POLITECNICO DI TORINO

Dipartimento di Automatica e Informatica

PhD in Computer and Control Engineering

XXXII cycle

Supervisor

Prof. Juan Carlos De Martin

From Linked Open Data to Big Data: Architectures and Methodologies

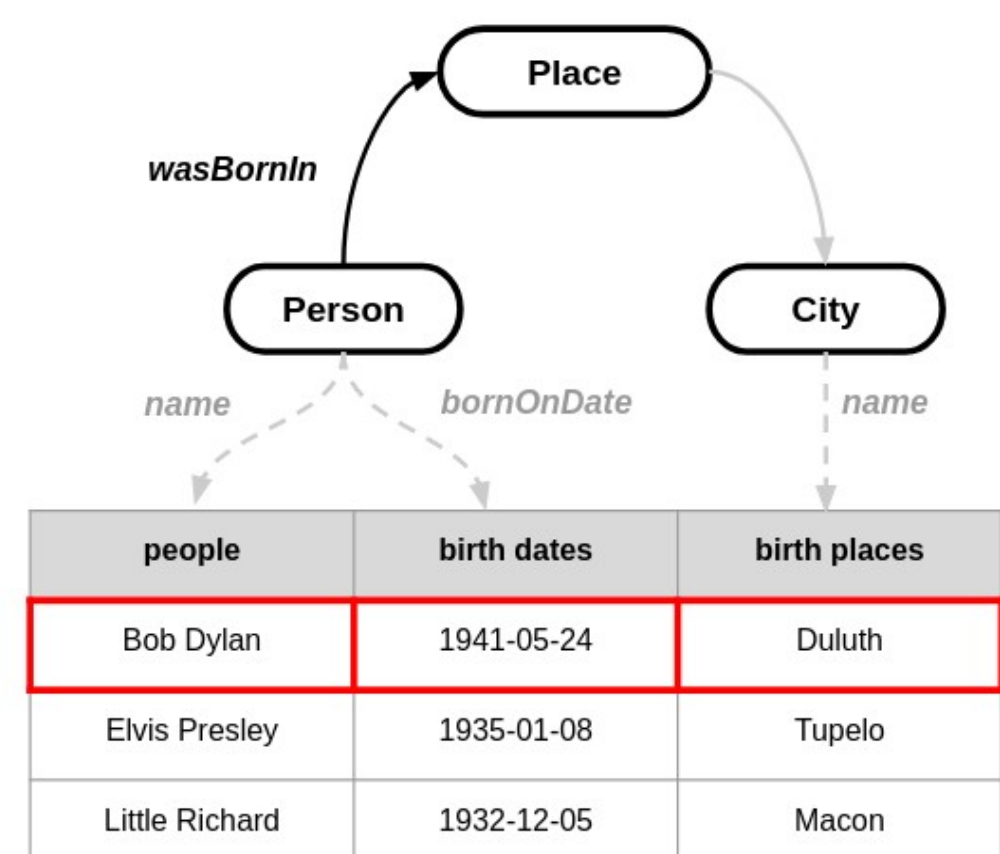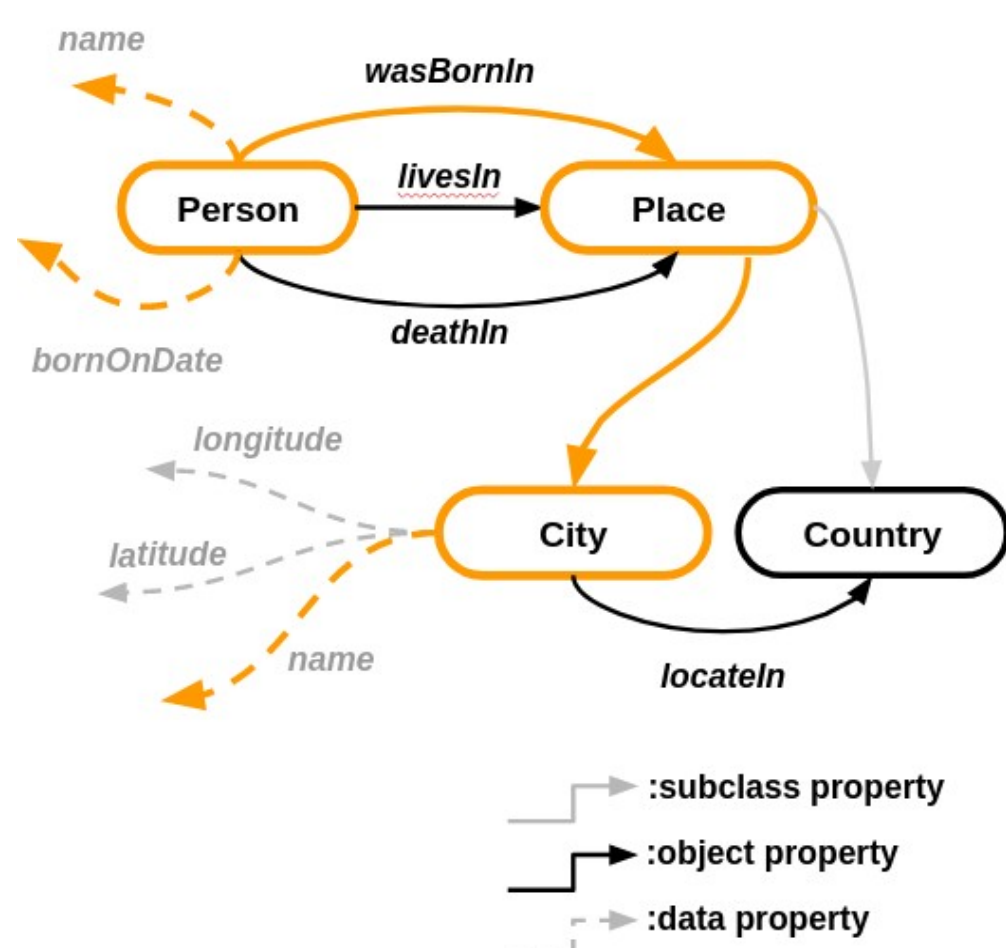PhD Candidate:

Giuseppe Futia

1. Context

The construction of **Knowledge Graphs (KGs)**, which are large networks of entities and relations relevant to a specific domain, is a complex task: it requires to **semantically integrate information** from heterogeneous data sources. Currently, such integration is not yet scalable, because it still requires significant manual effort and domain expertise.
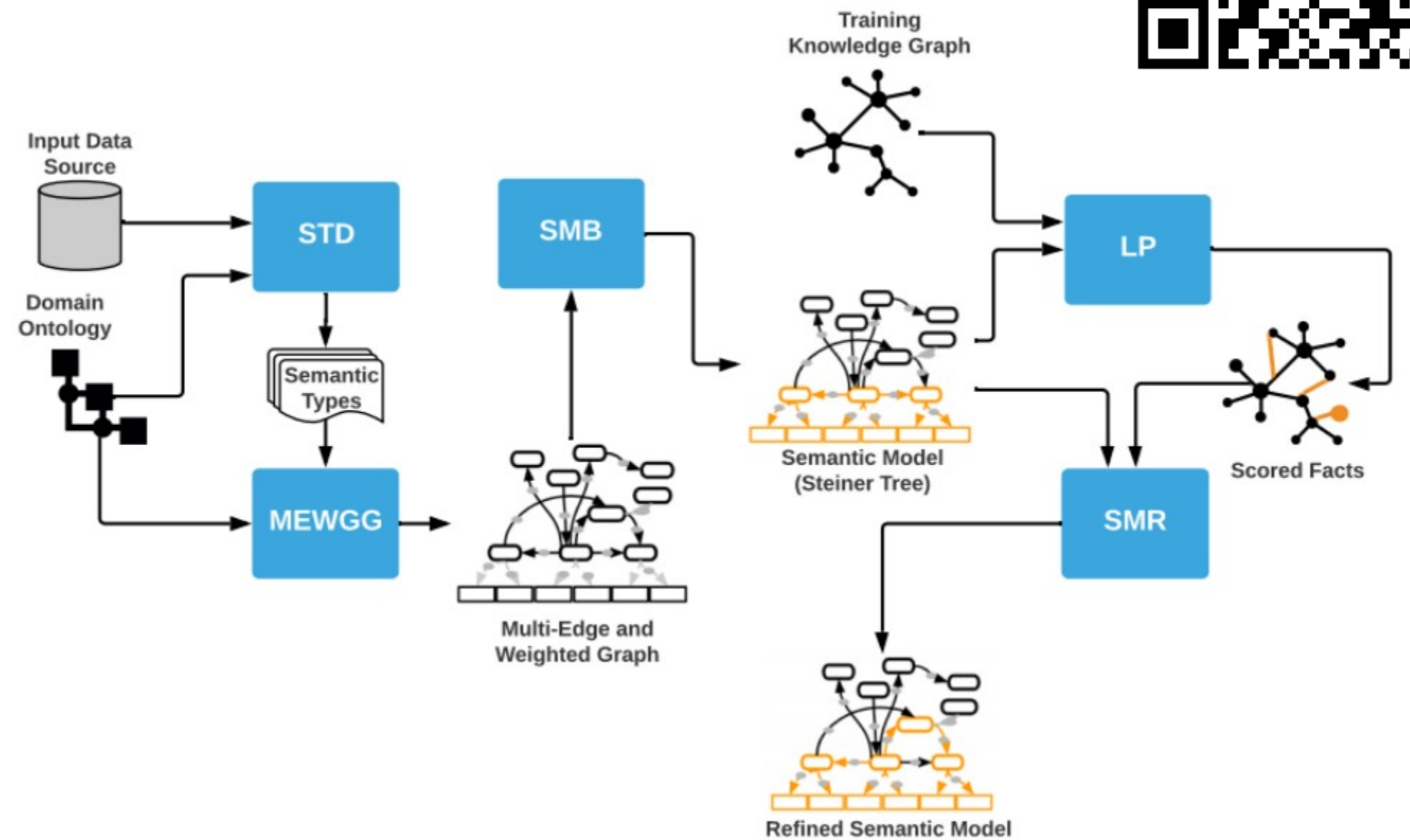
2. Goals

The main goal of the PhD research project is to design and develop a method for automatically building KGs. This method implies the construction of a map between the attributes of a data source and concepts and relations defined by a domain ontology: such map is called **semantic model**. Once the KG has been generated, a further step is to predict new relations between entities.

Geranium (geranium.nexacenter.org) is a platform to search for publications extracted from repositories of research institutions by **inferred semantic topics** and, in the future, to show **implicit connections** between researchers.

4. SeMi Evaluation

In the context of semantic modeling of relational DBs, results show that SeMi performs better than other tools where multiple JOINs among different tables are required to infer the correct semantic relation.

people	birth dates	birth places
Bob Dylan	1941-05-24	Duluth
Elvis Presley	1935-01-08	Tupelo
Little Richard	1932-12-05	Macon

- - -> :subclass property
- —> :object property
- - -> :data property

3. The SeMi Tool and the Geranium Platform

SeMi (**SE**mantic **M**odeling mach**I**ne) [1] is a tool for semantic modeling that extends a **steiner tree** detection algorithm with **deep learning** techniques on KGs. (Autoencoder --> **R-GCN** + **DistMult**).

References

1. Futia, G., Vetrò, A., & De Martin, J. C., SeMi: a SEmantic Modeling machIne to build Knowledge Graphs. Submitted to the SoftwareX journal [UNDER REVIEW]
2. Futia, G., Melandri, A., Vetrò, A., & De Martin, J. C., Training Neural Language Models with SPARQL queries for Semi-Automatic Semantic Mapping, In Proceedings of Semantics 2018, September 2018, Vienna (Austria)
3. **Mail**: giuseppe.futia@polito.it - **Twitter**: giuseppe_futia
4. **GitHub**: https://github.com/giuseppefutia