

# The Linked Data Benchmark Council (LDBC)

Irini Fundulaki<sup>1</sup>, Josep Larriba Pey<sup>2</sup>, David Dominguez-Sal<sup>2</sup>  
, Ioan Toma<sup>3</sup>, Dieter Fensel<sup>3</sup>, Barry Bishop<sup>4</sup>  
, Thomas Neumann<sup>5</sup>, Orri Erling<sup>6</sup>, Peter Neubauer<sup>7</sup>  
, Paul Groth<sup>8</sup>, Frank van Harmelen<sup>8</sup>, and Peter Boncz<sup>8</sup>

<sup>1</sup> Foundation for Research and Technology - Hellas

<sup>2</sup> Universitat Politcnica de Catalunya

<sup>3</sup> Universitaet Innsbruck

<sup>4</sup> Ontotext AD

<sup>5</sup> Technische Universitt Mnchen

<sup>6</sup> OpenLink Software

<sup>7</sup> Neo Technology

<sup>8</sup> Vrije Universiteit Amsterdam

## 1 Introduction

In the last years we have seen an explosion of massive amounts of graph shaped data coming from a variety of applications that are related to social networks (Facebook, Twitter, blogs and other on-line media) and telecommunication networks. Furthermore, the W3C Linking Open Data Initiative [8] has boosted the publication and interlinkage of a large number of datasets on the Semantic Web [2] resulting to the Linked Open Data Cloud. These datasets with billions of RDF triples such as Wikipedia[5], U.S. Census bureau [4], CIA World Factbook[1], DBPedia [5], and government sites<sup>1</sup> have been created and published online. Moreover, numerous datasets and vocabularies from e-science are published nowadays as RDF graphs most notably in life and earth sciences, astronomy [6][7][3] in order to facilitate community annotation and interlinkage of both scientific and scholarly data of interest. Technology and bandwidth now provide the opportunities for compiling, publishing and sharing massive Linked Data datasets. A significant number of commercial semantic repositories (i.e., RDF databases with reasoner and query-engine) which are the cornerstone of the Semantic Web exist.

Nevertheless at the present time, there is no:

- *comprehensive suite of benchmarks* that encourage the advancement of technology by providing both academia and industry with clear targets for performance and functionality.
- *independent authority* for developing benchmarks and verifying the results of those engines. The same holds for the emerging field of noSQL graph databases, which share with RDF a graph data model, pattern- and path-oriented query languages.

---

<sup>1</sup> <http://data.gov.uk/>, <http://www.data.gov/>

The **Linked Data Benchmark Council (LDBC)** project aims to provide a solution to this problem by making insightful the critical properties of graph and RDF data management technology, and stimulating progress through competition. This is timely and urgent since non-relational data management is emerging as a critical need for the new data economy based on large, distributed, heterogeneous, and complexly structured data sets. This new data management paradigm also provides an opportunity for research results to impact young innovative companies working on RDF and graph data management to start playing a significant role in this new data economy.

## 2 Objectives & Outcomes

The main technical objective of the Linked Data Benchmark Council (LDBC) is the development of benchmarks for different technology areas including *core data management* (query processing, query optimisation, transactions), *graph analysis*, *data integration* and *reasoning*. More specifically, LDBC aims at the:

- *development of new benchmarks* that will spur research and industry progress in large-scale graph and RDF data management. This includes setting challenges that will lead to significant progress in:
  - scalability, storage, indexing and query optimization techniques for RDF and graph database solutions beyond Terabyte scales.
  - quantitatively and qualitatively assess different solutions for data integration, and
  - computationally cheaper reasoning in RDF engines.
- establishment of an *industry-neutral entity*, the *LDBC foundation* for developing graph and RDF benchmarks, auditing benchmark results, and publishing audited results. The LDBC Foundation will work in the same spirit as the Transaction Processing Council (TPC) that has established a widely accepted by the industry, set of benchmarks for relational database management systems. It will be responsible for:
  - specifying benchmarks, benchmarking procedures and verifying/publishing results.
  - providing a TPC-style auditing service for certifying results published by vendors for benchmarks endorsed by LDBC.
  - training auditors for its benchmarking, creating a long-lasting business model for auditing benchmark results.

## 3 Target Audiences

The target audiences of LDBC that will comprise the core of the LDBC foundation as well as benefiting from and using the project results in the areas of technology, market and education are:

- *Technology Users*: This group includes both private and commercial users of RDF and graph databases that will use or integrate this technology for the benefits it has over traditional relational database management techniques.
- *Researchers*: This category includes a broad range of researchers from those who focus on graph-shaped data representations, query languages and optimisations, all the way to researchers from other fields who use this technology.
- *Technology Vendors*: This group is made up of commercial developers of RDF and graph database software components. It includes vendors who sell the software they produce as well as those who sell only services around their (open-source) products.

## References

1. Cia world factbook. <https://www.cia.gov/library/publications/the-world-factbook/index.html>.
2. T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. SCIENTIFIC AMERICAN-AMERICAN EDITION, 2001.
3. The UMD Astronomy Information and Knowledge Group. Astronomy Ontology in OWL. [archive.astro.umd.edu](http://archive.astro.umd.edu).
4. The 2000 US Census. [www.rdfabout.com/demo/census](http://www.rdfabout.com/demo/census).
5. DBPedia. [www.dbpedia.org](http://www.dbpedia.org).
6. Gene Ontology. [www.geneontology.org](http://www.geneontology.org).
7. UniProtRDF. [dev.isb-sib.ch/projects/uniprot-rdf](http://dev.isb-sib.ch/projects/uniprot-rdf).
8. W3C Linking Open Data. [esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData](http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData).