Pedro Guimarães

2023 – Today

PhD Candidate, Information Systems and Technologies

University of Minho

CCG /ZGDV | ICT Innovation Institute

2017 - Today

Researcher & Team Leader

20+ R&D Projects

15+ Scientific Publications

Google Data Analytics Certificate

Certified ScrumMaster® (CSM®)

https://prg.dsi.uminho.pt/

pedro.guimaraes@ccg.pt

Outline

Motivation and Goal

Domain reference model

Knowledge Graphs

Gaps and Research Opportunities

Our Proposal

Future Work

Motivation

Holly grail of data engineering may be **data integration**, a full interconnected and integrated world of data.

One of the main challenges of data integration is the existence of **many heterogeneous** data sources and data models (that are constantly evolving), even when **describing** the **same domain**.

This is due manly to the own perspective, knowledge and past experience from the domain expert (or body of domain experts) that is (are) modeling the data model.

https://oaei.ontologymatching.org/2022/knowledgegraph/index.html
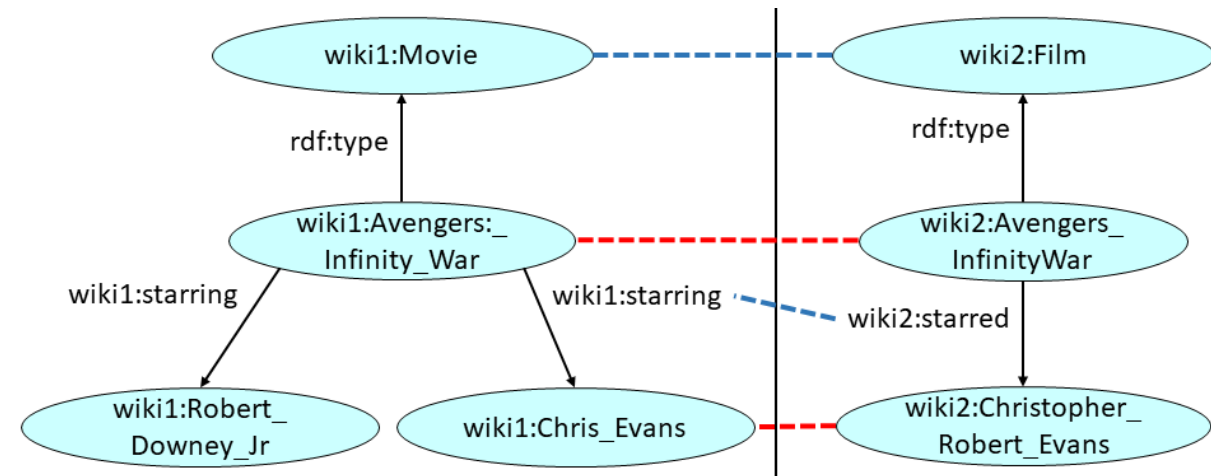
Goal
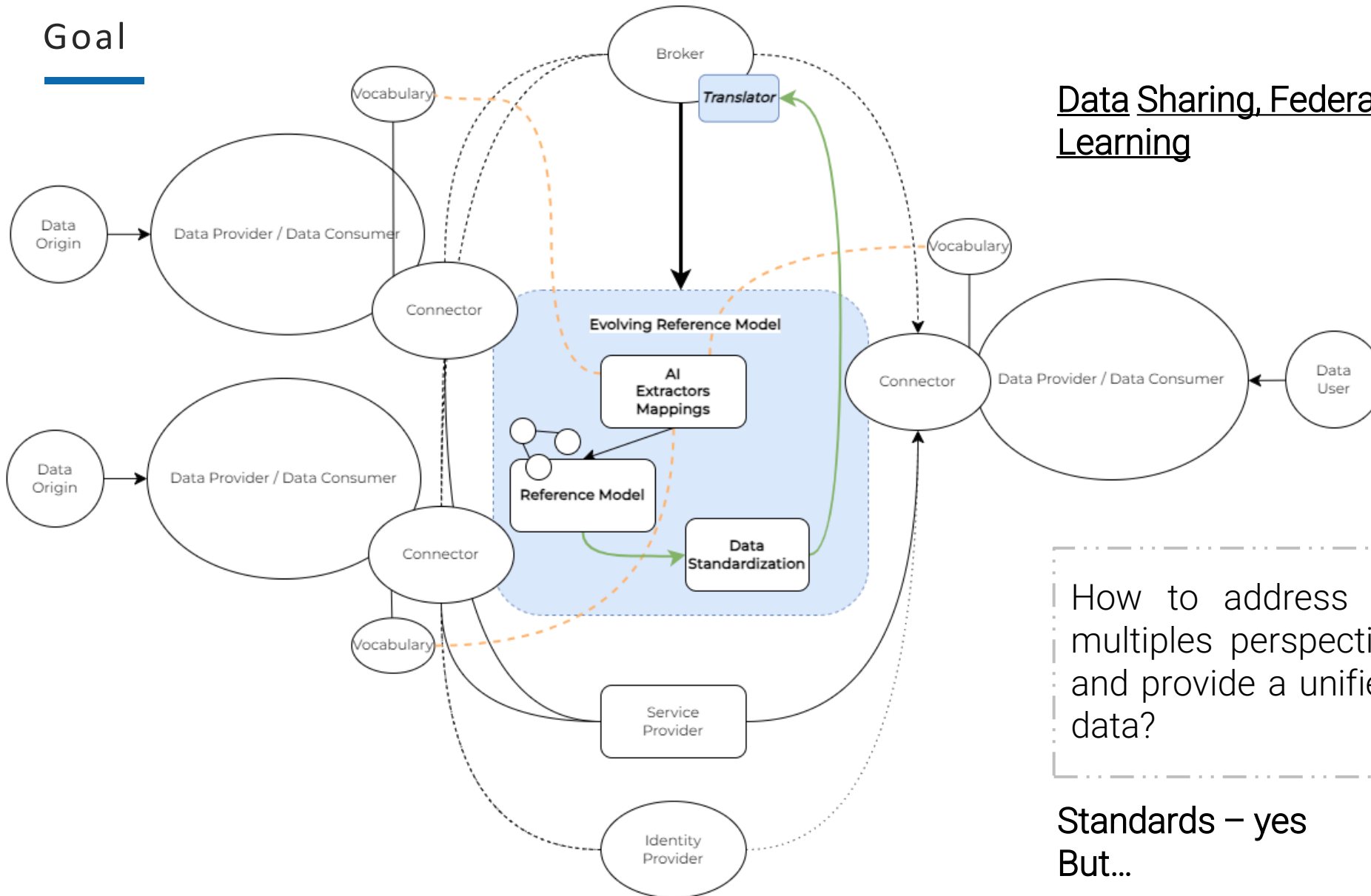
Data Sharing, Federated Data, Machine Learning

How to address the challenge of having multiples perspectives of the same domain and provide a unified view of knowledge and data?

Standards – yes
But…

Domain reference model

A **model** is an abstraction of something of the real-world that is created to better understand it

In software engineering, a **domain model** is a conceptual model of the domain that incorporates both behavior and data

Fowler, Martin. Patterns of Enterprise Application Architecture. Addison Wesley, 2003, p. 116.

Reference models are often called "universal models, generic models, or model patterns".

Fettke, P., & Loos, P. (2003). Classification of reference models - A methodology and its application. Inf. Syst. E-Business Management, 1, 35–53. https://doi.org/10.1007/BF02683509

Reference model as a domain specific structure that clearly expresses the domain expertise produced by an expert or body of experts.

Gray, J., & Rumpe, B. (2021). Reference models: how can we leverage them? Softw. Syst. Model., 20(6), 1775–1776. https://doi.org/10.1007/s10270-021-00948-0

Knowledge graphs

❑ Knowledge graphs (KGs) can scale to accommodate vast amounts of data and complex interrelationships.

❑ The schema-less nature of KGs allows for dynamic updates and additions.

❑ KGs enable the discovery of new relationships and patterns through inference and reasoning.

❑ KGs can be leveraged by NLP algorithms to enhance the understanding and processing of natural language data.
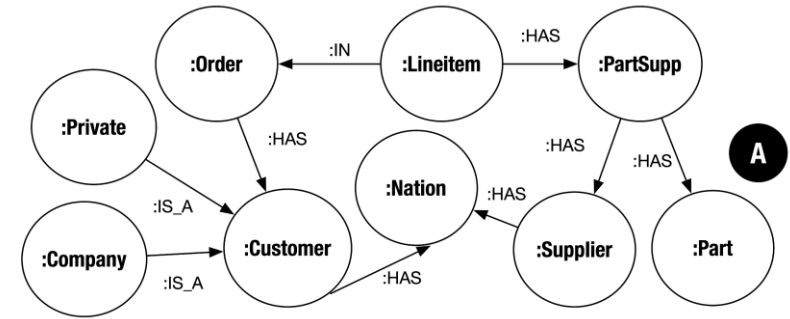

Labeled Property Graph
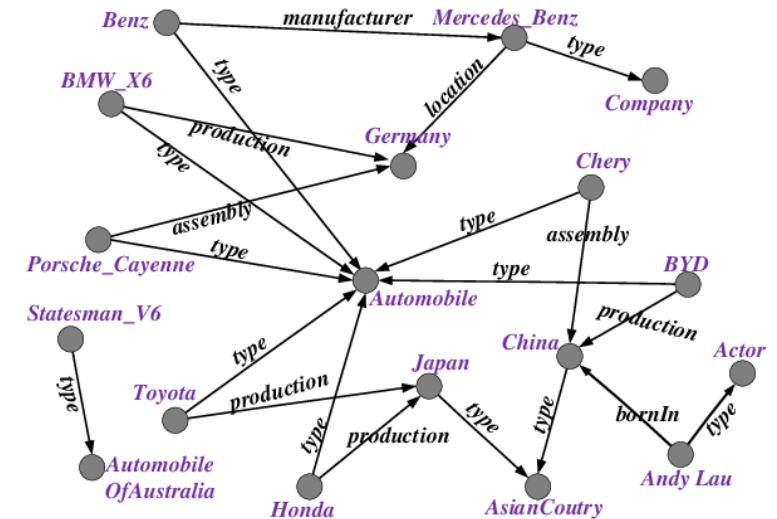
Zheng, W., Zou, L., Peng, W., Yan, X., Song, S., & Zhao, D. (2016). Semantic SPARQL Similarity Search Over RDF Knowledge Graphs. *Proc. VLDB Endow., 9*, 840-851.

Gaps and Research Opportunities

☐ What are the characteristics of an **evolving** domain reference model?

☐ How to combine multiple inputs such as ontologies, standards, vocabularies or data to enhance the domain reference model?

☐ Old facts, re-assess the prior knowledge?

Paper	Add New	Update Existing	Remove Outdated
Tu et al. [13]	●	●	●
Ilyas et al. [14]	●	●	●
Yang and Liao [41]	●	●	○
Grangel-González et al. [58]	●	●	○
Wang et al. [109]	◐	●	○
Yang et al. [101]	●	●	○
Wang et al. [93]	◐	●	◐
Obraczka et al. [44]	●	●	○
Wang et al. [95]	●	●	○
Liu et al. [111]	●	●	○
Jia et al. [85]	●	●	○

- 12 papers addressing some of the evolving characteristics of a KG.
- Reduced use of LLMs
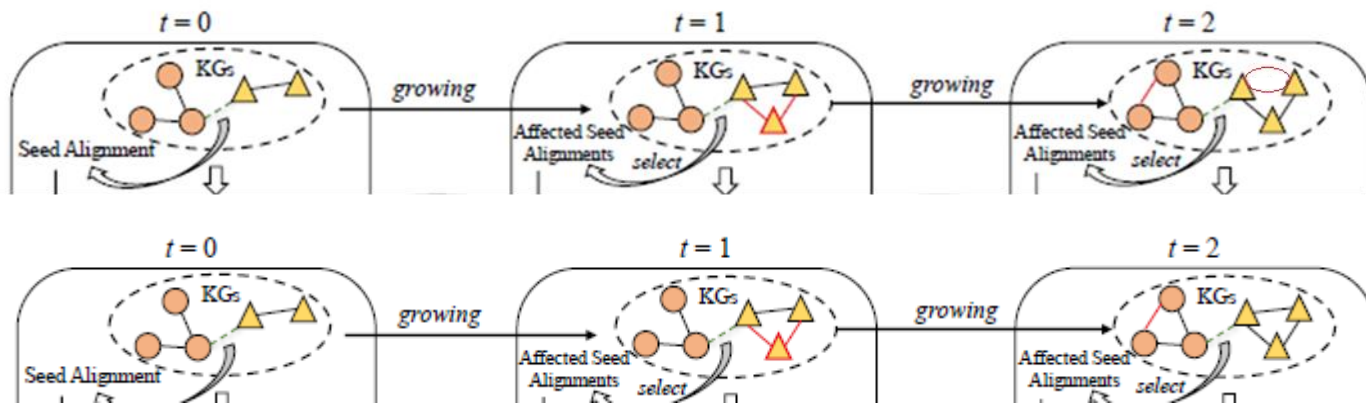
Re-assess the prior knowledge, old facts are still valid?

Wang, Y. *et al.* (2022). Facing Changes: Continual Entity Alignment for Growing Knowledge Graphs. In: Sattler, U., *et al.* The Semantic Web – ISWC 2022. ISWC 2022. Lecture Notes in Computer Science, vol 13489. Springer, Cham. https://doi.org/10.1007/978-3-031-19433-7_12
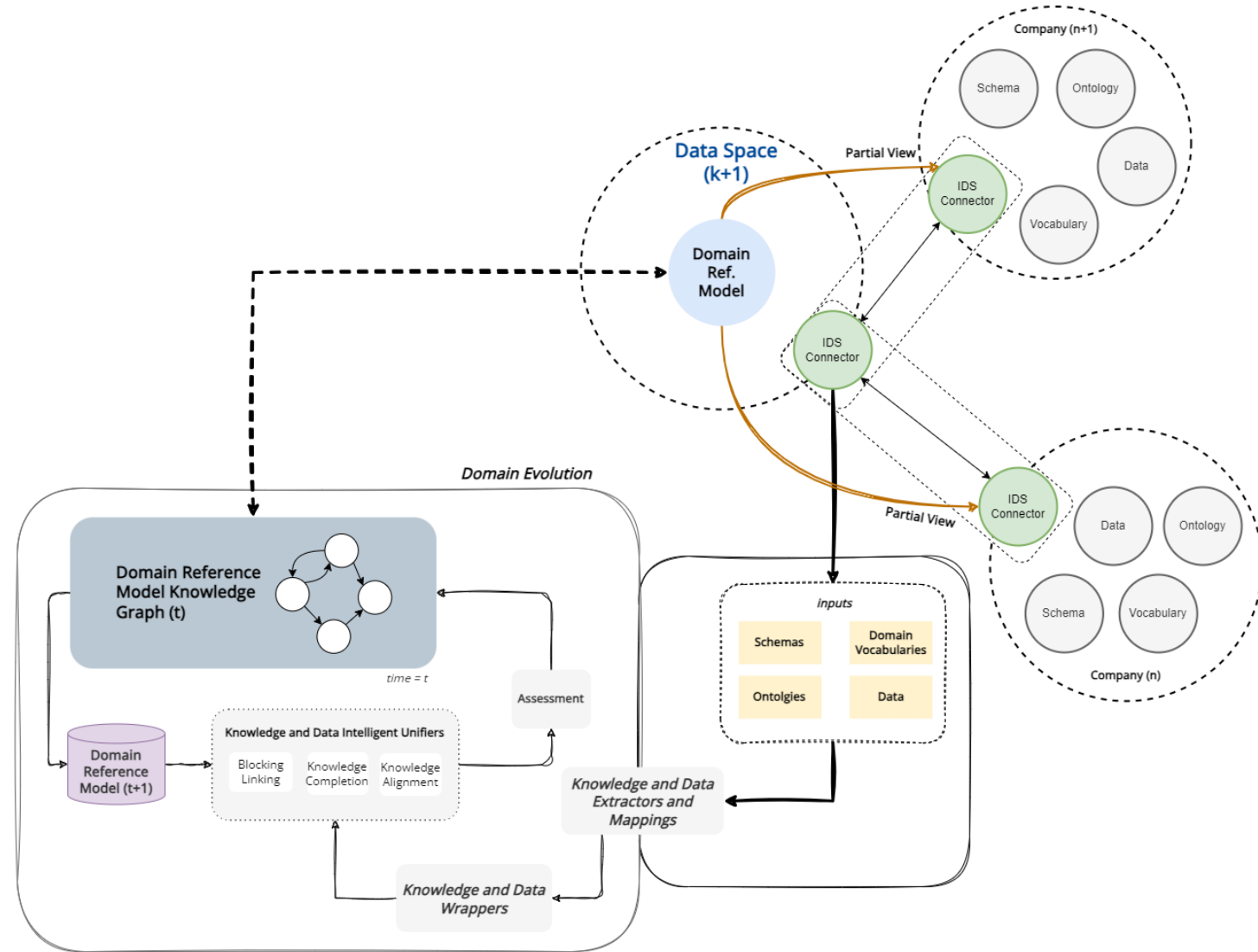
Our proposal

Incremental Evolution Continuously updating the domain reference model with new knowledge and data.

Intelligent Unification Ensuring the integrated knowledge is accurate, complete, and aligned.

Dynamic Adaptation Iteratively enhancing the reference model to keep pace with evolving domains and data sources

Future Work

•Address challenges of outdated or inaccurate information.

•Continuously improve alignment processes to manage the evolving nature of the domain.

•Develop adaptive artifacts (methods, models, components) for seamless integration and alignment.

•Refine knowledge and data extraction and mapping processes.

•Test the framework in real-world Data Space scenarios.

Thank you for your attention.