



Assessing the Viability of an OBDA Approach to Integrating Databases into a Virtual Knowledge Graph for Research Management at ITC

Phearun OL¹, Deperias KERRE², Christophe DEBRUYNE², Dona VALY³, and Sotheany NOU³

¹Graduate School, Institute of Technology of Cambodia, Phnom Penh, Cambodia

²Montefiore Institute, University of Liège, 4000 Liège, Belgium

³Research and Innovation Center, Institute of Technology of Cambodia, Phnom Penh, Cambodia



Motivation & Context

ITC runs three independent information systems that are valuable individually but hard to see across:

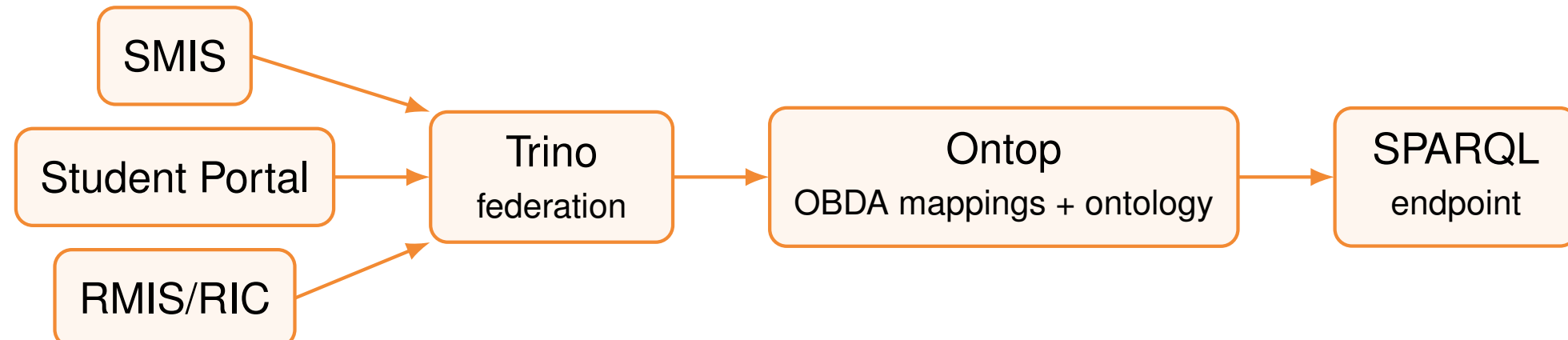
1. SMIS → students, staff, departments, academic records, ...
2. Student Portal → student-facing academic data.
3. RMIS/RIC → research profiles, projects, publications, ...

Challenges:

- Heterogeneous representations.
- No single global identifier is used across systems.

This separation limits the visibility and reuse of institutional knowledge.

Approach: OBDA / Virtual Knowledge Graph



This work is part of the ARES Institutional Support Program 2022–2027 at ITC, which favors reusing existing infrastructure over replacing it. A VKG offers a **cost-conscious** solution to the problem.

Ontology & Mapping

The focus was assessing the viability of VKGs, so the ontology was kept simple and built from scratch.

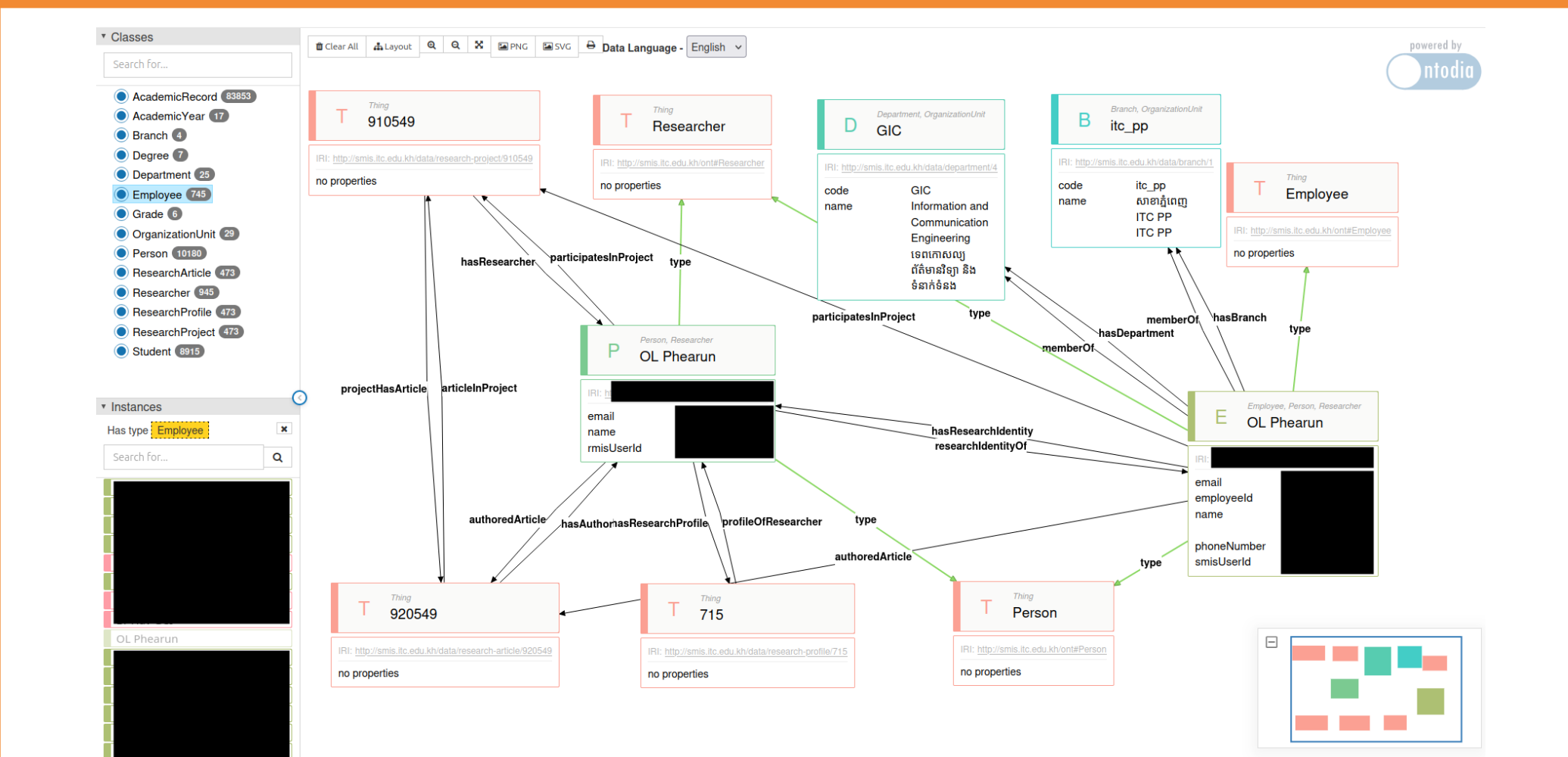
- Classes: Employee, Student, Researcher, Department, Branch, ...
- Properties: hasDepartment, hasAcademicRecord, participatesInProject, ...

Linking entities relies on source-specific IDs and bridge columns:

- Portal ↔ SMIS students via `id_card`
- SMIS ↔ RMIS/RIC via bridge table using Trino

Mappings were authored in Ontop's native OBDA syntax rather than R2RML → faster to write and iterate for the conditional joins needed here.

Demonstration



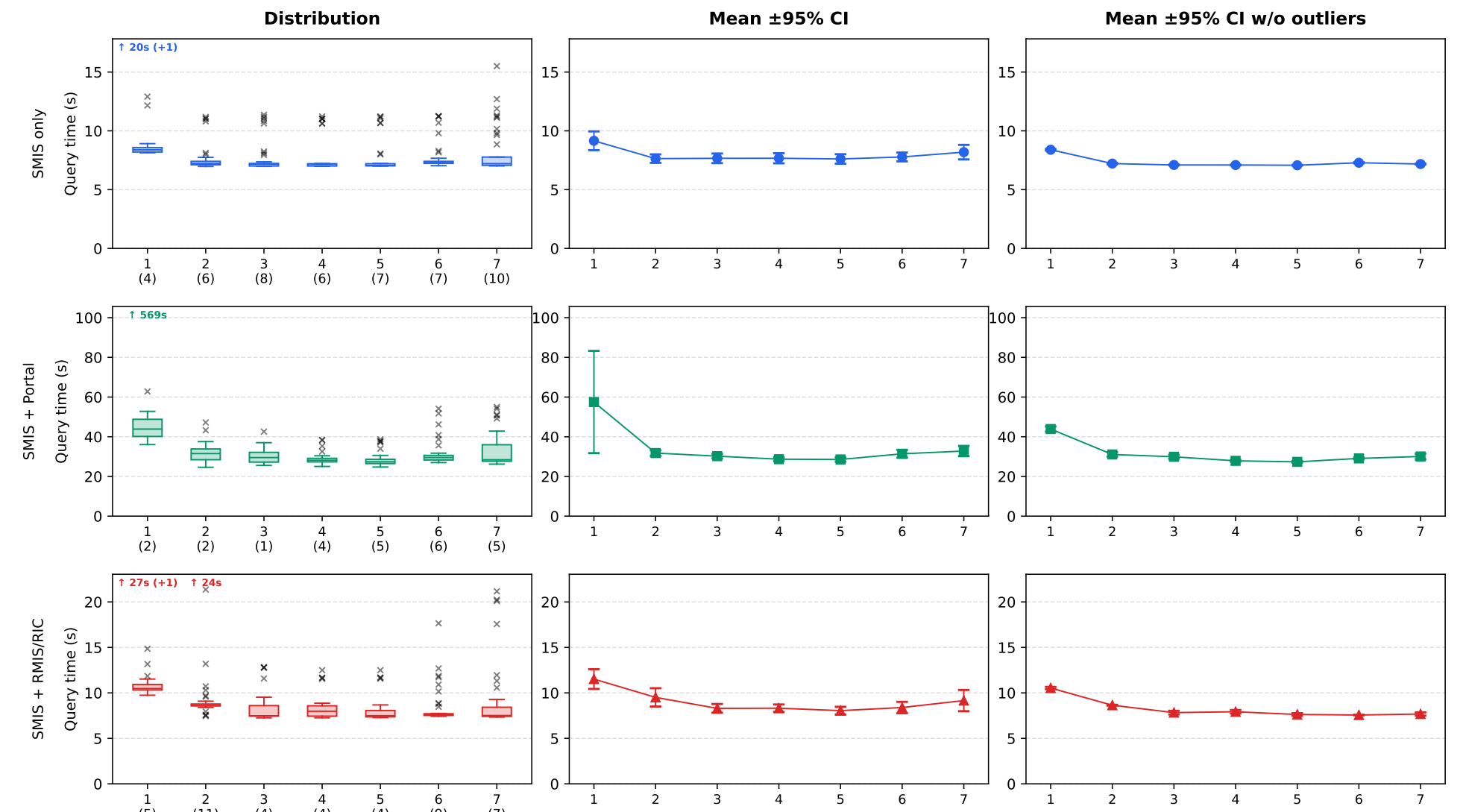
Ontodia demo of the VKG: cross-system semantic links between an SMIS employee, their department, and the corresponding RMIS/RIC research entities.

Resources

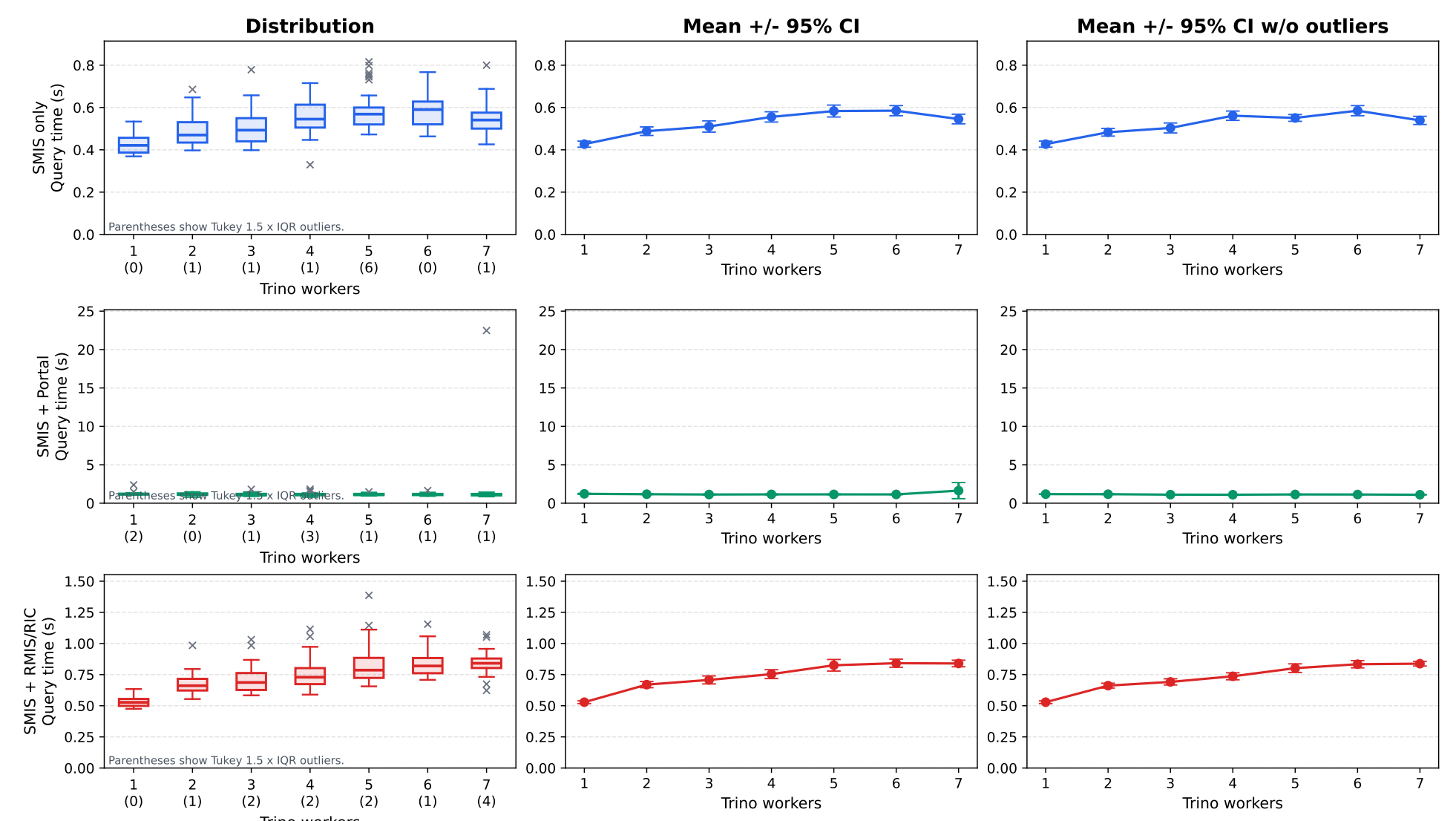
Ontology, mappings, and benchmark queries:
<https://github.com/Phearun2020/itc-rmis-kg>

Evaluation: Trino Worker Scaling

Three query cases were run over 1–7 Trino workers, 40 randomized runs each, under two network conditions.



Over VPN, network latency dominates → adding workers helps up to a point (3 workers capture most gains). *Machine: AMD Ryzen 7 5700G (8c/16t), 128 GB RAM, Windows 11.*



With direct access, latency drops, but a single worker becomes optimal → distributed-execution overhead now outweighs parallelism for these small queries. *Machine: KVM (QEMU), 16 vCPUs (host: Xeon Gold 6338), 31 GB RAM, Ubuntu 24.04.*

Lessons Learned in this Project

- VKG/OBDA **allows us to integrate** SMIS, Portal, and RMIS/RIC without touching the source systems. The approach is viable and allows us to avoid building new systems or extending existing ones.
- Whilst Trino supports querying different sources using different workers, it may not be necessary provided the queries are small and/or data is manageable.
- Ontop's mapping-based query rewriting adds SQL complexity that itself contributes to latency, independent of data volume or network.

Conclusions & Future Work

OBDA/VKG is a **viable** and **cost-conscious** foundation for institutional KG development at ITC, suited to read-only integration of live data.

Open challenges:

- Identity alignment across systems → institutional challenge
- More testing with additional competency questions

Next steps:

- Compare against a materialized RDF graph
- Evaluate reuse of established vocabularies (e.g., VIVO)



ACADÉMIE
DE RECHERCHE ET
D'ENSEIGNEMENT
SUPÉRIEUR