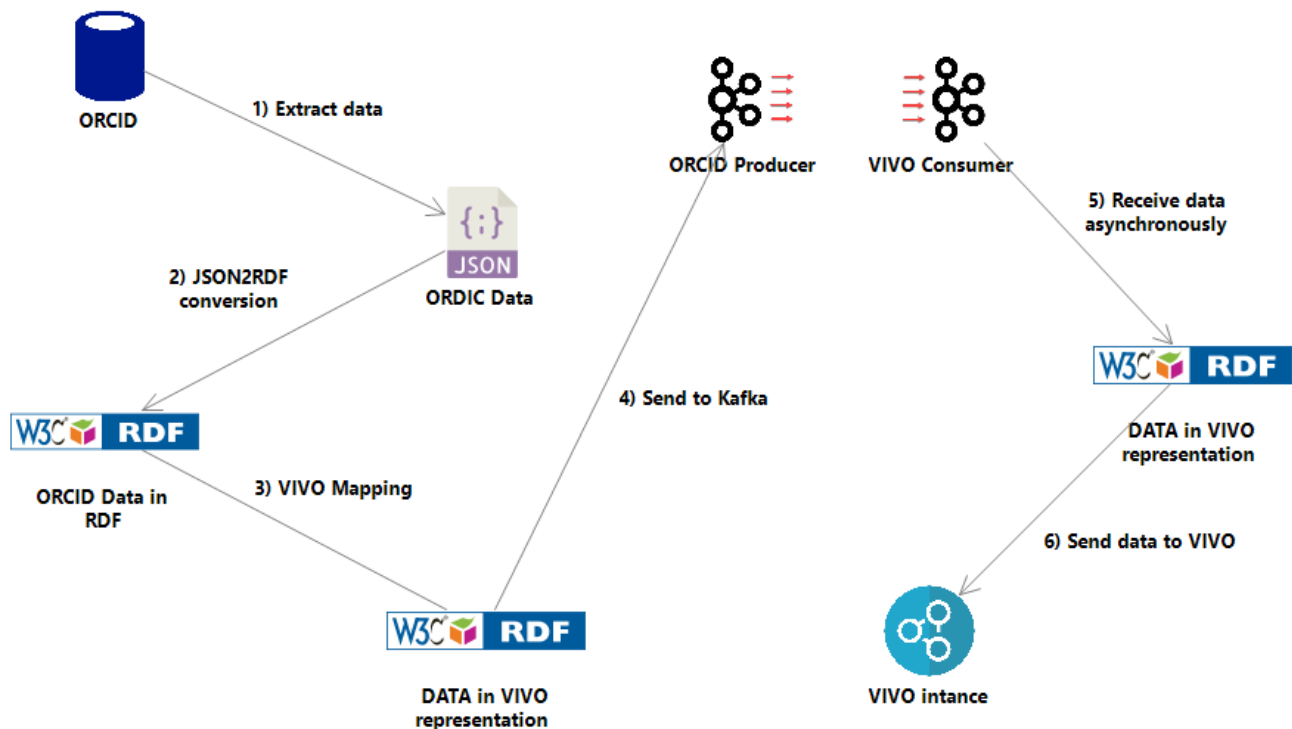
Event streaming

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 your data from other systems.
2. To **store** streams of events durably and reliably for as long as needed.
3. To **process** streams of events as they occur or retrospectively.



## 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

The screenshot displays the VIVO platform interface. At the top is a blue header with the UQAM logo, a search bar, and the word 'VIVO'. Below the header is a navigation bar with links: Accueil, Personnes, Organisations, Recherche, Événements, and Expertises. The main content area has a light blue background with text explaining VIVO's purpose: 'VIVO est une plateforme de valorisation de la recherche facilitant la collaboration entre chercheurs de toutes les disciplines. Consulter ou chercher de l'information sur les personnes, départements, cours, subventions et publications.' Below this is a search bar with a dropdown menu set to 'Tous' and a 'Chercher' button. The search results are displayed in three columns: 'Recherche' (with a 'Tout afficher' link), 'Personnes' (listing four individuals: Martine Delvaux, Dominic Beaulieu-Prévost, Marie-France Marin, and Tina Wey), and 'Unités' (showing 'Aucun département retrouvé.').

UQAM + VIVO Chercher dans ce site

UQAM > VIVO > VIVO Français (Canada) Index Administration du site root

Accueil Personnes Organisations Recherche Événements Expertises

VIVO est une plateforme de valorisation de la recherche facilitant la collaboration entre chercheurs de toutes les disciplines.

Consulter ou chercher de l'information sur les personnes, départements, cours, subventions et publications.

Tous Formulaire de recherche Chercher

Recherche Personnes Unités

Tout afficher

Martine Delvaux  
Professeur

Dominic Beaulieu-Prévost  
Professeur

Marie-France Marin  
Professeur

Tina Wey  
Research professional

Aucun département retrouvé.

## 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