

# Master Internship

## On Enhancing Knowledge Graphs with Provenance Support

Khalid Belhajjame  
PSL, Université Paris Dauphine  
kbelhajj@gmail.com

Dario Colazzo  
PSL, Université Paris Dauphine  
dario.colazzo@lamsade.dauphine.fr

**Background.** Knowledge graphs can be viewed as large collections of interconnected entities enriched with semantic annotations. They have become powerful assets for enhancing search and are now widely used in both academia and industry. Well-known knowledge graphs include Google’s Knowledge Graph, Facebook’s Graph Search and Yago (see (Pan et al. 2017)). In the context of this internship, the focus will be on knowledge graphs that are available or can be exported as RDF datasets (Manola and Miller 2004) enhanced with RDF Schema (Guha and Brickley 2014) statements that capture relevant domain background knowledge.

**Objective of the Internship.** By and large, available knowledge graphs lack provenance support (Cheney, Chiticariu, and Tan 2009). Provenance information informs on the how-about of entity, i.e., how they come to be, and can be used in a range of applications, e.g., to explain the results of a query or search, to propagate annotations among the entities that constitute the knowledge graph, to learn attribution information, to name a few. The main objective of the internship will be to investigate how RDF knowledge graphs can be enhanced with provenance support. In doing so, the candidate will examine the different kinds of provenance information that can be collected and recorded and the computational complexity incurred by each, design an algorithmic solution, and implement it. For validation purposes, we will be using real knowledge graphs that are freely available under the aegis of the open linked data initiative.

**Work environment.** The internship will last for 4 to 6 months, starting April the 1st 2020, and will be supervised by Khalid Belhajjame (associate professor at the Paris-Dauphine University), and Dario Colazzo (full professor at the same university). The internship will take place at the same University.

### References

- Cheney, James, Laura Chiticariu, and Wang Chiew Tan. 2009. Provenance in databases: Why, how, and where. *Foundations and Trends in Databases*, 1(4):379–474.
- Guha, Ramanathan and Dan Brickley. 2014. RDF schema 1.1. W3C recommendation, W3C. [Http://www.w3.org/TR/2014/REC-rdf-schema-20140225/](http://www.w3.org/TR/2014/REC-rdf-schema-20140225/).
- Manola, Frank and Eric Miller. 2004. RDF primer. W3C recommendation, W3C. [Http://www.w3.org/TR/2004/REC-rdf-primer-20040210/](http://www.w3.org/TR/2004/REC-rdf-primer-20040210/).
- Pan, Jeff Z., Guido Vetere, José Manuel Gómez-Pérez, and Honghan Wu, editors. 2017. *Exploiting Linked Data and Knowledge Graphs in Large Organisations*. Springer.