# Knowledge Graph Exploration:
# Semantic Web-Based Exploratory Data Analysis

*Satria Bagus Wicaksono*

*You Xu*

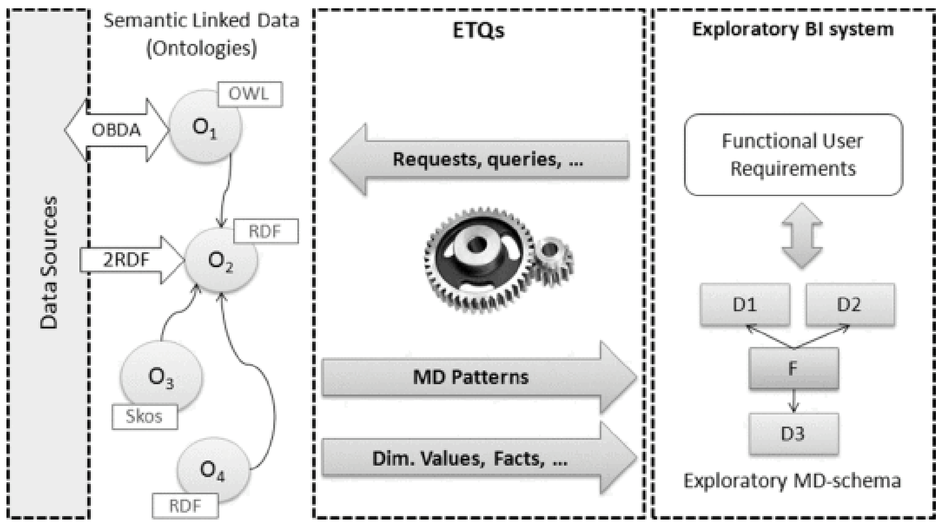
Universitat Politècnica de Catalunya

## I. Introduction

- Semantic data management tools such as Knowledge Graphs provide powerful solutions for data management and analysis.
- Challenges in knowledge graph: **heterogeneity and complexity.**
- Exploration aims to **understand, identify, and extract** insights from the graphs.

## II. Theories

- Semantic graphs
  - Schema-less
  - Flexible for modeling complex relationships.
- Online analytical processing (OLAP) with data warehouses are influential in decision-making requiring data aggregation.
- In Exploratory OLAP, ETQ (Query) replaces ETL




*Exploratory MD design for Semantic Web* [1]

- Approaches to achieve exploratory OLAP
  - Informational OLAP (I-OLAP): aggregate attributes
  - topological OLAP (T-OLAP): aggregate nodes/edges
- Essential components
  - Basic graph patterns (BGP)
  - Analytical Schema: schema for aggregation
  - Analytical Query: perform aggregation

## III. Application

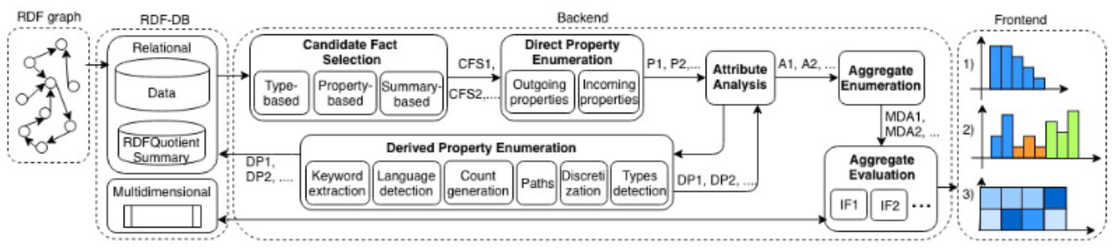
**Several tools** have been developed to explore and provide insight on the knowledge graph. It utilize pre-existing theory for knowledge graph exploration and add novel methods to analyze the graph. These tools also provide visualization of the analytics result.

- **Summarizing billions triples** [2], Summarizing triples using Usage summaries with *bisimulation label*.
- **Dagger** [3], a tool for generating top k most interesting aggregation over RDF graph space.
- **Spade** [4], Spade, an improvement of dagger by applying lattice based computations & novel early-stop technique.

## IV. Discussion

**Traditional OLAP** relies on the **Extract, Load, Transform** (**ETL**) process to convert data into a form suitable for multidimensional analysis, **Semantic web-based OLAP,** on the other hand, introduces the **Extract, Transform, Query** (**ETQ**) method. Both are uniquely suited to their respective data environments and come with their strengths and challenges. The decision to use one over the other primarily depends on the data's nature and the analysis process's specific needs.

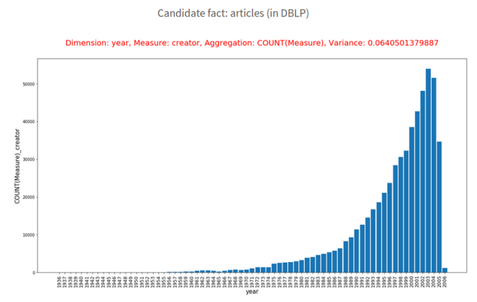
Both **Dagger** and **Spade**, utilized PostgreSQL for query aggregation and carefully translate RDF graphs to PostgreSQL to pertain its properties as a graph database [3], [4]. Additionally, both tools apply Analytical Query and perform Analytical Querying Answering strategies in aggregating the dimension by querying the graph components. Also, **Dagger** followed ETQ process in exploratory OLAP.




*Spade Architecture* [4]

## V. Conclusion

**The semantic web has potential** for articulating complex information but poses challenges for **redesigning OLAP** systems.




Dagger Aggregation Result [3]

**Novel approaches** like Dagger and Spade have been proposed, but the field is still early. **Robust tools and frameworks are needed** to navigate the semantic web's intricacies for advanced analytical tasks. Increasing demand for advanced analytics highlights the importance of research and development in this area.

## VI. Reference

[1] A. Abelló et al., "Using Semantic Web Technologies for Exploratory OLAP: A Survey," IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 2, pp. 571–588, Feb. 2015, doi: 10.1109/TKDE.2014.2330822.

[2] S. Khatchadourian and M. P. Consens, "Understanding billions of triples with usage summaries," Semantic Web Challenge, 2011.

[3] Y. Diao, I. Manolescu, and S. Shang, "Dagger: Digging for Interesting Aggregates in RDF Graphs," presented at the International Semantic Web Conference (ISWC), Oct. 2017.

[4] Y. Diao, P. Guzewicz, I. Manolescu, and M. Mazuran, "Spade: A Modular Framework for Analytical Exploration of RDF Graphs," Proceedings of the VLDB Endowment (PVLDB), vol. 12, no. 12, p. 1926, Aug. 2019, doi: 10.14778/3352063.3352101.