

1. Publishable summary

1.1. Summary description of project context and objectives

LDBC (<http://www.ldbc.eu/>) is a Collaborative Project (STREP) from the FP7-ICT-2011-8, funded and started in September 30, 2012 lasting for 30 months. It stands for “Linked Data Benchmark Council” and aims at creating the first comprehensive suite of open, fair and vendor-neutral benchmarks for RDF/graph databases together with the LDBC foundation which will define processes for obtaining, auditing and publishing results.

It addresses then the FP7 objective *ICT-2011.4.4 Intelligent Information Management* , outcome c), which calls for “Framework and tools for **benchmarking** and exploring information management diversity and comparing and optimising the **performance** of **non-mainstream data management** architectures and computing paradigms, novel data structures and algorithms on **extremely large volumes of data**. While methodological rigour and scientific quality and novelty are the main criteria for success, preference will be given to proposals that address a **clearly identified industrial, scientific or societal concern or opportunity** and/or **bring together hitherto unrelated scientific or software engineering communities.**” The initial consortium is composed by 8 preeminent institutions (academia, foundations and SMEs) from 8 EU different countries.

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 new RDF and graph data management technologies to start playing a significant role in this new data economy.

Standards and benchmarking are two of the most important factors for the development of new information technology, yet there is still no comprehensive suite of benchmarks and benchmarking practices for RDF and graph databases, nor is there an authority for setting benchmark definitions and auditing official results. Without them, the future development and uptake of these technologies is at risk by not providing industry with clear, user-driven targets for performance and functionality.

Providing a solution to this problem is the major goal of the project that seeks to make insightful the critical properties of graph and RDF data management technology, and stimulate progress through competition.

The LDBC project will bring together a community of top researchers and RDF and graph database vendors to develop and implement the first suite of open, fair and vendor-neutral benchmarks that will include workloads, data generators and test drivers for measuring the performance of RDF and graph databases. Good benchmarks contain relevant technical challenges that can only be overcome by innovation, i.e. they steer progress in research and industry. LDBC involves top research groups to identify these technical challenges (“choke points”) and continuously engage a user community that contributes scenarios, data and query workloads. LDBC plans separate work packages for different technology areas, including core data management (query processing, query optimisation, transactions), graph analysis, and data integration and reasoning; each of which is led by a scientific expert in this particular domain.