

LDBC

Collaborative Project

FP7 – 317548

D5.5.1 Dissemination Report Y1

Coordinator: Peter Boncz (VUA)

With contributions from: Irimi Fundulaki (FORTH), Miquel Ferrer (UPC), Ioan Toma (UIBK)

1st Quality reviewer: Josep Larriba Pey, UPC

2nd Quality reviewer: Hugh Williams, OGL

Deliverable nature:	R
Dissemination level: (Confidentiality)	CO
Contractual delivery date:	M12
Actual delivery date:	M12
Version:	1.0
Total number of pages:	25
Keywords:	

Executive summary

This report gives an overview of the dissemination activities in the first year of the LDBC project.

These activities follow four tracks:

1. Building the vendor community around LDBC – the group of potential members of the LDBC non-profit organization. This activity involved in Y1 a series of concalls, calls and email exchanges between database architects and performance engineers and product managers of various companies in the Graph or RDF data management space and LDBC members.
2. Engaging technology users to provide feedback and input via the Technical User Community (TUC). This activity involved the building of said TUC, as well as organizing two TUC meetings (November 2012 Barcelona, April 2013 Munich).
3. Engaging the academic community to both contribute technical expertise in LDBC benchmark development, and raising awareness for LDBC and its benchmarks with the goal of making these benchmarks the instrument of measuring scientific progress.
4. Engaging the larger IT public in LDBC and its benchmarks. This is addressed using a public relations strategy. Further, a questionnaire on graph and RDF benchmarking needs was organized right at the start of the project, whose results can be found in this report.

Document Information

IST Project Number	FP7 - 317548	Acronym	LDBC
Full Title	LDBC		
Project URL	http://www.ldbc.eu/		
Document URL			
EU Project Officer	Carola Carstens		

Deliverable	Number	D5.5.1	Title	Dissemination Report V1
Work Package	Number	WP5	Title	Dissemination

Date of Delivery	Contractual	M12	Actual	M12
Status	version 1.0		final x	
Nature	prototype <input type="checkbox"/> report <input type="checkbox"/> dissemination <input checked="" type="checkbox"/>			
Dissemination level	public <input type="checkbox"/> consortium <input checked="" type="checkbox"/>			

Authors (Partner)				
Responsible Author	Name	Peter Boncz	E-mail	p.a.boncz@vu.nl
	Partner	VUA	Phone	

Abstract (for dissemination)	This deliverable reports on the dissemination activities of the LDBC project. These activities cover (i) engagement of a community of industry layers that might become later LDBC members, (ii) engagement of RDF and graph data management practitioners for providing input for LDBC benchmarks as well as feedback on drafts, (iii) engagement of the academic community to help out with benchmark design and also to promote LDBC benchmarks as instruments for measuring scientific progress, and (iv) engaging the larger IT public with LDBC such that its work is appreciated and used.
Keywords	

Version Log			
Issue Date	Rev. No.	Author	Change
22.09.2013	0.1	Peter Boncz	Initial draft
23.09.2013	1.0	Peter Boncz	Final version after reviewer comments

Table of Contents

Executive summary	2
Document Information	3
Table of Contents	4
Abbreviations	5
1 Introduction	6
2 Vendor Community	8
2.1 Membership Audience	8
2.2 Membership Strategy	9
2.3 Potential Member Contacts	10
2.4 Vendor Strategy for LDBC Y2	11
3 Technical User Community	12
3.1 First TUC Meeting (Barcelona)	13
3.2 Second TUC Meeting (Munich)	14
3.3 Future TUC strategy	15
4 Scientific Meetings	16
4.1 Graph-TA Workshop	16
4.2 BerSys Workshop	17
4.3 GRADES Workshop	18
5 End User Dissemination	20
5.1 Public Relations	20
5.2 Early Questionnaire	20
6 Conclusion	25

Abbreviations

LDBC	Linked Data Benchmark Council
TUC	the Technical User Community of the LDBC
BerSys	Workshop on Benchmarking RDF data management Systems
GRADES	Workshop on Graph Data Management, Experiences and Systems
GRAPH-TA	Workshop on Graph-based Technologies and Applications
VUA	Vrije Universiteit Amsterdam
UPC	Universidad Politècnica de Catalunya
UIBK	University of Innsbrück
FORTH	Greek National Research Institute
TUM	Technical University Munich
OGL	OpenLink Group Ltd, which creates and sells among others the Virtuoso RDF Quad Store
ONTO	Ontotext, which creates and sells among others the OWLIM RDF engine
NEO	Neo Technologies, which provides commercial support for neo4j
SPA	Sparsity Technologies

1 Introduction

In this deliverable, we describe the dissemination activities of LDBC that go beyond those activities that are a normal component of every EU project. In particular, the core mission of LDBC depends on social activities where the LDBC:

- (i) seeks input from IT users, software and hardware vendors, academia and technical experts to define new benchmarking ideas,
- (ii) approaches those same parties to provide feedback on draft LDBC specifications, when available
- (iii) seeks participation in the non-profit LDBC legal entity from a strong coalition of industry partners and
- (iv) creates support for LDBC benchmark specifications among the public (i.e. IT practitioners and academia).

User input/feedback, vendor participation and benchmark acceptance are crucial to the success of the LDBC.

In this deliverable, we describe those specific actions LDBC undertakes to be successful in its mission:

- we describe the activities of the Technical User Community (TUC) of the LDBC.
- we describe the efforts on engaging the vendor community.
- we also provide details on the academic outreach, which aims at helping build and helping achieve acceptance of LDBC benchmarks.
- we finally provide detail on our end-user engagement activities.

We note that this deliverable *does not report on* dissemination activities such as **scientific publications** by LDBC members, **speaking engagements** and **conference and workshops visits**. These “usual” dissemination activities up until M12, are covered by the management deliverable D7.7.3.

An important aspect, if not the core, of current and all future research and development activities are the activities undertaken by **Benchmark Task Forces**. A benchmark task force conducts all tasks needed to develop a benchmark specification. This entails

- gathering user requirements,
- gathering and analyzing datasets and workloads,
- identifying so-called “choke points” by involving data management systems experts,
- designing benchmark datasets, workloads and metrics,
- creating data and workload generators and packaging these,
- documenting the benchmark and its rules (creating the specification),
- producing initial results and engaging with TUC members for feedback.

In this EU project, the task force participation of the members provides a “matrix organization” that overlays orthogonally on the work packages. That is, activities of task forces during work on the development of future LDBC benchmarks, unite the activities of the various project participants in different technical areas (work packages). Also after the end of the EU project, Benchmark Task Forces are expected to be the main tool of the non-profit LDBC organization to develop new specifications.

After consultation at the first TUC meeting, two Benchmark Task Forces were formed, and thus play a role in the M12 reports:

- the Social Network Benchmark Task Force
- the Semantic Publishing Benchmark Task Force

The specific activities of the Benchmark Task Forces, while playing an important role as the subject matter of discussion in TUC meetings, as well as in academic and industrial exchanges with LDBC members with

the outside world, are also **not described** in this particular document. Rather, each Benchmark Task Force has provided a specific M12 report on its activities, describing the rationale, benchmark specification, and preliminary results of the two benchmarks under development. These reports have been released in November 2013 as public documents of the LDBC non-profit organization – reading is recommended.

This deliverable is organized along the main tasks in WP5:

- Task 5.1.1 Industry Consortium Formation (Start M1, End M30 / Lead VUA:5). Apart from the activities in establishing a non-profit organization in the form of a UK company of limited guarantee (the topic of D6.6.1) the activities have taken the form of scientific director and project coordinator contacts with industry players, briefing them on LDBC, investigating their motives with the general strategy of building an LDBC non-profit foundation with broad membership. This is discussed in Section 2 of this deliverable.
- Task 5.1.2 and Technical User Community RDF (Start M4, End M30 / Lead OGL:4, Participants ONTO:2) and Task 5.1.3 Technical User Community Graphs (Start M4, End M30 / Lead NEO:4, Participants UPC:2). At the start of the project, we have decided not to split the users into graph and RDF groups. There are benchmark demands, and in particular the social network benchmark, where both communities have a stake. This is discussed in Section 3 of this deliverable.
- Task 5.1.4 Scientific Outreach (Start M4, End M30 / Lead VUA:3, Participants TUM:2,FORTH:2). The main activity of this task force has been organizing workshops: Graph-TA in February 2013 (Barcelona), BerSys in May at ESWC 2013 (Montpellier) and GRADES in June at SIGMOD 2013 (New York). We have also published project overview papers about LDBC to the Datenbanken and SIGMOD Record journals, and described the project in an IDEAS 2013 keynote. This is discussed in Section 4 of this deliverable.
- Task 5.1.5 End User Community Outreach (Start M1, End M30 / Lead IUBK:4) Further, dissemination activities (Events and Campaigns) will be designed with the general purpose to raise awareness among the technology end-user public for the LDBC non-profit organization. This both involves the designing of a PR Strategy as well as an End User Engagement Strategy to make end users interact with the LDBC and its benchmarking products. It was decided to lay more emphasis on this in Y2 than in Y1 of the project, because the LDBC did not have draft benchmark specifications available until M12. Therefore, the discussion of this in Section 4 of this deliverable is relatively limited.

We note that the final Dissemination task of WP5, Task 5.2 Dissemination Portal (Start M1, End M30 / Lead UIBK:6, Participants VUA:3,ONTO:2,NEO:2,UPC:3,OGL:2) is described in the separate deliverables D5.5.3 and D5.5.4 rather than here.

2 Vendor Community

The goal of this project is to establish the LDBC non-profit organization and make it a successful and lasting organisation, supported by the RDF and graph database industry. For this purpose, LDBC has chosen to incorporate a “UK company limited by guarantee”, which is the established form in the UK for non-profit organizations. The reason to do so in the UK is that the Articles and Bylaws, which govern LDBC membership, governance and conflict resolution must be understandable to an international audience, hence these are best phrased directly in English.

In contrast to the EU project LDBC, the LDBC non-profit organization is meant to last. Also, its membership should be broader than that of the EU project. In order to foster an LDBC non-profit organisation that gains broad participation from the RDF and graph database industry, it is important to create LDBC as a vendor-neutral organisation, where we must ensure that LDBC does not favour the industrial partners of the consortium. The scientific director (VUA) and project coordinator (UPC) have undertaken actions in the first year of LDBC to involve industry players outside of the consortium to become members, when the non-profit organization gets created.

2.1 Membership Audience

The industry member audience for the LDBC can be divided into three main classes:

- RDF store software vendors (products),
 - Oracle (Oracle RDF, Oracle noSQL)
 - IBM (DB2 RDF)
 - OpenLink Software (Virtuoso)
 - Ontotext (OWLIM)
 - Franz Inc (AllegroGraph)
 - Garlik (4Store/5Store– now part of Experian)
 - Systap LLC (BigData)
 - Clark&Parsia (Stardog)
 - Dydra (Dydra)
 - SparqlCity (SparqlBase)
 - MarkLogic (MarkLogic7)
 - [open-source: Apache Foundation (Jena/TDB), OpenRDF (Sesame)]

Note: in 2013 StrixDB disappeared, IBM, SparqlCity and MarkLogic are new entrants.

- graph database vendors (products)
 - Neo Technology (Neo4j),
 - Objectivity(InfiniteGraph),
 - Sparsity Technologies (DEX),
 - Kobrix Software (HyperGraphDB),
 - Orient Technologies (OrientDB),
 - YarcData (Urika),
 - Aurelius (Titan/Faunus),
 - [open-source: Apache Foundation (Hama, Giraph)]

Note: new prototypes in the distributed graph programming area are appearing regularly, e.g. Signal/Collect (University Zurich), Spark (Berkeley), Grappa (University Washington) and Green Marl (Oracle Labs) and also graph-enabling extensions in Stratosphere (Humboldt University Berlin). Some of these prototypes might mature into widely used open source systems or even start-ups. They can play a role in scientific uptake of LDBC benchmarks.

- and some of the longer-existing players in the object-oriented database space (who might be interested in graph data management benchmarking):
 - Actian (Versant, USA)
 - InterSystems (Caché, USA)

Other Parties. However, the above only focuses on software vendors. The ecosystem around the LDBC non-profit organization further contains the following actors:

- academic institutions, or individual academics
- independent consultants who made graph or RDF data management systems their field of expertise
- hardware vendors.

In the case of the TPC organization (www.tpc.org), which has been successful in relational database systems benchmarking in the past two decades, a large part of the membership comes from the hardware community. Recognizing this, the LDBC consortium has tried to extend the project under the respective 2013 EU program, proposing to extend the consortium with HP Italy, IBM Labs in Haifa and University of Cyprus. This EU proposal failed, but the contacts in the hardware community will still be developed further over 2014. Nevertheless, in the remainder we focus on the software industry member audience.

2.2 Membership Strategy

LDBC is to incorporate the company at M12, and expects the EU project members to join the company by December 2013. At a first stage, LDBC will be incorporated with two members, i.e. Peter Boncz and SPA, the new partner of LDBC from July 1st 2013¹. The directors of the company at this stage will be Peter Boncz and Josep L. Larriba-Pey representing SPA. From there on, the expansion should continue and take shape over 2014. The communication strategy toward the relationship with these potential members must, in the interest of LDBC, be to convince a large part of them to join LDBC quickly.

We consider the following requirements for LDBC to make this possible:

- The benchmark specifications and publications of LDBC should be of high technical quality (relevant, transparent, portable, scalable, economic and understandable) and of high software quality.
- The legal structure of the LDBC non-profit organization should be transparent and resistant against unfair treatment and potential conflicts of interest.
- The annual membership cost of LDBC both in the monetary and time sense should be reasonable.

To address these, the LDBC project has established contacts with world-recognized technical experts, outside the centers of its own competence (e.g. Carlos Guestrin of the GraphLab community and Luis Ceze).

To obtain high-quality benchmarks, the LDBC has developed a new “choke point” driven benchmark development strategy that, apart from user input and feedback (through the TUC) relies on access to technical expert in database architecture, to ensure that the benchmarks contain relevant technical challenges.

To ensure a good legal structure, LDBC has specifically drawn on the expertise of experienced TPC members through interviews and advisor roles (e.g. Berni Schiefer, Karl Huppler). The UK style Articles of the LDBC organization are also supplemented by an explanatory memorandum that explains the main concepts in terms of US company structures, to ensure that the LDBC is understandable for a US company.

During 2014, we will make membership of LDBC free, to take that potential obstacle to joining off the table, and finance the (limited) legal and financial expenses of the LDBC non-profit organization through a gift from the EU LDBC project (the remainder of the EUR 15K legal fund for incorporation).

Phrased positively, reasons to join the LDBC could be the following:

- A company wants to influence the benchmark adoption policy and becomes a member to obtain a vote. This could take the form of a company pushing for a new benchmark, or pushing to change an existing benchmark, but also take the negative form of pursuing the intention to block a benchmark. The (perceived) notion that the benchmark could show their own products in a positive light when compared to competitors. It could also be the conviction that streamlining development by using benchmarks would improve the quality of future products.
- A company may simply want to show that it belongs to the Graph or RDF market segment, by joining the LDBC.

¹ Sparsity is Third Party type B of LDBC since July 1st 2013.

The third motivation is one that will take on strength once the LDBC has a critical mass. The first two motivations underline that for LDBC, it is important start benchmark adoption of a few benchmarks by having these reach draft stage. This will both lead to certain products wanting to advertise their score as well as potentially lead to outside parties acting on their desire to change these benchmarks.

2.3 Potential Member Contacts

Until now the scientific director and project coordinator have kept contacts with a number of graph and RDF software companies, which comprise the great majority of those in these (small) markets. A first group was formed when writing the LDBC EU proposal, start of 2012, when several of these supplied support quotes. During the first year, we have used conference calls and individual telephone calls and email exchanges to maintain and extend contacts. The table below gives an overview of the current state.

Organisation	Name	Role & Comments
Oracle	Xavier Lopez	Runs the Oracle RDF products. Spoke at TUC2; supportive. Poised to come on the LDBC Advisory Board.
Oracle	Souripriya Das	Architect in the Oracle RDF products.
Oracle	Timothy Harris	Works with Sedlar and Ceze
Oracle	Eric Sedlar	Head of Oracle Labs. Runs the Green Marl program. Supportive of LDBC. At GRADES2013.
UWashington	Luis Ceze	Oracle commissioned him for a graph programming framework benchmark. (parts of) this workload could become the SNB graph analytics workload. Has accepted to be on the LDBC advisory board.
Objectivity	Patricia Stamos	
Objectivity	Darren Wood	Provided LDBC quote.
Objectivity	Leon Guzenda	
VelocityDB	Mats Persson	
IBM	Berni Schiefer	IBM benchmark expert; now runs BLU development. Was briefed on LDBC and conducted a 1.5h interview in July.
IBM	Mario Ds Briggs	On vendor call.
IBM	Jason Crawford	IBM Lab researcher interested in graph programming frameworks.
IBM	Kavita Srinivas	Architect in the IBM DB2 RDF product. On Vendor call.
IBM	Karl Huppler	Just retired. >10years president of TPC. Approached at TPCTC2013.
Bigdata	Bryan Thompson	Architect/founder of BIGDATA. Provided quote for LDBC. On vendor call and at GRADES2013.
Clark & Parsia	Kendall Clark	Architect/founder of Clark&Parsia. Provided LDC quote. Wants strong emphasis on reasoning and has since been critical on LDBC.
Franz	Jans Aasman	Architect/founder of Allegrograph. Provided LDBC quote.
Versant	Steve Shine	CEO of Actian, who bought Versant in 2013. Is supportive of LDBC. Conversation in Amsterdam May 2013.
Versant	Volker John	Runs Versant development in Hamburg.
SparqlBase	Barry Zane	Ex architect/founder of Netezza and Paracel. We conducted two concalls while spinning off SparqlCity.
Dydra	James Anderson	Architect/founder of Dydra, a cloud-based sparql solution. Critical – would have preferred to be EU project partner.
Garlik	Steve Harris	Founder of Garlik, has since left the company for a new startup. Provided LDBC quote.
Garlik	Siva Nagappan	Technical architect on 5store. On vendor call.
Garlik	Dave Challis	Technical architect on 5store. On vendor call.
YarcData	Alyssa Jarrett	Email contacts.
OrientDB	Luca Garulli	Architect/founder of OrientDB. Provided LDBC quote. OrientDB has been silent in the GraphDB market in 2013.

2.4 Vendor Strategy for LDBC Y2

In 2014, the LDBC Company will exist and the membership expansion outside the project should take shape. A number of elements in these are:

- Appointment of Advisory Board members. Advisory Board membership of the EU LDBC project is a lightweight membership that company individuals can enter into relatively easily. Thus, advisory board membership can be used as a stepping stone towards LDBC membership, in addition to the primary purpose of advising LDBC on strategic matters.
- Organization of two TUC meetings, is being used not only to seek input and feedback from IT practitioners on benchmark development, but also as a venue to invite and meet prospective LDBC members. There will be a third TUC meeting co-located with Neo Technologies' GraphConnect in London (UK) on November 19th. A fourth TUC meeting will follow in April 2014.
- Open source release of benchmark kits for the Social Network Benchmark (SNB) and Semantic Publishing Benchmark (SPB). The LDBC will pursue a campaign of bringing these kits, and the task force reports that describe their design rationale, to the attention of outside partners that may be prospective members.
- Publicity campaign on the creation of the LDBC non-profit organization. A kick-off for this will be held during the GraphConnect event. Further press releases, and also attention at e.g. keynote speeches in conferences (IDEAS 2013, Barcelona + EDBT 2014 Athens).
- Continuation of the soft campaign between scientific director and project coordinator towards outside parties, now concretely suggesting outside partners to come on board. An additional argument will be an offer to participate in the writing of the Bylaws, which describe the rules that govern benchmark submissions that will be completed in the first half of 2014.

3 Technical User Community

The Technical User Community (TUC) is the primary way for the LDBC to obtain user input (workloads, datasets) and feedback (on benchmark proposals and results). The TUC gathers in TUC meetings, and two of those were organized in Y1 of LDBC. TUC meetings are also a good place to invite people from outside to take a look at what LDBC is doing. A specific case is companies that are potential LDBC non-profit organization members.

Until now, the TUC organization has been arranged at LDBC project partners in-house, such that meeting organization cost have been zero, except for travel expenses of the LDBC participants. For organizing the TUC event, there is significant budget. Part of the budget is for travel support for invited participants. This is used on limited basis, offering EUR 700 maximum travel reimbursement, for TUC participants where this helps arranging their presence. A detailed total of the amount of money spent on this in Y1 can be found in D7.7.3.

Further, the LDBC has a web-portal where TUC members can share information related to benchmarks, workloads and datasets. It is typically through the involvement of selected TUC members in LDBC benchmark development task forces that continuous cooperation is assured. This means that TUC members are drawn into the benchmark development task forces, or regularly consulted by these.

3.1 First TUC Meeting (Barcelona)

The first TUC meeting was organized in relatively little time after the start of the LDBC EU project on September 30, 2012. Still, the various industry partners in LDBC (“host” column) were able to draw from their customer bases a number of interested organizations, who presented their use cases in the two day event in Barcelona on the 19/20th November 2012. This willingness to even travel in at short notice from the USA in various cases (Elsevier, RJ Lee Group, Yale) shows the interest in LDBC among graph and RDF technology users.

Organisation	Person(s)	Host	Presentation
Connected Discovery	Bryn Williams-Jones	VUA	The OpenPHACTS project
BBC	Jem Rayfield	ONTO	BBC Dynamic Semantic Publishing
Yale University	Lec Maj	ONTO	Linked Data For researchers, YCBA Pilot Project
Press Association	Jarred McGinnis	ONTO	Semantic Publishing at the Press Association
R.J. Lee Group	David Neuer	ONTO	RJ Lee Group, Themis, and the Semantic Web
Elsevier	Alan Yagoda	OGL	Elsevier Health Sciences, Smart Content Drives Smart Applications- The Future Of Using Knowledge In Healthcare
ACCESO Group	Carlos Gutierrez	UPC	-
Media Planning Group	Santi Murillo	UPC	-
CA Technologies	Victor Muntés	UPC	Graph Database Use Cases at CA
Bio4J	Eduardo Pareja	NEO	Next Generation Sequencing and bio4j
Innoquant	Oleg Morajko	UPC	Innoquant DEX and neo4j Graph Use Cases

The slides and presentation can be found on the public TUC portal:

<http://www.ldbc.eu:8090/display/TUC/First+TUC+meeting+Nov+2012>

The goals of the first TUC meetings were to:

- Obtain customer workloads and datasets
- Obtain ideas and feedback on the first Benchmark Task Forces to be formed.

The important outcomes from this meeting where:

- Healthy indications of broad demand for better data management benchmarks.
- Two Spanish marketing companies (Havas, Accesso) as well as CA technologies provided datasets and workloads to UPC.
- Both BBC and Press association provided workloads and datasets to ONTO
- Two Benchmark Task Forces were formed: the Social Network Benchmark (SNB) task force and the Semantic Publishing Benchmark (SPB) task force.

** In this first TUC meeting there was also participation from Paola di Maio, who argued for driving the project from a technical towards a sociological direction. After this idea was not embraced by the LDBC members present, she opened a discussion on the W3C semantic web mailing list, attacking the project.

3.2 Second TUC Meeting (Munich)

The second TUC meeting was held in April 2013, ad joint to the M6 plenary meeting of the LDBC project. By that time, the two benchmark development task forces had been operating already for five months, such that in the TUC meeting there were also status presentation of these given by the two task force leaders.

The TUC members who visited the meeting are given in the below table, as well as the partner (Host) who invited them and the title of their presentation.

Organisation	Person	Host	Title
Wolters Kluwer	Edward Thomas	VUA	Semantic Publishing at Wolters Kluwer
Mediapro	Gustavo González	UPC	Graph-based User Modeling through Realtime social streams
BBC	Dave Rogers	ONTO	Linked Data Platform at BBC
R.J. Lee Group	David Neuer	ONTO	-
MPI	Fabian Suchanek	TUM	YAGO: A large knowledge base from Wikipedia and WordNet
VUA	Antonis Loizou	VUA	The OpenPHACTS approach to data integration
Oracle	Xavier Lopez		Graph Database Performance - an Oracle Perspective
University of Cyprus	Pedro Trancoso	UPC	Benchmarking and computer architecture: the research side
Dshini	Klaus Großmann	NEO	Neo4j at www.dshini.net
Brox	Mirko Kämpf	OGL	GeoKnow - Spatial Data Web project and Supply Chain Use Case
ACCESO	Jesús Lanchas	UPC	Use case of ACCESO in Social Network Analytics
Actify	Andrew Sherlock	UPC	Visual Insights for Manufacturing

The slides of these talks can be found on:

<http://www.ldbc.eu:8090/display/TUC/Second+TUC+meeting%2C+April+2013>

There was good feedback on both benchmarks under development. For the Social Network Benchmark, this came from from Dshini and Mediapro. For the Semantic Publishing Benchmark, from BBC and Wolters Kluwer.

Oracle was also present with a presentation that covered performance experiments both in the Oracle RDF support, including new reasoning support , as well as Oracle NoSQL. This contact led to closer communication between Oracle and LDBC, that will likely result in greater involvement of Oracle in the project.

The OpenPHACTS presentation at this second TUC was much more detailed than the one at the first TUC meeting (which explained the goals and methods of OpenPHACTS). In this talk, Antonis Loizou provided detailed examples of the methodology and techniques used for data integration. These methods and datasets will be inspiration for the data integration and ETL benchmarking work in Y2 of LDBC.

3.3 Future TUC strategy

In Y2 of LDBC two more TUC meetings will be organized, one on November 20 in London, coinciding with Neo Technologies' GraphConnect event, and the other ad joint with the 18M plenary (probably in Amsterdam, April 2014).

At GraphConnect, there will be (partial) drafts of the Social Network Benchmark and Semantic Publishing Benchmark available, such that obtaining concrete feedback on these will be one of the focal points. The current state of these benchmarks will be critically examined to see if they improve the state-of-the-art. This could possibly lead to a second round, e.g. the SPB possibly being extended in as diverse directions as faceted search, integration of keyword search, advanced reasoning and/or geographical querying. Similar for SNB, though by November this will only have reached partial draft state, and further development towards a final draft will lead into 2014 anyway.

In case of the Social Network Benchmark, the consensus is to have three separate workloads: interactive (OLTP), analytical (OLAP) and graph analytics. The SNB task force currently lacks coverage in the area of graph programming frameworks, such that involving other groups involved in this area is a second goal.

A third goal for the next series of TUC meetings is broadening, and identify 1-2 new benchmark areas and start task forces for these. Possible new directions that have been suggested are (i) cultural heritage in RDF, (ii) data integration, (iii) advanced reasoning.

Finally, in this meeting we may also touch on the issue of the policies and rules for the LDBC non-profit organization, whose creation will also be announced in London. This will also lead to the situation that non EU project partners can become members. As such the TUC events will also be used to draw in such partners.

4 Scientific Meetings

The scientific dissemination activities have as goal to establish the LDBC as the authority for graph and RDF benchmarks and make its benchmarks a well-known and recognized instrument for measuring scientific progress. It also has as goal to draw into the project talented academics who can materially contribute to the benchmark design and evaluation.

One aspect is doing so using classical dissemination activities such as giving talks (or keynote talks) at well-known conferences, as well as publishing scientific papers in which LDBC is described, or its benchmarks are described or used. For the overview of such activities, we refer however to D7.7.3.

We focus here on the activities related to scientific community building, which is addressed mainly by organizing workshops, both in the area of Graph data management and RDF data management.

4.1 Graph-TA Workshop

Graph-TA was an informal workshop celebrated on Feb 19th, 2013. The Graph-TA workshop was organized by the DAMA-UPC (Data Management) group of Universitat Politècnica de Catalunya – BarcelonaTech and it aimed at encouraging researchers working with graphs and graph-based technologies, to present their current work and research lines in order to foster collaboration among them.

The workshop hosted more than 40 researchers from Spain, Switzerland, Chile, and The Netherlands, and a total of 18 talks in three sessions were given. Each session allowed 10 minutes per talk plus 1 hour of discussion in a poster session where all the researchers were able to interact. Among the topics covered in the workshop, a prominent one was graph benchmarking, with three papers, and LDBC was presented to the audience by the LDBC coordinator.

Author(s)	Title
Horst Bunke	Towards a Quadratic Time Approximation of Graph Edit Distance
Leen Lambers	Automated Reasoning for Graph Properties and its Applications to Graph Database Design and Analysis
Josep Ginebra	A Cluster Analysis of Vote Transitions
Marta Pérez	Marshall-Olkin Extended Zipfian Distribution
Susana Ladra	k2-tree for Web graph representation
Josep Lluís Larriba-Pey	Linked Data Benchmark Council
Romualdo Pastor-Satorras	Modeling face-to face social interaction networks: The role of human dynamics
Francesc Comellas	Reconstruction of Complex Networks
Xavier Martínez	Dynamic Data Partitioning for Distributed Graph Databases
Ricard Gavaldà	Mining Frequent Closed Graphs on Evolving Data Streams
Renzo Angles	Benchmarking essential graph queries
Anjan Dutta	A symbol spotting approach in graphical documents by hashing serialized graphs
Nieves Brisaboa	Variants of k2-tree for RDF graphs
Josep Lladós	Improving Fuzzy Multilevel Graph Embedding through Feature Selection Technique
David Domínguez	Shaping communities out of triangles
Joan Vilatella	A Simple Heuristic Algorithm for Fast Edge-coloring of Large Graphs
David Francis Nettleton	Some Issues related to the Mining of OSNs represented as Graphs
Francesc Serratos	Active & Interactive Graph Matching

The slides of these presentations can be found on:

. <http://www.dama.upc.edu/seminars/1st-graph-ta>

4.2 BerSys Workshop

The BeRSys 2013 workshop, held in May 2013 at ESWC in Montpellier, is a forum where topics related to the evaluation (included, but not limited to, expressive power, usability and performance) of RDF data management platforms can be discussed and elaborated. The objectives of this new workshop are to: (i) create a discussion forum where researchers and industrials can meet and discuss topics related to the performance of RDF systems, and (ii) expose and initiate discussions on best practices, different application needs and scenarios related to RDF data management. BerSys 2013 was organized by Iriini Fundulaki (ONTO), Ioan Toma (IUBK) and Ioana Manulescu (INRIA, Paris). The workshop drew less papers than expected, and had to be merged with the USEWOD workshop.

During this first edition of BeRSys, the keynote talk was given by Prof. **Manolis Koubarakis on “Benchmarking Geospatial RDF Stores”** who presented a functional and performance benchmark used to compare the functionality and performance of most of the geospatial RDF stores that are available today.

Further presentations focussed on key issues in creating benchmarks for different data types illustrated by different application types, and some featured results of benchmarking different RDF data management systems. **Emanuele Della Valle** pointed out the differences in existing systems’ operational semantics when processing data streams and proposed remedies in *“On the need to include functional testing in RDF stream engine benchmarks”* (co-authored with Daniele Dell’Aglia and Marco Balduini).

Martin Przyjaciel-Zablocki argued that social-network data have many relevant characteristics, for example in terms of graph properties, but rarely are available in RDF form. He and his co-authors Alexander Schätzle, Thomas Hornung, and Io Taxidou therefore created an RDF benchmark from a large site’s (non-RDF) social-network data, and presented lessons learned and first results in *“Towards a SPARQL 1.1 Feature Benchmark on Real-World Social Network Data”*.

Tatiana Tarasova presented *“ParlBench: a SPARQL-benchmark for electronic publishing applications”* (co-authored with Maarten Marx), data characterised by large amounts of structured text and metadata on parts and on authors of these texts.

In the **closing discussion**, opinions were gathered on what makes a good benchmark and/or a good benchmark-creation methodology. The need to differentiate by use-cases was pointed out, and the advantages of using real-life queries were made explicit. It was also emphasized that use cases, data modelling, queries need to be taken into account when creating benchmarks. Clear definitions of quality measures (aka key performance indicators) were found lacking in some evaluation. The need to report tests and differential results on different criteria (rather than produce one-dimensional rankings of systems) was also stressed.

Author(s)	Title
Manolis Koubarakis	Benchmarking Geospatial RDF Stores (keynote)
Martin Przyjaciel-Zablocki, Alexander Schätzle, Thomas Hornung, Io Taxidou	Towards a SPARQL 1.1 Feature Benchmark on Real-World Social Network Data
Tatiana Tarasova, Maarten Marx	ParlBench: a SPARQL-benchmark for electronic publishing applications
Johannes Lorey, Felix Naumann	Caching and Prefetching Strategies for SPARQL Queries
Axel-C. Ngonga Ngomo	DBpedia SPARQL Benchmark - Performance Assessment with Real Queries on Real Data
Daniele Dell’Aglia, Marco Balduini, Emanuele Della Valle	On the need to include functional testing in RDF stream engine benchmarks

Further details: The workshop schedule and slides can be found on the workshop web pages at <http://ldbc.eu/events/bersys-2013>. The contributed papers and an abstract of Manolis Koubarakis’ Keynote can be found in the joint proceedings in the CEUR Volume 981 at <http://ceur-ws.org/Vol-981/>.

4.3 GRADES Workshop

The first GRADES workshop was organized by Peter Boncz (VUA) and Thomas Neumann (TUM) at SIGMOD/PODS 2013. This is the most prestigious venue for data management research. The place being in the US allowed us to attract many visitors from the bustling Big Data scene in Silicon Valley. This was further helped by teaming up with the GraphLab group of Carlos Guestrin (then Carnegie Mellon University, now moved to University Washington, and starting a start-up company). GraphLab is a graph programming framework originating from the data mining community.

The workshop solicits contributions from two perspectives:

- (i) Experiences. This includes topics that describe use case scenarios, datasets, and analysis opportunities occurring in real-life graph-shaped, as well as benchmark descriptions and benchmark results
- (ii) Systems. This includes topics that describe data management system architectures for processing of Graph and RDF data, and specific techniques and algorithms employed inside such systems.

Author(s)	Title
Carlos Guestrin (GraphLab)	Graphs at scale with GraphLab (keynote)
Adam Welc, Raghavan Raman, Zhe Wu, Sungpack Hong, Hassan Chafi, Jay Banerjee (Oracle Labs)	Graph Analysis. Do We Have to Reinvent the Wheel?
Eiko Yoneki (Univ of Cambridge); Amitabha Roy (EPFL)	Scale-up Graph Processing: A Storage-centric View
Andrey Gubichev (TU Munich); Srikanta Bedathur (IIIT-D); Stephan Seufert (Max Planck Institute for Informatics).	Sparqling Kleene - Fast Regex Paths in RDF-3X
David DeBoer, Wenchao Zhou, Lisa Singh (Georgetown University)	Using Substructure Mining to Identify Misbehavior in Network Provenance Graphs
Alex Averbuch (Neo Technology); Marco Quaggiotto, Ciro Cattuto, Andréanison (ISI Foundation)	Time-varying Social Networks in a Graph Database – a Neo4j Use Case
Roberto De Virgilio, Antonio Maccioni, Riccardo Torlone (Universita' Roma Tre)	Converting Relational to Graph Databases
Georgia Koloniari (University of Macedonia); Evaggelia Pitoura (University of Ioannina)	Partial View Selection for Evolving Social Graphs
Renzo Angles (Universidad de Talca); Arnau Prat (UPC Barcelona); David Dominguez (Sparsity Technologies); Josep Lluís Larriba Pey (UPC Barcelona).	Benchmarking database systems for social network applications
Elena Vasilyeva (SAP AG); Maik Thiele (Technische Universität Dresden); Christof Bornhövd (SAP Labs, LLC); Wolfgang Lehner (Technische Universität Dresden)	Leveraging Flexible Data Management with Graph Databases
Yongming Luo, George H.L. Fletcher (TU Eindhoven); Jan Hidders (TU Delft); Paul De Bra (TU Eindhoven); Yuqing Wu (Indiana University, Bloomington)	Regularities and dynamics in bisimulation reductions of big graphs
Cliff Joslyn, Sutanay Choudhury, David Haglin, Bill Howe, Bill Nickless, Bryan Olsen (Pacific Northwest National Lab)	Massive Scale Cyber Traffic Analysis: A Driver for Graph Database Research
Ma Haixin, Qian Weining, Xia Fan, Wei Jinxian, Yu Chengcheng, Aoying Zhou (East China Normal University)	On Benchmarking Online Social Media Analytical Queries
Nilesh Jain, Guangdeng Liao, Theodore Willke (Intel Corp)	GraphBuilder: Scalable Graph ETL Framework
Reynold Xin, Joseph Gonzalez, Michael Franklin (UC Berkeley)	GraphX: A Resilient Distributed Graph System on Spark
Michael Rudolf, Marcus Paradies, Christof Bornhövd (SAP); Wolfgang Lehner (Technische Universität Dresden)	SynopSys: Large Graph Analytics in the SAP HANA Database Through Summarization
Sungpack Hong, Jan Van Der Lugt, Adam Welc, Raghavan Raman, Hassan Chafi (Oracle Labs)	Early Experiences in Using a Domain-Specific Language for Large-Scale Graph Analysis

The workshop drew 30 paper submissions, which is much more than expected for a first SIGMOD workshop and the submissions were generally of high quality. Many of the presentations were industrial.

The workshop was visited by over 50 people, which is a lot given that this is a first time workshop with 8 concurrent SIGMOD/PODS workshops at the same time, and with its 50 participants was among the top-2 best attended events that day. Given that 16 paper were accepted, a nonconventional format was used, similar to Graph-TA, where after series of short 10-minute talk there would be a longer post-style presentation session for the papers. The reaction of the audience was supportive, and we plan to prepare another edition of this workshop in 2014.

Many of the papers are interesting. The Chinese paper “On Benchmarking Online Social Media Analytical Queries” is very close to the SNB topic area and includes analysis of a Chinese twitter-style website. In Y2 of LDBC we intend to start cooperating with these authors.

Similarly, LDBC people established further contacts with the GraphLab group, which also sponsored the workshop in return of LDBC sponsoring the GraphLab workshop. The latter, Organized July 3, 2013 in San Francisco drew 500 visitors. Three LDBC researchers went there and presented LDBC.

Further, at GRADES, LDBC established contact with Eric Sedlar, the director of Oracle Labs. This group is active in the graph programming framework field through its Green Marl prototype. Oracle Labs put us into contact with Luis Ceze from Washington University (authors of the graph programming framework Grappa), who they commissioned to create a graph algorithm benchmark. This benchmarking effort could supplement the SNB task force black spot in the area of graph programming frameworks, and possibly contribute the graph algorithm workload.

5 End User Dissemination

The goal of this task is to engage the broader IT public with LDBC. Generally speaking, in Y1 of LDBC we help back on this, since there were no LDBC benchmark drafts available, hence it was difficult to communicate much beyond the intention of the project. We do intent in Y2 to step up efforts in this area.

5.1 Public Relations

The project started out with a Kick Off press release, performed by all partners. This led to 9 announcements on various media. We note that a detailed list of PR events can be found in D7.7.3.

In February 2013, OGL and VUA also announced new BSBM results on RDF database systems on an extremely large size (150 billion triples). This result came out of the LOD2 EU project, however the lod2 blog posting on this (<http://lod2.eu/BlogPost/1584-big-data-rdf-store-benchmarking-experiences.html>) prominently announced LDBC as well.

In February as well, Peter Neubauer of Neo Technologies discussed the LDBC project on a premier forum of open source developers (FOSDEM). We hereby note that LDBC has promised to release all its benchmarks in open source.

Other PR activities involved various blog post (NEO,OGL) as well as a social media campaign during the GRADES and GraphLab workshops in summer 2013; which focused on promoting the then just released first version of the SNB data generator.

5.2 Early Questionnaire

In order to investigate the characteristics of the TUC membership, we ran a questionnaire among those customers of ONTO, OGL and NEO interested in LDBC. The questionnaire was organized quickly and therefore response was limited to eighteen filled in forms, which covered the TUC around October 2012.

<https://docs.google.com/spreadsheet/viewform?formkey=dGxuNORZbzdxaGlVMEZSeUNBZVIHcGc6MQ>

What follows is a graphical summary of the questionnaire outcome:

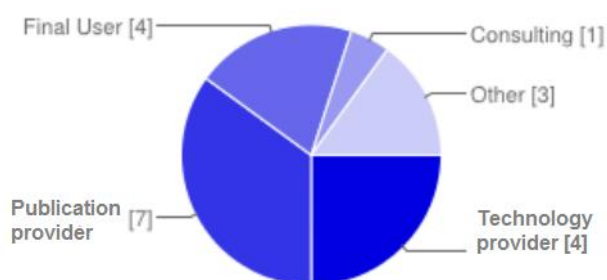


Figure 1: Different Sectors of the Respondents



Figure 2: Different Roles of the Respondents

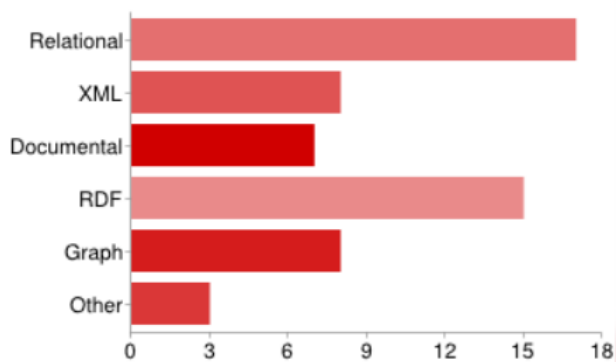


Figure 3: Different database models

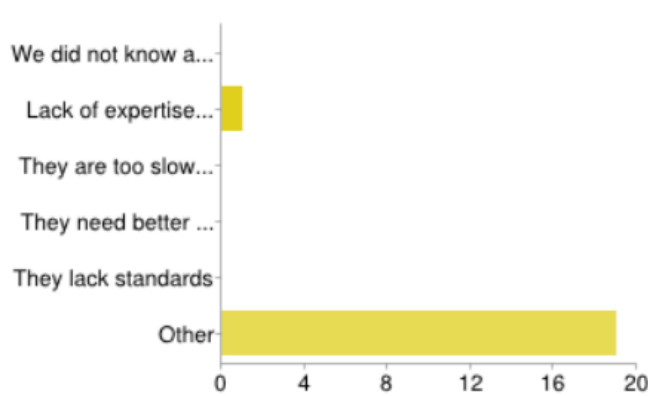


Figure 4: why not using Graph/RDF model now?

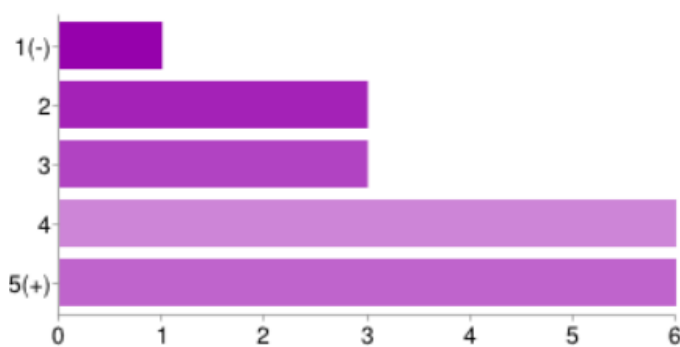


Figure 5: Familiarity with RDF?

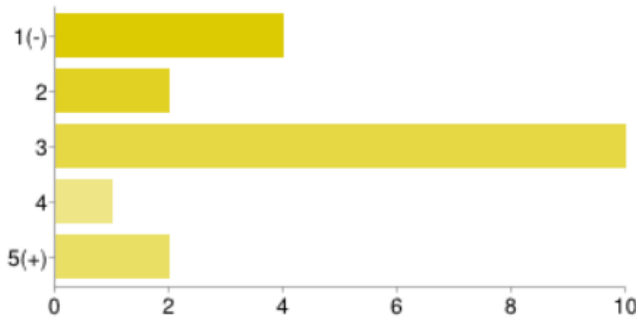


Figure 6: Familiarity with Graph Database Systems?

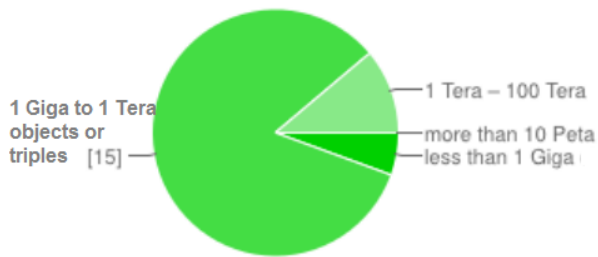


Figure 7: Size of the datasets?

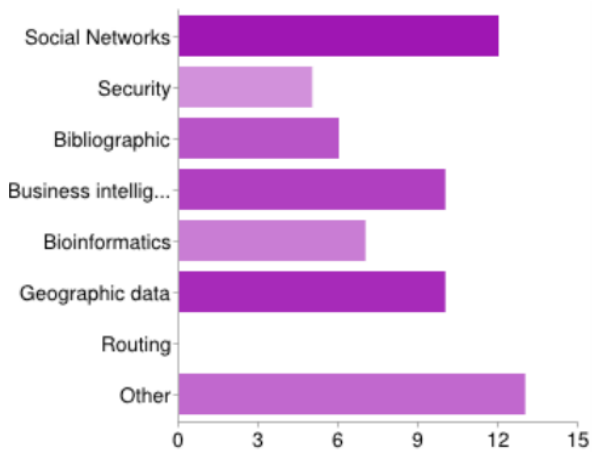


Figure 8: Data Domains?

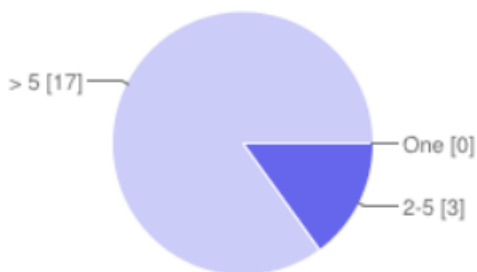


Figure 9: Number of data sources?

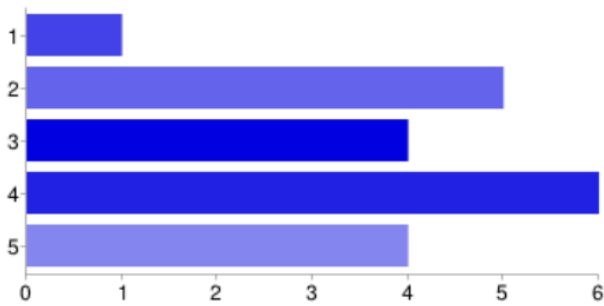


Figure 10: Importance of ACID capabilities (Y-axis)

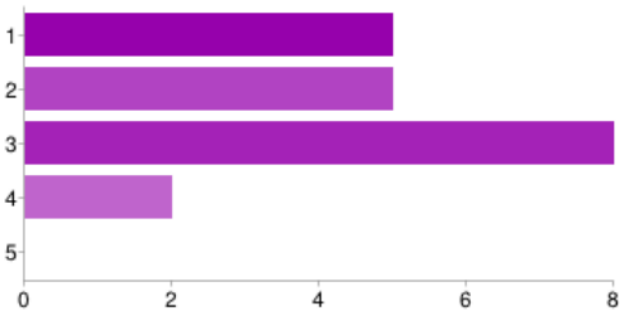


Figure 11: Application is Analytical (Y-axis)

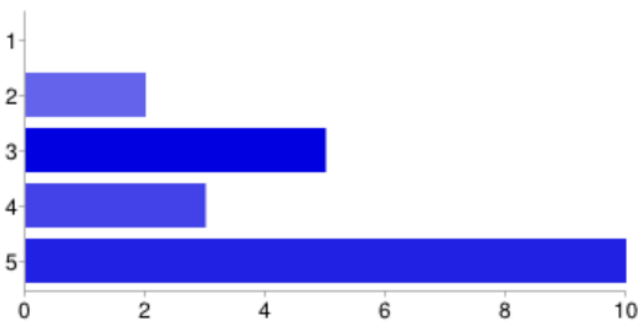


Figure 12: Importance of High Availability (Y-axis)

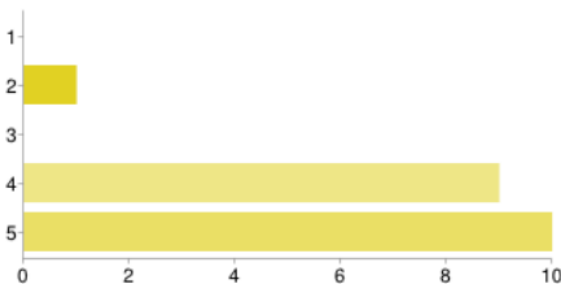


Figure 13: Importance of low-latency/real-time queries (Y-axis)

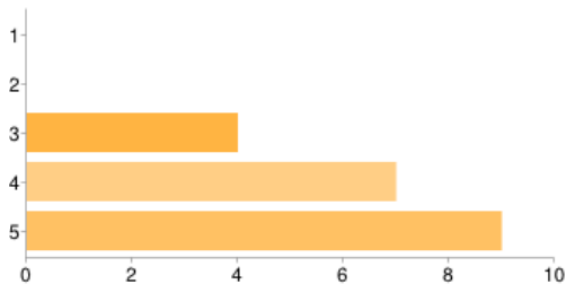


Figure 14: Importance of benchmarks to select a product (Y-axis)

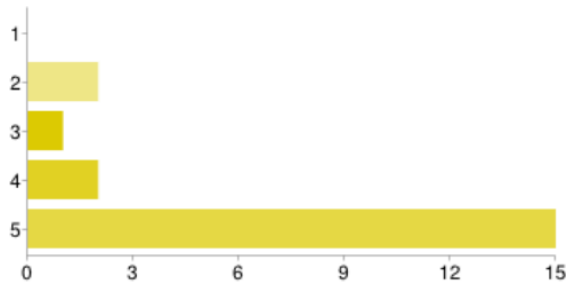


Figure 15: Importance of a standard query language (Y-axis)

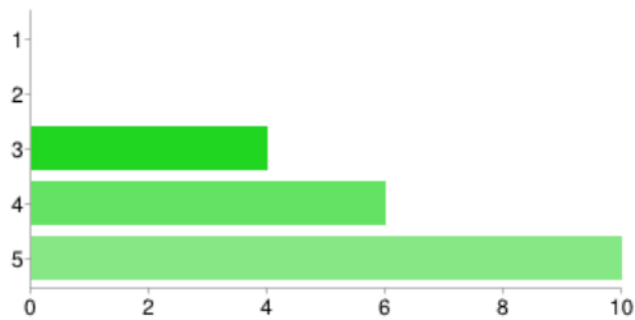


Figure 16: Benchmarks datasets and queries should resemble industry workloads?

From this questionnaire we concluded that among the TUC, which typically consist of data practitioners dealing with one of the database companies in LDBC has strong industrial needs and depend on benchmarking for product selection. They support a large variety of data-intensive applications. Their usage patterns are a mix of transactional and analytical queries, frequently needing ACID guarantees.

6 Conclusion

This deliverable has presented the dissemination activities for the LDBC project in Y1. We note that certain detailed information, such as a list of presentations and scientific publications is found elsewhere in D7.7.3.

Rather, in this document we commented on the specific LDBC activities for (i) engagement of a community of industry layers that might become later LDBC members, (ii) engagement of RDF and graph data management practitioners for providing input for LDBC benchmarks as well as feedback on drafts, (iii) engagement of the academic community to help out with benchmark design and also to promote LDBC benchmarks as instruments for measuring scientific progress, and (iv) engaging the larger IT public with LDBC such that its work is appreciated and used.

The overall conclusion is that dissemination in LDBC is on track. In the coming year, the fruits of the vendor engagement should be found in new members. The TUC is working well and is supporting the benchmark design task forces; and Y2 may allow us to broaden the development. End user engagement is expected to intensify in Y2 as usable (draft) benchmark specification will become available, and first audited benchmark results appear on the LDBC benchmark leaderboards.