

# LDBC

Collaborative Project

FP7 – 317548

---

## D5.5.2 Dissemination Report Y2

---

**Coordinator: Peter Boncz (VUA)**

**With contributions from: Irini Fundulaki (FORTH), Arnau Prat (UPC), Ioan Toma (UIBK), Damaris Coll (UPC), Serge Tymaniuk (IUBK), Ricard Tapias (UPC), Ilya Stefanov (ONTO)**

Deliverable nature:	R
Dissemination level: (Confidentiality)	CO
Contractual delivery date:	M24
Actual delivery date:	M24
Version:	1.0
Total number of pages:	50
Keywords:	

## Executive summary

This report gives an overview of the dissemination activities in the second 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 Y2 a series of meetings at TUC events or scientific conferences, 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 and feedback and input via the Technical User Community (TUC). There were two TUC meetings organized (November 2013 London, April 2014 Amsterdam).
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. This happened at regular scientific meetings, and at scientific workshops co-organized by LDBC.
4. Engaging the larger IT public in LDBC and its benchmarks. This is addressed using a public relations strategy aimed at the larger public coordinated by a newly created Dissemination Team (DT). This involved the creation of a website for the Linked Data Benchmark Council ([ldbcouncil.org](http://ldbcouncil.org)) including a new public image for the organization and the creating of a blog series as well as social media activities coordinated by community managers.

The appendix of the document is the detailed report on the activities of the DT.

## Document Information

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

<b>Deliverable</b>	<b>Number</b>	D5.5.2	<b>Title</b>	Dissemination Report Y2
<b>Work Package</b>	<b>Number</b>	WP5	<b>Title</b>	Dissemination

<b>Date of Delivery</b>	<b>Contractual</b>	M24	<b>Actual</b>	M12
<b>Status</b>	version 1.0		final x	
<b>Nature</b>	prototype <input type="checkbox"/> report <input type="checkbox"/> dissemination <input checked="" type="checkbox"/>			
<b>Dissemination level</b>	public <input type="checkbox"/> consortium <input checked="" type="checkbox"/>			

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

<b>Abstract (for dissemination)</b>	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.
<b>Keywords</b>	

<b>Version Log</b>			
<b>Issue Date</b>	<b>Rev. No.</b>	<b>Author</b>	<b>Change</b>
22.09.2014	0.1	Peter Boncz	Initial draft
29.09.2014	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 Vendor Strategy for LDBC Y3 .....	9
3 Technical User Community .....	11
3.1 Third TUC Meeting (London) .....	12
3.2 Fourth TUC Meeting (Amsterdam) .....	13
3.3 Future TUC strategy.....	14
4 Scientific Meetings.....	15
4.1 SIGMOD 2014 Programming Contest.....	15
4.2 LDBC Talks and Visits .....	15
4.3 Second Graph-TA Workshop.....	16
4.4 BerSys Workshop .....	17
4.5 GRADES 2014 Workshop.....	17
5 End User Dissemination.....	19
5.1 Dissemination Plan Progress.....	19
5.2 Dissemination Team Report.....	20
6 Conclusion.....	21
Annex A Dissemination Team Report Y2 LDBC .....	23
A.1 Global Methodology for the Dissemination team.....	23
A.2 Dissemination plan of actions .....	25
A.2.1 Blog Schedule.....	25
A.2.2 List & plan of dissemination platforms .....	26
A.3 Weekly dissemination efforts.....	27
A.3.1 Publications from 07/10/14 to 07/16/14 .....	27
A.3.2 Publications from 07/24/14 to 07/30/14 .....	28
A.3.3 Publications from 08/07/14 to 08/13/14 .....	31
A.3.4 Publications from 08/14/14 to 08/20/14 .....	32
A.3.5 Publications from 08/27/14 to 09/03/14 .....	33
A.3.6 Publications from 08/27/14 to 09/03/14 .....	34
A.3.7 Publications from 09/04/14 to 09/10/14 .....	34
A.4 Monitoring .....	35
A.4.1 Monitoring Tracking and Management Tools State of the art.....	35
A.4.2 Monthly reports .....	37
A.4.3 Monthly report 07/11/2014 – 08/11/2014.....	37
A.4.4 Monthly report 08/11/2014 – 09/11/2014.....	43

## 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
FOSDEM	Open Source Developers European Meeting
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
SNB	The LDBC Social Network Benchmark
SPB	The LDBC Semantic Publishing Benchmark

# 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 M24, are covered by the management deliverable D7.7.5.

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 M24 reports:

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

Both task forces delivered in Y2 their first benchmark specifications, which are now in Draft state. This means that the benchmark software (generator, driver, specific database implementations) as well as the

specification documents have been released. In particular, we note that the latter documentation, though not an explicit part of the LDBC Y2 EU deliverable set contains valuable information (like in Y1 the task force reports did) that are recommended reading material. See:

- SPB Draft Specification  
[https://github.com/ldbc/ldbc\\_spb\\_bm/blob/master/doc/LDBC\\_SPB\\_v0.1.pdf](https://github.com/ldbc/ldbc_spb_bm/blob/master/doc/LDBC_SPB_v0.1.pdf)
- SNB Draft Specification  
[https://github.com/ldbc/ldbc\\_snb\\_docs/blob/master/LDBC\\_SNB\\_v0.1.pdf](https://github.com/ldbc/ldbc_snb_docs/blob/master/LDBC_SNB_v0.1.pdf)

This deliverable D5.5.2 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 2014 (Barcelona), BerSys in June at VLDB2014 (Hangzhou, China) and GRADES in June at SIGMOD 2014 (Snowbird, USA). We have also published project progress papers about LDBC to the SIGMOD Record and ERCIM News journals, and described the project in keynotes at IDEAS 2013 and EDBT 2014. This is discussed in Section 4 of this deliverable.
- Task 5.1.5 End User Community Outreach (Start M1, End M30 / Lead IUBK:4). According to plan, but reinforced by Y1 review feedback, we created and executed an end-user dissemination plan that consists of a new website released jointly with the first benchmark drafts of the SNB and SPB (and materials such as a blog series), LDBC presence at technology user events (FOSDEM, GraphConnect, GraphLab Conference, EDF), mini benchmarks included in products and more.

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 deliverable 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. LDBC was incorporated as a non-profit organization in the UK in September 2013. 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 second 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 (*current LDBC members in italics*):

- RDF store software vendors (products),
  - *Oracle (Oracle RDF, Oracle noSQL)*
  - *IBM (DB2 RDF)*
  - *OpenLink Software (Virtuoso)*
  - *Ontotext (OWLIM)*
  - *Systap LLC (BigData)*
  - *SparqlCity (SparqlBase)*
  - MarkLogic (MarkLogic7)
  - Franz Inc (AllegroGraph)
  - YarcData (Urika)
  - Garlik (4Store/5Store– now part of Experian)
  - Clark&Parsia (Stardog)
  - Dydra (Dydra)
- graph database vendors (products)
  - *Neo Technology (Neo4j)*,
  - *Sparsity Technologies (Sparksee)*,
  - Objectivity(InfiniteGraph),
  - Kobrix Software (HyperGraphDB),
  - Orient Technologies (OrientDB),
  - Aurelius (Titan/Faunus),
- graph programming frameworks (products)
  - *Oracle (Green Marl)*
  - *IBM (System G)*
  - Facebook (Giraph core team)
  - GraphLab (GraphLab startup)
  - Flink (Stratosphere)

The appearance of commercial products in the area of graph programming frameworks is a new phenomenon for Y2 of the LDBC. In the case of Oracle, the Green Marl work on expression graph algorithms is being integrated into its products, and has now appeared first in the Oracle RDF engine. GraphLab and Flink are new start-ups in this area (the latter is not squarely directed at graph problems but Stratosphere has iterative extensions that allow it to run graph algorithms).

IBM has been working on its System G product and this IBM group is the one that contacted LDBC for membership.

New prototypes in the distributed graph programming area are still appearing, e.g. Signal/Collect (University Zurich), Spark (Berkeley), and Grappa (University Washington). 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.

## 2.2 Membership Strategy

LDBC allows three classes of members: Company Members, Non-Profit Members and Individual Members. Non-Profit Members have the same rights as Company Members but may pay a lower fee. Individual Members have less formal power as other members, but may still participate in Board Meetings and access all LDBC information. LDBC was incorporated at M12 with two initial members:

- Sparsity (Company Member) and
- Peter Boncz (Individual Member).

Still in 2014, we added

- STI2 (Non-Profit Member),
- FORTH (Non-Profit Members),
- Neo Technologies (Company Member),
- Openlink Systems (Company Member) and
- Ontotext (Company Member).

Over the summer of 2014 four US companies became a member:

- Oracle Labs (Company Member),
- IBM Corp. (Company Member),
- Systap LLC (Company Member) and
- SPARQLcity (Company Member) \*\* still in process of membership acquisition

From the fact that LDBC acquired these new members in Y2 we conclude that the strategy laid out in D5.5.1 to create an environment where companies want to join as member of LDBC is working. In order to main the current levels of confidence, LDBC does need to focus on its momentum and progress with the benchmark adoption process such that by spring 2015 the SPB and SNB reach adoption with full audited results out.

## 2.3 Vendor Strategy for LDBC Y3

In 2015, the LDBC will pursue the membership strategy (aimed at companies) that increasingly leverages its benchmarking activities to attract new members. Like the past years, there will of course be continued personal engagement by the LDBC project members as well as LDBC board members, but the fact that benchmarks are out and companies want to participate for marketing purposes is going to take a more important role than before. The benchmark-based strategy is summarized as follows:

- On the SNB Interactive Workload, the focus will be in engaging parties to submit full benchmark results. We expect OpenLink (Virtuoso), Sparsity (Sparksee) to be the first. Other parties close to engaging would be Neo Technologies (neo4j), Oracle and SPARQLcity. This engagement can then further be used to draw Franz (Allegrograph), YarcData and MarkLogic into LDBC benchmarking (&membership). Same holds for IBM.

- On the SNB Business Intelligence Workload there will be significant development in Y3. Forerunners are expected to be Openlink and SPARQLcity. We expect MarkLogic and YarcData to engage on this workload as well as Oracle and IBM.
- With the help of funding by Oracle, funds have been secured to continue research needed for the creation of the SNB Graph Analytics workload in LDBC, in addition to the already existing Interwative Workload and the just mentioned Business Intelligence Workload. The groups involved are UPC/VUA on the one hand, but also the TU Delft group of Alexandru Iosup (Graphitti project). Also the Universty of Washington is involved (Luis Ceze – graphbench.org). From the LDBC point of view, the goal is a set of graph algorithms and evaluation metrics. UPC and LDBC specifically commit to making modifications to the SNB data generator such that the results of these make sense. Thanks to the Oracle funding, this commitment is extended beyond the EU project end in April 2014.
- In the SPARQ and RDF arena, we expect in Y3 full results on the SPB from a number of vendors. The first ones will be Ontotext (OWLIM) and Openlink (Virtuoso), but we also expect Oracle (RDF) and SPARQLcity to execute the benchmark. We will approach Franz (Allegrograph) and YarcData in the respect as well.
- Concerning RDF, the SPB task force will have resources in Y3 to consider broadening the functional space. The three WP4 deliverables of M24 describe three of the possible directions. A possible course of action is to enrich SPB with rich reasoning tests. This might lead to LDBC throwing its weight behind the SPIMBench (Semantic Publishing Instance Matching Benchmark) initiative. Such a course of action should preferably include participation of Clark&Parsia, whose Stardog product has advanced reasoning functionality.

The above strategy will continue to be executed using personal contacts from LDBC project members, directors, as well as face-to-face meetings at TUC meetings and LDBC Board Meetings. But, we think that the availability of benchmark results and population of LDBC by a sizeable part of the graph and RDF software industry itself will form an additional stimulus to broaden the vendor membership of LDBC.

### 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 Y2 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 Y2 can be found in D7.7.5.

Further, the LDBC has a web-portal where TUC members can share information related to benchmarks, workloads and datasets. In Fall 2014 this portal was moved to <http://wiki.ldbcouncil.org>. 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 Third TUC Meeting (London)

This meeting was held on site with Neo Technologies [www.graphconnect.com/london](http://www.graphconnect.com/london) on November 19 in London. The meeting was packed with more than 40 participants, and was preceded by a 15 minute section of Peter Boncz inside the keynote presentation of the CEO of Neo Technology at GraphConnect, which introduced the LDBC to the neo4j user audience (>300 persons).

Notable presentations were reflections by the BBC on the SPB benchmark, which follows its use case, and usage of the SNB data generator by the team from Chinas well as from IBM. The IBM System G group later in Y2 applied for LDBC membership. There was also a status update by Luis Ceze on the graphbench.org initiative. The director of the Oracle RDF product (Xavier Lopez – also on teh LDBC Advisory Board)gave a presentation about an effort to standardize the data model of graph databases (Property Graphs) through the W3C. This effort to our knowledge later has not prospered inside W3C but is being supported by LDBC as we see value in more compatibility between RDF and graph database systems.

Organisation	Person(s)	Host	Presentation
BBC	Johan Hjerling	ONTO	<a href="#"><u><i>BBC Linked Data and the Semantic Publishing Benchmark</i></u></a>
Andreas Both	Unister	OGL	<a href="#"><u><i>Ontology-driven applications in an e-commerce context</i></u></a>
Nuno Carvalho	Fujitsu Laboratories Europe	OGL	<a href="#"><u><i>Fujitsu RDF use cases and benchmarking requirements</i></u></a>
Europeana	Robina Clayphan	OGL	<a href="#"><u><i>Europeana and Open Data</i></u></a>
East China Normal University	Minqi Zhou, Weining Qian	VUA	<a href="#"><u><i>Elastic and realistic social media data generation</i></u></a>
Shapespace	Andrew Sherlock	OGL	Shapespace Use Case
Telenor	Sebastian Verheughe	NEO	<a href="#"><u><i>Real-time Resource Authorization</i></u></a>
IBM	Keith Houck	UPC	<a href="#"><u><i>Benchmarking experiences with System G Native Store</i></u></a>
University of Zurich	Abraham Bernstein	VUA	<b><i>Streams and Advanced Processing: Benchmarking RDF querying beyond the Standard SPARQL Triple Store</i></b>
University of Washington	Luis Ceze	VUA	<b><i>Grappa and GraphBench Status Update</i></b>
Ontotext	Atanas Kiryakov	ONTO	<a href="#"><u><i>Large-scale Reasoning with a Complex Cultural Heritage Ontology (CIDOC CRM)</i></u></a>
National and Kapodistrian University of Athens / CWI	Kostis Kyzirakos	VUA	<a href="#"><u><i>Geographica: A Benchmark for Geospatial RDF Stores</i></u></a>
Oracle	Xavier Lopez	UPC	<b><i>W3C Property Graph progress</i></b>
University Zurich	Thomas Scharrenbach	IUBK	<b><i>PCKS: Benchmarking Semantic Flow Processing Systems</i></b>
Unit for Digital Documentation, University of Oslo, Norway	Christian-Emil Ore	ONTO	<b><i>CIDOC-CRM</i></b>

### 3.2 Fourth TUC Meeting (Amsterdam)

The fourth TUC meeting was held in April 2014, ad joint to the M18 plenary meeting of the LDBC project. At this time, the Draft versions of both SNB (Interactive Workload) and SPB had shaped up and were presented by LDBC members in TUC talks. The final release of the draft benchmarks was in both cases hampered by software availability and documentation, and in the end took place in June.

In this TUC, again many different graph and RDF use cases were presented and feedback was requested on the LDBC benchmarks. Significant talks where the ones by Alexandu Iosup. The work of his group on benchmarking graph programming frameworks (Graphitti project) was picked up later for LDBC cooperation, specifically on the development of the Graph Analytics workload for the SNB. We note that the alternative cooperation with graphbench.org had not progressed much, because the UWashington project is delayed. Again, it was also Oracle who stimulated this cooperation (Oracle also funds graphbench.org).

Another important presentation was the one about Swissprot; this represents a quite demanding SPARQL use case and provided many examples of where RDF technology currently breaks.

Organisation	Person	Host	Title
TU Delft	Alexandru Iosup	VUA	<a href="#">Towards Benchmarking Graph-Processing Platforms</a>
Kings College	Mike Bryant	NEO	<a href="#">The EHRI Project: Archival Integration with Neo4j</a>
University of Magdeburg	Thilo Muth	NEO	<a href="#">MetaProteomeAnalyzer: a graph database backed software for functional and taxonomic protein data analysis</a>
Janssen Pharmaceutica / Johnson & Johnson	Davy Suvee	ONTO	<a href="#">Euretos Brain - Experiences on using a graph database to analyse data stored as a scientific knowledge graph</a>
TU Eindhoven	Yongming Luo	VUA	<a href="#">Regularities and dynamics in bisimulation reductions of big graphs</a>
TU Delft	Christopher Davis	NEO	<a href="#">Enipedia - Enipedia is an active exploration into the applications of wikis and the semantic web for energy and industry issues</a>
Sysunite	Bastiaan Bijl	OGL	<a href="#">Using a semantic approach for monitoring applications in large engineering projects</a>
Geodan	Frans Knibbe	OGL	<a href="#">Benchmarks for geographical data</a>
University of Rome & FAO	Armando Stellato	ONTO	<a href="#">VocBench2.0, a Collaborative Environment for SKOS/SKOS-XL Management: scalability and (inter)operatibility challenges</a>
TopQuadrant	Ralph Hodgson	OGL	<a href="#">Customer experiences in implementing SKOS-based vocabulary management systems</a>
European Bioinformatics Institute	Simon Jupp	ONTO	<a href="#">Delivering RDF for the life science at the European Bioinformatics Institute: Six months in</a>
Swiss Institute of Bioinformatics	Jerven Bolleman	ONTO	<a href="#">Breakmarking UniProt RDF. Sparql queries that make your database cry...</a>
Digital Heritage Netherlands	Rein van 't Veer	OGL	<a href="#">Time and space for heritage</a>

### **3.3 Future TUC strategy**

The fifth TUC meeting, will be held Nov 14 2014 in Athens, organized by FORTH at Athens University. The overriding themes of that fifth TUC meeting are gathering feedback and results from running the SPB and SNB Interactive Workload, as well as continuing the momentum in committing graph and RDF data management companies to join the LDBC. It is anticipated that within the lifetime of the EU project LDBC, a sixth TUC meeting will be organized at the end of March 2015.

After the EU project ends, the TUC meetings are recommended continue under the banner of [ldbncouncil.org](http://ldbncouncil.org) (the then independent Linked Data Benchmark Council).

## 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 D7.7.5.

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 SIGMOD 2014 Programming Contest

The premier database research conference SIGMOD each year organizes a programming contest. This year's contest was centered around graph analytics. The organizers from Albany University approached the LDBC for help, and LDBC provided support so that the dataset used would be generated by the SNB data generator.

We consider this an important success for LDBC, concerning its scientific goals.

The usage of the SNB data generator provided important visibility for LDBC and more importantly for the SNB data generator and the SNB benchmarks as tools for benchmarking graph analytics, in the database research community. During a plenary session the organizers complimented LDBC for the "state of the art" graph generator.

### 4.2 LDBC Talks and Visits

While deliverable D7.7.5 will give a full overview of papers and presentation, here a some additional information on notable presentation in scientific audiences on LDBC.

- **LDBC Keynotes at IDEAS 2013, EDBT/ICDT 2014 and SEMANTICS2014.** Peter Boncz (VUA) gave the keynote address at the IDEAS conference on October 2011, and another one on March 28 2014 at the EDBT/ICDT conference in Athens. The latter is considered a major database research conference, where EDBT is the applied side and ICDT the theoretical. In both cases these keynote were fully dedicated to LDBC, and described in detail all work of the SNB, mostly. At SEMANTICS2014 there was also a keynote talk by Orri Erling (OLG) that prominently.
- **Activities by the coordinator (UPC).** Josep L. Larriba Pey has been in Chile from July 11th until August 24th in a research visit to the "Universidad de Talca". There, he has worked in four aspects regarding LDBC. On one side, he has given three talks at "Universidad de Chile" and "Universidad de Talca" at its Talca, and Curicó campuses, where the SNB has been explained in detail. Second, he has given a course on "Advanced databases" explaining the different aspects of Graph databases and the use of LDBC as a benchmark for such DBMSs. Third, in collaboration with Renzo Angles (who worked at VUA in the calendar year 2013 on LDBC), the students of course "Advanced Databases" at the Computer Science Engineering curriculum will be using LDBC SNB in their classes to implement it on different DBMSs like Neo4j and Sparksee. Finally, he has been working with Renzo Angles ("Universidad de Talca") on benchmark follow ups, related to the use of LDBC's SNB data generator, DATAGEN.

Josep L. Larriba has been invited to the Scalperf 2014 workshop (Bertinoro, Italy, Sept 22-26 2014) to give a presentation on benchmarking for a scientific audience, which focused on LDBC. See the abstract here: <http://www.dei.unipd.it/~versacif/scalperf14/abstracts/index.html>.

- **Reasoning Web Summer School.** DAMA-UPC and Sparsity were invited to give part of the Reasoning Web Summer School (Athens, Sept. 8-11, 2014) on the topic Graph databases. The three hour part of the course was oriented partly to give a review of technologies and research on Graph Databases and 30 minutes of the course were devoted to explain the importance of benchmarking and the use of LDBC as a benchmark. This course provided an interesting contact with Dr Anastasios Kementsietsidis from Google, who is leading in benchmarking at Google.

LDBC's Semantic Publishing **Instance Matching Benchmark** (SPIMBench) was also presented at the poster session of the Reasoning Web Summer School by Irini Fundulaki (FORTH). The reaction from researchers was very positive since SPIMBench introduces interesting tests that take into account rich schema information expressed in the form of OWL reasoning rules.

- **LDBC Booths:** were present at: EDF2014, ESWC2014 and SEMANTICS2014 (via IUBK). At all these events we had a table, poster and we distributed flyers. The LDBC Poster attracted a lot of positive comments from both the research and industry. At ESWC2014, LDBC was also briefly presented in the EU networking sessions (IUBK & FORTH).

LDBC was also present in other events, here we list the science and research oriented venues (see the next chapter).

### 4.3 Second Graph-TA Workshop

DAMA-UPC and Sparsity have organised the second Graph-TA workshop on February 21st 2014. The objective of this workshop is to create synergies among the different research groups that participate. This is achieved by setting up sessions with short (5 minutes) presentations and 1 hour poster session after those presentations. In this specific workshop instance, there were 4 presentations about benchmarking, of which two were on the progress of the SPB and the LDBC project.

Author(s)	Title
Hassan Chafi	<a href="#">"Graph Analytics Research at Oracle Labs"</a>
Fernando Orejas and Elvira Pino	<a href="#">"Incremental Model Synchronization with Triple Graph Grammars"</a>
Peng Wang, Veronique Eglin, Christophe Garcia, Christine LARGERON, Josep Lladós and Alicia Fornés	<a href="#">"Handwritten Word Spotting Based on Graph Representation"</a>
A. Godoy, M. Sales-Pardo and R. Guimerà	<a href="#">"Long-Term Evolution of Email Communication Networks"</a>
Ricard Gavaldà and Alberto Lumberras	<a href="#">"Trust Metrics and User Recommendation in Social Networks"</a>
Xavier Muñoz, Elisabet Burjons and Juraj Hromkovic	<a href="#">"Bounds for the required advise for 1-competitiveness in the k-server problem in a k-star"</a>
Roar Audun Myrset and Sebastian Verheughe	<a href="#">"Using a graph database for resource authorization"</a>
Arnau Prat-Perez	<a href="#">"High Quality, Scalable and Parallel Community Detection for Large Real Graphs"</a>
Peter Boncz.	<a href="#">"LDBC: Benchmarking Graph Data Management Systems"</a>
Francesc Serratosa	<a href="#">"Learning Weights for Graph Matching Edit Costs"</a>
Anjan Dutta	<a href="#">"Product Graph based Subgraph Matching"</a>
Arnau Padrol, Guillem Perarnau, Julian Pfeifle and Victor Muntés-Mulero	<a href="#">"Overlapping communities"</a>
Emili Sapena, Lluís Padró and Jordi Turmo	<a href="#">"A Global Relaxation Labeling Approach to Coreference Resolution"</a>
Glyn Morrill and Oriol Valentín	<a href="#">"Towards Logical Syntactic Structures as Graphs"</a>
Norbert Martínez	<a href="#">"Towards query algebras for graph operations"</a>
Toyotaro Suzumura	<a href="#">"ScaleGraph: A Billion-Scale Graph Analytics Library"</a>
Venelin Kotsev	<a href="#">"LDBC - Semantic Publishing Benchmark"</a>
Joan Guisado	<a href="#">"Massive Query Expansion by Exploiting Graph Knowledge Bases"</a>
Christophe Scholliers, Coen De Roover, Elisa Gonzalez Boix.	<a href="#">"A GraphDatabase Storage Model for Change-Oriented Software Development"</a>
Sergio Gomez	<a href="#">"Community structure in complex networks at different resolution levels"</a>
J. Barceló, M.P. Linares and O.Serch	<a href="#">"On some Graph Related Problems in Transportation Analysis"</a>
Ali Naderi, Jordi Turmo and Horacio Rodríguez	<a href="#">"Graph-Based Entity Linking"</a>
Javier de San Pedro, Jordi Cortadella and Toni Roca	<a href="#">"Hierarchical Floorplanning of Chip Multiprocessors using Subgraph Discovery"</a>
David F. Nettleton	<a href="#">"User and adversary driven anonymization of online social networks represented as graphs"</a>
Toni Valles	<a href="#">"Extracting information from a network by grouping nodes into different partitions"</a>

## 4.4 BerSys Workshop

BeRSys 2014 was held together with the SSW 2014 workshop, this time in conjunction with VLDB 2014 on September 5th, 2014. We received a very small number of submissions (only six) from which we accepted the following three papers:

Author(s)	Title
Fopa Leon Constantin, Fabrice Jouanot, Alexandre Termier, Maurice Tchuente and Oleg Iegorov	Benchmarking of triple stores scalability for MPSoC trace analysis
Shi Qiao and Meral Ozsoyoglu	RDF Benchmarks vs Real RDF Datasets: How Different Are They Really?
Diego Ernesto Rosa Pessoa, Marcelo Iury de Sousa Oliveira, Bernadette Farias Lóscio and Ana Carolina Salgado	A proposal for RDF data integration benchmarking

Fei Wu from Google gave the keynote talk "Semantics: Bridge to Search of the Future "; the joint workshops were relatively well attended considering that the workshops were held at the same time as TPC-TC workshop.

## 4.5 GRADES 2014 Workshop

The first GRADES workshop was organized by Peter Boncz (VUA) and Josep Lluís Larriba Pey (UPC) at SIGMOD/PODS 2014. The keynote talk was given by Avery Ching from Facebook, and the VP of Apache Giraph, on the use scenarios of Giraph at Facebook.

Author(s)	Title
Zhisong Fu, Bryan Thompson, Michael Personick (SYSTAP, LLC)	<a href="#">MapGraph: A High Level API for Fast Development of High Performance Graph Analytics on GPUs</a>
Norbert Martinez-Bazan, David Domínguez-Sal (Sparsity Technologies)	<a href="#">Using semijoin programs to solve traversal queries in graph databases</a>
Vasiliki Kalavri (KTH); Stephan Ewen, Kostas Tzoumas ( TU Berlin); Vladimir Vlassov (KTH); Volker Markl (TU Berlin); Seif Haridi (KTH)	<a href="#">Asymmetry in Large-Scale Graph Analysis, Explained</a>
Nandish Jayaram (University of Texas); Arijit Khan (ETH); Chengkai Li (University of Texas); Xifeng Yan (UCSB); Ramez Elmasri (University of Texas)	<a href="#">Towards a Query-by-Example System for Knowledge Graphs</a>
Semih Salihoglu, Jennifer Widom (Stanford University)	<a href="#">HELP: High-level Primitives For Large-Scale Graph Processing</a>
Ramesh Subramonian (LinkedIn/Oracle Labs)	<a href="#">Graph Processing on an "almost" Relational Database</a>
Andrey Gubichev, Manuel Then (TU Munich)	<a href="#">Graph Pattern Matching -- Do We Have to Reinvent the Wheel?</a>
Ilie Tanase, Yinglong Xia (IBM); Lifeng Nai (Georgia Institute of Technology); Yanbin Liu, Wei Tan, Jason Crawford, Ching-Yung Lin (IBM)	<a href="#">A Highly Efficient Runtime and Graph Library for Large-Scale Graph Analytics</a>
Arnau Prat (DAMA-UPC); David Domínguez-Sal (Sparsity Technologies)	<a href="#">How community-like is the structure of synthetically generated graphs</a>
Raghavan Raman, Oskar van Rest, Sungpack Hong (Oracle Labs); Zhe Wu, (Oracle); Hassan Chafi, (Oracle Labs); Jay Banerjee (Oracle)	<a href="#">PGX.ISO: Parallel and Efficient In-Memory Engine for Subgraph Isomorphism</a>
Yodsawalai Chodpathumwan (University of Illinois at Urbana-Champaign); Arash Termehchy (Oregon State University); Yizhou Sun (Northeastern University); Amirhossein Aleyasin (University of Illinois at Urbana-Champaign); Jose Picado (Oregon State University)	<a href="#">Toward General Similarity Search Over Graphs</a>
Marcus Paradies, Michael Rudolf (SAP AG); Christof Bornhövd (SAP Labs); Wolfgang Lehner (TU Dresden)	<a href="#">GRATIN: Accelerating Graph Traversals in Main-Memory Column Stores</a>

The workshop was visited by about 50 people, which is a lot given that there are 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 12 papers 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. We have applied to organize a third GRADES workshop in 2015 (beyond the scope of the EU LDBC project).

In the GRADES2014 workshop there was participation of new members IBM, Oracle Labs and Systap. This also led to a separate meeting over the topic of taking initiative to bringing the property graph and RDF data models (and hence also their query languages SPARQL and Cypher) closer together. This led to an ongoing collaboration into the foundations of such a PG-RDF data model mapping with Olaf Hartif of University Waterloo.

We specifically raise attention to the DAMA-UPC paper (“How community-like is the structure of synthetically generated graphs?”) since it compares a number of graph structure characteristics between real social network datasets, the LDBC SNB dataset and the output of other synthetic graph generators. The outcomes of these tests indicate that the LDBC SNB is much more realistic than previous synthetic graph generators (though they also provide interesting directions for improvement).

From conversations with the keynote speaker Avery Ching (Facebook, Apache Giraph) the plan was formed to run similar tests (using giraph) on the real Facebook graph. This would happen at Facebook, but could be assisted by summer interns. It is planned that some LDBC students will go to Facebook in summer 2015 for this purpose. These activities align with the (Oracle-financed) initiative to enhance the SNB data generator while defining its Graph Analytics workload over the year 2015 (i.e. also after the EU project LDBC has finished).

The GRADES workshop also led to a conversation with Facebook on LDBC membership which is generally positive and is being continued at time of this writing.

## 5 End User Dissemination

The goal of this task is to engage the broader IT public with LDBC. We have from the start chosen to initiate the end-user campaign at the moment the LDBC would have something to show for, this being a first set of draft benchmark (specification and software). This moment was reached in June 2014.

### 5.1 Dissemination Plan Progress

The end-user dissemination activities were written up in a dissemination plan that was shared with the PO and the project reviewers; also after their feedback at the Y1 review.

We list below the proposed actions and briefly report on their status:

- **Documentation on Github:** The SNB and SPB are available in the Github at [github.com/ldbc/ldbc\\_spb\\_bm](https://github.com/ldbc/ldbc_spb_bm) respectively [github.com/ldbc/ldbc\\_snb\\_datagen](https://github.com/ldbc/ldbc_snb_datagen). The documentation of both benchmarks was improved and is available at: [github.com/ldbc/ldbc\\_spb\\_bm/tree/master/doc](https://github.com/ldbc/ldbc_spb_bm/tree/master/doc) respectively [github.com/ldbc/ldbc\\_snb\\_docs](https://github.com/ldbc/ldbc_snb_docs).
- **Dissemination Materials:** there are “Getting Started with” blogs ([ldbcouncil.org/blog/getting-started-semantic-publishing-benchmark](http://ldbcouncil.org/blog/getting-started-semantic-publishing-benchmark) and [ldbcouncil.org/blog/getting-started-snb](http://ldbcouncil.org/blog/getting-started-snb)) to help people get started with the SPB and SNB. This material will be supplemented in October 2014 with presentations on SNB and SPB plus a webinar.
- **Mini Benchmarks:** OWLIM now comes standard with a script that downloads, compiles and runs the SPB benchmark installation. A similar feature will be added for Virtuoso.
- **Professionally designed website:** The LDBC consortium subcontracted Berta Hernández, a professional designer from Barcelona, to create the graphic design of LDBC corporate image and its new web site. Note that the LDBC EU project keeps using the old LDBC logo, whereas the new LDBC logo is branded to the industry-council by the same name, which will outlive the EU project. The website respects the fact that there are many audiences for the LDBC (IT professionals looking for benchmark results, Industry players considering membership, General public wondering what is the use of LDBC and Developers involved in benchmarking graph/RDF software).
- **Community Managers.** Two community managers were hired, namely Ricard Tàpias (UPC), taking care of promoting the portal and disseminating the benchmarking results among the Graph community and Iliya Enchev (ONTO) taking care of promoting the portal and disseminating the benchmarking results among the RDF community.
- **Blog Series.** Participants in the LDBC project will prepare a blog series highlighting the newly developed benchmarks and achievements of the LDBC. This blog series will be published on the new LDBC legal entity website and supported in social media (pointing to articles and engaging in discussion around them); the Community Manager is coordinating this.
- **Email Actions.** The LDBC dissemination team is actively promoting LDBC on social media channels and forums in order to attract more people to the new web site of the LDBC non-profit entity and to the TUC meetings. Targeted mailing actions to the customers of the LDBC partners have been performed and will continue in the future.
- **Community Events.** Rather than focusing on only scientific events, LDBC has been present in developer community events:
  1. FOSDEM2014 – Peter Boncz (VUA) and Peter Neubauer (NEO) represented LDBC giving invited talks; The former talk was encouraging developers in the room to download SPB and SNB software, while focusing on the SNB data generator. The latter talk was on progress with the neo4j graph database system, highlighting NEO’s involvement in LDBC.
  2. GraphConnect2014 – Peter Boncz (VUA) gave a 15 minute talk section in the keynote by Emil Eifrem (CEO of NEO), explaining the mission and goals of the LDBC, pointing to the open-source resources available.

3. GraphLab2014 – Arnau Prat (UPC) represented LDBC at the event promoting the LDCB results. GraphLab Conference is an event organized by GraphLab, a world class leaders of graph programming frameworks industry. In this event, data scientist, engineers and big data industry technical stakeholders joint to show and discuss their latest experiences and products. GraphLab Conference 2014, the third edition of the GraphLab Conference series, was held in San Francisco on 21st July 2014. An LDBC member was present in the event, which was announced in the social networks (twitter), to draw the attention of the audience in the event. We distributed large brochures (DinA4) explaining the different initiatives LDBC promotes, and actively explained them to the audience. This audience was composed by both relevant linked data technology vendors that could potentially implement the benchmarks developed by LDBC or join the LDBC consortium (Titan, Allegro Graph, GraphLab, Oracle, MongoDB), and to technology users that could use the benchmark to decide which technology to adopt. The audience looked receptive and interested to the initiatives promoted by LDBC, as well as the software provided in the LDBC repositories such as the data generators.

## 5.2 Dissemination Team Report

In order to streamline the end-user dissemination activities, LDBC decided to install a dissemination team that concalls once a week and coordinates a social media strategy.

- Ricard Tàpias (UPC) – community manager graph
- Iliya Enchev (ONTO) – community manager RDF
- Serge Tymaniuk (IUBK) – website manager
- Damaris Coll (UPC) – strategic marketing and DT lead

The Dissemination Team (DT) has a weekly short concall to coordinate its activities of the week. In this concall, usually the LDBC Project Management Board is also present:

- Ioan Toma (IUBK)
- Peter Boncz (VUA)
- Josep Lluís Larriba Pey (UPC)

The blog series published at the [ldbncouncil.org](http://ldbncouncil.org) website is the backbone of the communication strategy. The appendix that follows is the detail report written by the Dissemination Team over its activities in LDBC Y2.

## 6 Conclusion

This deliverable has presented the dissemination activities for the LDBC project in Y2. We note that certain detailed information, such as a list of presentations and scientific publications is found elsewhere in D7.7.5.

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, and specifically in the area of end-user engagement has significantly increased its activities. In Y3 we will be looking to further expand the membership base as well as further improve the effectiveness of the end-user communication campaign.



## Annex A Dissemination Team Report Y2 LDBC



*Dàmaris Coll (UPC)*  
*Iliya Stefanov (ONTO)*  
*Ricard Tapias (UPC)*  
*Serge Tymaniuk (IUBK)*

In this deliverable, we describe the dissemination activities of LDBC on the social networks, the LDBC website blog and third parties forums and blogs, which represent the effort for the project to disseminate its work with the most up to date social platforms and with a special intensity to build a community around benchmarking. 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.

In particular the topics covered in this document include all the documentation generated by the dissemination team when analysing, creating and performing its actions:

- (i) Creation of a global methodology of work for the Dissemination team.
- (ii) Creation of the Dissemination Plan of Action.
- (iii) Analysis of the Weekly Dissemination Efforts.
- (iv) Analysis of the monitoring tools.
- (v) Monitoring reports.

### A.1 Global Methodology for the Dissemination team

This section describes the amount of work that was undertaken by the team in order to set up an initial startup actions that were going to help to elaborate a methodology of work for the dissemination activities to be performed by the Community Managers (Ricard Tapias and Iliya Stefanov) which include the corresponding reporting of the actions performed:

- (i) **Initial actions.** In order to organise the work along the course of the remaining months of the project:
  - **Marketing procedures.** Making a state-of-the-art guidelines based on best practices, effective messages and methods of interaction in social media.
  - **Content.** The partners of LDBC have the compromise to create an initial number of blog entries <sup>1</sup>(1). Those blog entries are the basis for the content to be distributed to the Social Networks.
  - **Social networks considered.** Establishing the Social Networks in which the dissemination work will be executed. The initial actions regarding those SNs were:
    - **All the CMs will be registered.** All the CMs registered and have access to the LDBC accounts in the different SNs selected.
    - **All the CMs will subscribe to the relevant groups for LDBC.** There will be a constant search for relevant groups/forums/websites to LDBC where the CMs will interact.
  - **Decision on the tools to monitor the activity of the SNs.** A number of tools to monitor the activity generated will be used. The initial task involved to create a simple state-of-the-art of monitoring and reporting tools in order to select the ones more appropriate to our needs.
  - **Tools to decide who is influential to be sent Social Messages to.** Additional tools to help the CMs

---

<sup>1</sup> The list of blog posts and their schedule is available in the second section “Dissemination plan of actions”

- (ii) **Blog at the LDBC organization web page.** The Blog will provide of enough & updated interesting contents to the CMs to elaborate the messages for the social networks.
    - **Access to the blog.** All the CMs will have access to the LDBC organisation blog.
  
  - (iii) **Templates for monitoring and reporting.** During the first two weeks of the dissemination task, the DT created the templates and information to be included in those.
  
  - (iv) **Methodology for the weekly actions.** Every week, the members of the team take the following actions in the form of a pipeline as shown in Figure 1:
    - **Meetings of the DT.** Each week there is a coordination meeting to summarise the work undertaken during the previous week and prepare the work to be done during the following week.
    - **Content.** The content generated by the CMs will be based on the before mentioned blog posts. Every week, a blog entry is selected by the CMs based on a set of priorities provided during the coordination meeting. The work of the CMs and the DT generates a set of contents used to disseminate information in the SNs.
    - **Emission.** Each author publishes the blog posts in the LDBC organisation blog. The dates for each blog entry were decided and distributed in the intranet of the project for every author to be seen. The DT will send reminders to all authors prior to their publishing date. When each post is published the CMs will:
      - **Interact with Social Networks.** The content to be published in each of the Social Networks addressed will be obtained from the blogs of the present or previous weeks. The content published in the SNs will make reference to the LDBC organization web page to attract attention to the web.
      - **Interact with response from the public.** In the event that there are questions or interaction from the public, the CMs will respond if this is clear to them.
      - **Take part of the relevant discussions** in the followed forums as well as being active in the blogs of interest, either commenting or posting in case it would be possible.
      - **Technical questions.** The answer to technical questions will be deferred to the technical staff if the questions are difficult due to their technical content.
  
  - (v) **Monitoring.** Based on the fact that LDBC needs to obtain feedback about the dissemination actions taken, the DT sets of monitoring actions undertaken weekly.
    - **Accounting.** All the messages injected in the Social Networks, as well as the different actions taken are accounted using a number of tools.
    - **Emission.** The exact blog entries and answers are included in the weekly monitoring actions document to have them ready for the reporting performed monthly.
1. **Reporting actions.** In order to have a summary of the impact of the actions taken, the DT generates a monthly report for the Monthly consortium meeting. The report contains:
- A pointer to the blogs published.
  - An account of all the actions taken on the different Social Media Platforms.

- An account of the response obtained, both in terms of direct reaction from the platforms, and in terms of the increment in the number of visits to the LDBC organisation web site.
  - A summary of the advances obtained from previous reporting periods.

	W1	W2	W3	W4	W5	W6	W7	W8
<b>Content generation</b>	Content W1	Content W2	Content W3	Content W4	Content W5	Content W6	Content W7	Content W8
<b>Emission and interaction</b>		Emission W1	Emission W2	Emission W3	Emission W4	Emission W5	Emission W6	Emission W7
<b>Accounting</b>			Monitoring	Monitoring	Monitoring	Monitoring	Monitoring	Monitoring
<b>Monthly reporting</b>				Reporting M1				Reporting M2

Figure 1. Pipeline of actions to be taken by Community Managers.

## A.2 Dissemination plan of actions

The plan created for the dissemination includes the following:

- (i) Blog Schedule. Each author from the project proposed an article to publish and the DT scheduled them.
- (ii) List & plan of social networks, forums, and websites where the dissemination will take place.

### A.2.1 Blog Schedule

Every Thursday the author or the CMs will publish a new blog entry, and until the next Thursday, their actions will focus in disseminating the new post along the social networks. The table below represents how the work is distributed every week in order to divide equally the number of posts assigned to each CM. The DT established this calendar until November, on a Monthly Consortium meeting the calendar for November to the end of the project will be created.

	Post	Community Manager	Scheduled date
1	New website online & LDBC benchmarks reach Public Draft	Iliya	06/23/14
2	Chokepoint based Benchmark Development	Iliya	07/14/14
3	Social Networks Benchmark goals	Ricard	07/24/14
4	Is SNB like Facebook’s LinkBench?	Ricard	07/31/14
5	Getting Started with SPB	Iliya	08/07/14
6	SNB Data Generator - Technical	Ricard	08/14/14
7	Getting Started with SNB	Ricard	08/21/14
8	Using SPB to Find OWLIM Performance Issues	Iliya	08/28/14
9	SNB Interactive workload	Ricard	09/04/14
10	SNB Driver	Ricard	09/11/14 delayed by the author to 09/26/14
11	SNB Data Generator - Getting Started	Ricard	09/18/14
12	Graph Analysis on SNB	Ricard	10/02/14

13	SNB BI workload	Ricard	10/09/14
14	SNB Temporal Phenomena	Ricard	10/16/14
15	Running SPB on various databases	Iliya	10/23/14
16	Industry relevance of SPB	Iliya	10/30/14
17	Graph-TA conclusions/reporting on Benchmarking talks	Ricard	11/06/14
18	SNB benchmark progress	Ricard	11/13/14

### A.2.2 List & plan of dissemination platforms

The DT team decided on the list of social networks, forums and websites where the CMs should work with the dissemination tasks and created a plan for each one of them. Following the list of social networks and the planned activities for each is provided:

#### (i) Twitter

Twitter is one of the most well-known social platforms. It allows microblogging and sharing content via images and links. It was identified that the audience of the LDBC daily shares in this social network their opinion, results and questions in this social network. Twitter is settled as the main social network to build community and where to identify interesting leaders of opinion regarding benchmarking.

Plan:

- A daily tweet will be published about something related to the blog entry of the week. The DT wants to make a daily effort in this social network because it is identified as our primary dissemination platform.
- Interaction with the users who talk about LDBC and optionally retweets about related information when appropriate.
- CMs will constantly try to grow the number of followers by following, retweeting and declaring its favorites.

#### (ii) Facebook

Facebook is the biggest social platform. Although its orientation is of a more personal and leisure type it was identified as relevant to build LDBC's own page to share our activity because researchers may find their niche there.

Plan:

- A weekly update will be published announcing our new blog entry. Depending on the feedback, our activity could be increased or decreased.
- The updates will be shared through a personal profile in the relevant discussion groups.

#### (iii) LinkedIn

LinkedIn is one of the biggest networking sites which allows sharing, commenting and searching for professional information. Developers will find information about the LDBC benchmarks in our discussion group. It is still under the DT consideration to create an organization page instead of sharing the information in the discussion group.

Plan:

- A weekly update will be published announcing our new blog entry. Depending on the feedback, our activity could be increased or decreased.
- The updates will be shared through a personal profile in the relevant discussion groups.

#### (iv) Google+

Google+ is a social network similar to Facebook but with an orientation to share different information regarding the social circle it is directed. Google+ has been proved as one good tool to improve SEO. LDBC would be benefited by increasing the SEO of its website and blog.

Plan:

- At least a weekly update will be published announcing our new blog entry. Depending on the feedback, our activity could be increased or decreased.
- The updates will be shared through a personal profile in the relevant discussion groups.

#### (v) Other Publications

Apart from the blog and microblogging in our social networks accounts & groups, the CMs will look out for the opportunity to publish on the following channels:

Wikipedia: LDBC will try to publish an article explaining what is the LDBC council and its mission.

Dzone: CMs will share the link of each new entry on our blog and consider publishing there the more research relevant posts instead of using our own blog.

ISTC Big Data blog: Let them know about the existence and mission of the LDBC and propose to write a post for their blog in case they may be interested. (<http://istc-bigdata.org/index.php/category/istc-blog/>)

NOSQL My popescu website: Let them know about our blog so they can share our content in case they may be interested (<http://nosql.mypopescu.com/>)

W3 org website: Add LDBC to the benchmarks' list (<http://www.w3.org/wiki/RdfStoreBenchmarking>)

## A.3 Weekly dissemination efforts

The CMs elaborate each week the specific messages that will be sent the next week for all the social networks in a document which is uploaded to the intranet for the review of the DT. The messages are constructed with the information available from the blog posts. Every author is in charge of providing the article to be published at least one week before its publication. We have noticed that for several reasons some publications have been delayed thus affecting the work flow of the CMs. For the next term we are working on a contingent plan that will avoid stopping the dissemination task in the social networks when one incident occurs in the blog posts scheduling.

### A.3.1 Publications from 07/10/14 to 07/16/14

CM assigned: Iliya Stefanov

Blog post: "Chokepoint based Benchmark Development"

#### Twitter

Thursday: *not provided by CM*

Friday: #LDBC the soon-to-be-standard #benchmark for #Graph and #RDF has its own new site <http://ldbcouncil.org/>

Monday: *not provided by CM*

Tuesday: Sometimes less is more, thus we strive to limit the number of queries in the LDBC benchmarks to what's essential <http://ow.ly/ziAP6>

Wednesday: *not provided by CM*

Extra messages: LDBC aims at #BigData Graph problems <http://ldbcouncil.org/public/why-graph>

### Facebook

“Here are the most critical points in database benchmark design captured by LDBC <http://ldbcouncil.org/blog/choke-point-based-benchmark-design>”

### LinkedIn

“The most critical points for database systems captured in LDBC benchmark design <http://ldbcouncil.org/blog/choke-point-based-benchmark-design>”

Shared in the following groups:

<https://www.linkedin.com/groups/OWL-RDF-119102>

<http://www.linkedin.com/groups/Semantic-Web-49970>

### Google+

“Weekly update publication: We are extremely proud to announce you our new LDBC website <http://ldbcouncil.org/>. We aim to produce industry standards for benchmarking RDF and Graph technologies. We are looking to reach to as many interested in the field as possible. Check out what we've done so far and share your opinion and questions on our forum <http://ldbcouncil.org/forum>”

### Forums

The new blog post was also announced in the following forums/Q&A pages:

[https://groups.google.com/forum/#!forum/semantic\\_web](https://groups.google.com/forum/#!forum/semantic_web)

[Quora](#)

## A.3.2 Publications from 07/24/14 to 07/30/14

CM assigned: Ricard Tapias

Blog post: “Social Network Benchmark goals”

### Twitter

Thursday: One of the main objectives of LDBC is to shape an up-to-date [#SocialNetwork #Benchmark](#) (SNB). Read more about it here [ow.ly/1234567](http://ow.ly/1234567)

Friday: The [#SocialNetwork #Benchmark](#) is going to include Interactive, Business Intelligence and Analytical workloads [ow.ly/1234567](http://ow.ly/1234567)

Monday: The SNB uses a powerful data generation tool that mimics the data managed by real [#SocialNetworks](#) [ow.ly/1234567](http://ow.ly/1234567) [#BigData](#)

Tuesday: [#SocialNetworks](#) are technologically challenging, thus an up-to-date Benchmark with significant challenges is needed [ow.ly/1234567](http://ow.ly/1234567)

Wednesday: You can find the [#SocialNetwork](#) Benchmark data generator created by LDBC here [ow.ly/1234567](http://ow.ly/1234567) [#BigData](#)

### Facebook

“Being part of the Big Data arena, Social Networks are technologically challenging for the data management solutions in the market. With the objective of shaping an up-to-date benchmark that poses significant technological challenges, the LDBC consortium decided to create the Social Network Benchmark (SNB) with three different workloads: Interactive, the Business Intelligence and the Analytical use cases & which includes a powerful synthetic data generation tool. Check out our new blog entry to read more about it”

### LinkedIn

“Weekly update publication: Being part of the Big Data arena, Social Networks are technologically challenging for the data management solutions in the market. With the objective of shaping an up-to-date benchmark that poses significant technological challenges, the LDBC consortium decided to create the Social Network Benchmark (SNB) with three different workloads: Interactive, the Business Intelligence and the Analytical use cases & which includes a powerful synthetic data generation tool. Learn more about it reading our new blog post”

Shared in the following groups: *Social Network Analysis in Practice*, *Social, Media Research*, *Analysis and Insight*, *Graph Databases*

### Google+

“ Being part of the Big Data arena, Social Networks are technologically challenging for the data management solutions in the market. With the objective of shaping an up-to-date benchmark that poses significant technological challenges, the LDBC consortium decided to create the Social Network Benchmark (SNB) with three different workloads: Interactive, the Business Intelligence and the Analytical use cases & which includes a powerful synthetic data generation tool. Read more about it our new blog entry”

### Forums

The new blog post was also announced in the following forums/Q&A pages: DZone.

## 1.3. Publications from 07/31/14 to 08/06/14

CM assigned: Ricard Tapias

Blog post: “Is SNB like Facebook’s Linkbench?”

## Twitter

Thursday: Why is [#SocialNetwork #Benchmark](#) not like [#Facebook](#)’s [#LinkBench](#)? Check out our new blog post to find it out [ow.ly/1234](#)

Friday: For end users facing graph processing tasks, SNB could be used to compare merits of different products and technologies.

Monday: For vendors of [#GraphDB](#) tech SNB provides a checklist of features and performance characteristics that can serve to guide new development

For researchers SNB [#enchmark](#) provides interesting challenges: query optimization, (distributed) graph analysis & transactional throughput

Tuesday: The [#SocialNetwork #Benchmark](#) interactive workload tests a system’s throughput with relatively simple queries & concurrent updates. // The SNB interactive workload is waiting your feedback, the data generator & driver software stack are ready here [ow.ly/1234](#)

Wednesday: The [#BusinessIntelligence](#) workload consists of complex structured queries to analyze online behavior of users for marketing [#SNB](#)

*Extra:*

The [#graphanalytics](#) workload under development will consist of algorithms like PageRank, Clustering and Breadth First Search [#SNB](#)

[#SNB](#) data generator works for all workloads while [#LinkBench](#) manages the OLTPs on its friends graph with a close tuning to low-level mysql query patterns.

## Facebook

“Check out our new blog post discussing in some detail the rationale and goals of the design of the Social Network Benchmark (SNB) and explain how it relates to real social network data, and in particular how it differs in the type of data generation from Facebook’s own graph benchmark called LinkBench making it in contrast suitable for more workloads”

## LinkedIn

“Check out our new blog entry where we discuss in some detail the rationale and goals of the design of the Social Network Benchmark (SNB) and explain how it relates to real social network data and in particular, how it differs from Facebook’s own graph benchmark called LinkBench.

The benchmark includes three different workloads regarding the type of queries: the Interactive, the Business Intelligence and the Analytical. Each of the workloads produces a single metric for performance at the given scale and a price/performance metric at the scale.

SNB is addressed to end users facing graph-processing tasks and as an aid to technology selection, to vendors providing a checklist of features and performance characteristics for product positioning and for

researchers providing interesting challenges to objectively compare the effectiveness and efficiency of new and existing technology in these fields”

### **Google+**

“New blog post: Is SNB like Facebook’s LinkBench? We discuss in some detail the rationale and goals of the design of the Social Network Benchmark (SNB) and explain how it relates to real social network data and how it differs from Facebook’s own graph benchmark called LinkBench”

### **Forums**

The new blog post was also announced in the following forums/Q&A pages: DZone.

## **A.3.3 Publications from 08/07/14 to 08/13/14**

CM assigned: Iliya Stefanov

Name of the post: “Getting Started with the Semantic Publishing Benchmark”

### **Twitter:**

Thursday: How to get started with our Semantic Publishing Benchmark #LinkedData #SemanticWeb

Friday: The Semantic Publishing Benchmark covers nearly all features of #SPARQL 1.1 (link to the article)

Monday: Some interesting choke points of SPARQL query optimisation are regarded in SPB.

Tuesday: Check out how #RDF Stores deal with nested optional clauses, subqueries, full-text search and other advanced #SPARQL features <http://ow.ly/AuF5b>

Wednesday: #SPB is distributed as a jar file and provides a reference dataset based on BBC ontologies <http://ow.ly/A3fuA>

### **Facebook:**

“The Semantic Publishing Benchmark is aimed at measuring the performance of CRUD operations of SPARQL endpoints. It provides 9 queries covering most advanced features of SPARQL 1.1, such as subqueries, regular expressions, limit, sorting. The queries are particularly aimed at choke points related to query optimization. You can have a look at the example query and results in our new blog post - Getting Started with SPB (link)”

### **LinkedIn:**

“New Blog Post: Getting Started with SPB

The Semantic Publishing Benchmark is aimed at measuring the performance of CRUD operations of SPARQL endpoints. It provides 9 queries covering most advanced features of SPARQL 1.1, and is aimed at choke points relevant to query optimisation. You can find more information in our new blog post (link)”

**Google +**

“New Blog Post: Getting Started with SPB

The Semantic Publishing Benchmark is aimed at measuring the performance of CRUD operations of SPARQL endpoints. It provides 9 queries covering most advanced features of SPARQL 1.1, and is aimed at choke points relevant to query optimisation. You can find more information in our new blog post (link).”

**A.3.4 Publications from 08/14/14 to 08/20/14**

CM assigned: Ricard Tapias

Name of the post: “DATAGEN: The SNB data generator”

**Twitter**

Thursday: Read our new blog post to know which are the main features of [#DATAGEN](#), our [#SocialNetwork](#) data generator [Ow.ly/1234](#)

Friday: [#DATAGEN](#) has been carefully designed with the following goals in mind: realism, scalability, determinism and usability. [#SNB](#)

Monday: [#DATAGEN](#) is deterministic regardless of the num cores/machines used to produce data, making [#SNB](#) reproducible & fair: [ow.ly/1234](#)

Tuesday: [#DATAGEN](#) is implemented following the MapReduce paradigm, allowing for the generation of large datasets on commodity clusters [#SNB](#)

Wednesday: In [#DATAGEN](#), output attributes, cardinalities, correlations and distributions have been finely tuned to reproduce real [#SocialNetworks](#)

Extra: Our data generator could be used in any research or evaluation it is not restricted to LDBC’s SN Benchmark. You can download it from: [ow.ly/1234](#)

**Facebook**

“Take a look to our new blog entry discussing the main features of LDBC’s Social Network Benchmark (SNB) data generator DATAGEN. One of the most important components of a benchmark is the dataset, however, directly using real data in a benchmark is not always possible. For these reasons, LDBC-SNB provides DATAGEN which is the synthetic data generator responsible for generating the datasets for the three LDBC-SNB workloads: the Interactive, the Business Intelligence and the Analytical.”

**LinkedIn**

“Which are the main features of DATAGEN, the LDBC-SNB data generator?”

One of the most important components of a benchmark is the dataset, however, directly using real data in a benchmark is not always possible. For these reasons, LDBC-SNB provides DATAGEN which is the synthetic data generator responsible for generating the datasets for the three LDBC-SNB workloads: the Interactive, the Business Intelligence and the Analytical.

Its four main characteristics are:

- Reality: Provides all the elements to mimic a real social network with output attributes, cardinalities, correlations and distributions.
- Scalability: allows generating datasets of different sizes, from a few Gigabytes to Terabytes.
- Determinism: making the comparisons between different systems fair and the benchmarks' results reproducible
- Usability

Read our new blog entry to know more about DATAGEN and its main features”

### Google+

“Check out our new blog post discussing the main features of LDBC’s Social Network Benchmark (SNB) data generator DATAGEN. One of the most important components of a benchmark is the dataset, however, directly using real data in a benchmark is not always possible. For these reasons, LDBC-SNB provides DATAGEN which is the synthetic data generator responsible for generating the datasets for the three LDBC-SNB workloads: the Interactive, the Business Intelligence and the Analytical”

### Forums

The new blog post was also announced in the following forums/Q&A pages: DZone.

## A.3.5 Publications from 08/27/14 to 09/03/14

CM assigned: Iliya Stefanov

### Blog post: Using SPB to Find OWLIM Performance Issues

*Note: This post was published in advance, couple of weeks before it was due. For this reason, the dissemination was disturbed. We have published one message on twitter in the week when the post was due and another posts on Facebook and LinkedIn in the week of 22 – 29 September*

### Twitter:

Day 1 : “Using LDBC-SPB to Find OWLIM Performance Issues”

### Facebook:

“The LDBC Semantic Publishing Benchmark used in the development and release process of one of the leading RDF Stores - GraphDB (formerly OWLIM) <http://ldbouncil.org/blog/getting-started-semantic-publishing-benchmark>”

### LinkedIn:

“ The LDBC Semantic Publishing Benchmark used in the development and release process of one of the leading RDF Stores - GraphDB (formerly OWLIM) <http://ow.ly/BNPKL>

Graph Databases Group: The LDBC Semantic Publishing Benchmark used in the development and release process of OWLIM (now GraphDB) <http://ldbouncil.org/blog/using-ldbc-spb-find-owlim-performance-issues>”

### A.3.6 Publications from 08/27/14 to 09/03/14

*Note: For reasons beyond CMs control this post was published in advance only one day after de previous post. For this reason, the messages promoting this post where published 2 weeks later. We haven't published on other social networks except Twitter (LinkedIn, Facebook and Google +) because another post was published one day before and it would have been unproductive.*

CM assigned: Ricard Tapias

Blog post: "Social Network Benchmark goals"

#### Twitter

Day 1: Getting started with #SNB: <http://ow.ly/ABmzC> Click to know more about its essential forming parts. #Benchmark #SocialNetworks

Day 2: Take a look to the schema of the datasets produced by #DATAGEN, in terms of entities and their relations. #SNB

Day 3: The LDBC execution driver automatically generates the benchmark workload and gathers the benchmark results <http://ow.ly/Bv3Sx> #SNB

Day 4: The #SocialNetworks #Benchmark Analytical workload will aim at exploring the characteristics of the underlying structure of the network

Day 5: For a detailed description of the #SNB #Benchmark do not hesitate to read the official SNB specification draft: <http://ow.ly/BDiXU>

### A.3.7 Publications from 09/04/14 to 09/10/14

CM assigned: Ricard Tapias

Blog post: *Introducing SNB Interactive, the LDBC Social Network Benchmark online workload*

#### Twitter

Thursday: New blog post: Introducing #SNB Interactive, the LDBC Social Network #Benchmark Online Workload [ow.ly/1234](http://ow.ly/1234) #SocialNetworks

Friday: The #SNB Interactive workload measures the speed of queries of medium complexity against a Social Network being constantly updated.

Monday: #SNB Interactive is an example of LDBC's choke point driven design methodology. Read more about it in our blog post: [ow.ly/1234](http://ow.ly/1234)

Tuesday: #SNB Interactive gives end users and developers a reference workload to compare the merits of different graph data management technologies

Wednesday: The #SNB Interactive Workload is defined in natural language with sample implementations in #SPARQL and #cypher.

## Facebook

“New entry on the blog! The LDBC Social Network Benchmark (SNB) is composed of three distinct workloads, interactive, business intelligence and graph analytics. Check our new blog post introducing the Interactive Workload”

## LinkedIn

“Check out our new blog post introducing the Social Network Benchmark (SNB) Interactive, one of the three distinct workloads composing SNB, along with business intelligence and graph analysis.”

## Google+

“Check out our new blog post introducing the Interactive Workload. The LDBC Social Network Benchmark (SNB) is composed of three distinct workloads, interactive, business intelligence and graph analytics. Read more about it on the blog!”

## Forums

The new blog post was also announced in the following forums/Q&A pages: DZone.

## A.4 Monitoring

Monitoring the relevant sites for the LDBC community is also another important task for the CMs. For any of the identified channels when there are any active topics related with benchmarking, graph DBs or RDF, the CMs should be active on them replying to comments or sharing their relevant publications on our social networks. Google Alerts, RSS feed aggregators or any of the tools listed in the “Monitoring, tracking and management tools” section will be used as monitoring tools.

The monthly reports with all the results tracking visits on the website and engagement in social networks are also included in this section.

### A.4.1 Monitoring Tracking and Management Tools State of the art

From the first list of tools proposed by the Dissemination Team, the CMs selected 6 tools to use in a regular basis to manage, monitor, track, social media channels and social networks activity. These are described in the table below:

Tool	Features	Pros	Cons	Chosen pricing plan
<i>Paper.li</i>	Content curation tool. Monitor twitter users, google+ accounts, YouTube channels and RSS alerts. Search for keywords on tweets and Facebook updates.	Easy way to have a general vision of what is going on in the social networks in relation with a chosen topic.	Monitoring Facebook pages or profiles is not possible. No analytics nor statistics. The criteria followed by the system to select the relevant content is unknown. The paper created is public.	FREE

Tool	Features	Pros	Cons	Chosen pricing plan
<i>Hootsuite</i>	Web and mobile social media management system. It Integrates Twitter, Facebook, LinkedIn, Google+, Foursquare, Myspace, WordPress, TrendSpottr and Mixi accounts. Browser-based dashboard that allows the user to monitor mentions, send and schedule tweets, monitor feeds on Facebook, keyword tracking, monitor industry conversations on LinkedIn, schedule content of WordPress accounts.	Very powerful management tool to schedule the emission of content and to synchronize accounts from different social networks. It also offers analytics and stats.	Only few tracking and monitoring features.	PRO (7.19€/month): includes interesting features such as unlimited RSS monitoring, more reporting options, advanced message scheduling and the possibility to manage the accounts as a team and not as a single user.
<i>Google alerts</i>	Content detection and notification service. The user receives a notification every time there's new content regarding the chosen keywords. It searches for updates in webpages, blogs, news, groups and videos.	Easy way to have a panorama of what is going on in the web in regarding a chosen topic.	No social network tracking/management.	FREE
<i>Twazzup</i>	Keywords and hashtags real-time monitoring. It offers information about top influencers users, top RT links, top RT links and google new alerts.	Useful to search for most followed users for a certain topic and most recent users talking about the topic.	Only twitter monitoring service, no stats nor analytics provided.	FREE
<i>Google analytics</i>	Web stats and analytics tool for websites. Regarding social media, it allows you to analyze how visitors interact with sharing features on your site (like the Google +1 or the Like Facebook button) and engage with your content across social platforms. It also offers a way to track your online campaigns such as monthly newsletters or web ads to evaluate how many traffic are they bringing to a certain URL.	Analytics of the interaction between social networks accounts and the website. Campaign tracking.	Most of the features are focused on web analysis more than social networking analysis.	FREE
<i>Hashtagify.me</i>	A tool to know the popularity of a hashtag, other hashtags related to the chosen one, their evolution in terms of popularity over time and the users that use it the most.	Accurate information related with hashtags and its use among users.	-	FREE

## A.4.2 Monthly reports

To the date of this deliverable the DT has completed two monthly reports which are included following. The reports include the engagement and variables analyzed for each of the social networks and then a

- |    |                |    |                  |    |            |  |
|----|----------------|----|------------------|----|------------|--|
| 1. | <b>COUNTRY</b> | 2. | <b>Followers</b> | 3. | <b>GMT</b> | complete detail of the most relevant variables in the website tracking using Google Analytics. |
|    |                |    | (>2%)            |    |            |  |

## A.4.3 Monthly report 07/11/2014 – 08/11/2014

One of the CMs tasks it has been to monitor LDBC's social media channels with the aim of elaborating detailed reports in order to get information about the effects of the work performed by the DT and the feedback of our audience. The reports analyse a series of variables, which were identified during the first stage of the dissemination strategy planning and agreed with all the members of the DT.

### Twitter

Apart from the actions described in the "Dissemination actions plan" and the "Weekly actions" sections, the CMs also used other tools to reach a broader audience. In first place, we used *Followerwonk* and the twitter search engine to find and follow relevant users that could be interested in LDBC so that they know about us. In second place, we used *Justunfollow* to track our followers and build a profile with a balanced followers/followed ratio. As a result, we gained 25 followers in a month, which means a 15.1% growth. This is a notable step forward given the low Twitter activity in August involving research and business related users.

### Overview:

**FOLLOWERS: 190**      **FOLLOWING: 98**      **FOLLOWER'S GROWTH: 15.1%**

### Engagement:

A larger number of followers can be pointless if it does not go along with a real repercussion in terms of community engagement. Having our tweets retweeted and being mentioned by other users is what makes us reach a broader audience. For the moment, LDBC's twitter profile has had a notable repercussion (51 retweets and 20 mentions in one month) given the amount of people following it.

**RETWEETS: 51**      **MENTIONS: 20**

### Followers by Region

:



4.	USA	5.	25.4%	6.	-4 to -10
7.	UK	8.	16.7%	9.	+1
10.	Germany	11.	11.1%	12.	+2
13.	Spain	14.	8.7%	15.	+2
16.	France	17.	6.3%	18.	+2
19.	Netherlands	20.	6.3%	21.	+2
22.	Austria	23.	4%	24.	+2
25.	Bulgaria	26.	2.4%	27.	+3
28.	Denmark	29.	2.4%	30.	+2

Considering the geographic distribution of our followers, it would be convenient to schedule some of our tweets so that they are published at a reasonable hour in the USA. Publishing messages when your followers are active increases the probability of your tweets to be seen before they sink in the users news feed.

### On Twitter:

This is a record of the most clicked links shared on Twitter. One should bear in mind that the number of clicks on the links is not the same as the number of views of the post. Also, a high number of clicks does not necessarily only mean better content, because factors like the way the tweet has been written and the moment it has been published have to be considered. For the moment we don't have any detailed stats of the blog posts views, deeper analytics will be included on the next deliverable.

#### 1. LDBC blog post: Choke-point based benchmark design

Created: Jul 18, 2014 6:52 AM

Total Clicks: **15**

#### 2. LDBC blog post: Social Network Benchmark Goals

Created: Jul 24, 2014 2:26 PM

Total Clicks: **11**

#### 3. *Database Architects* blog post: The LDBC Social Network Benchmark

31. Created: Aug 11, 2014 7:56 AM

32. Total Clicks: **10**

#### 4. Original paper: Foundations of an Alternative Approach to Reification in RDF

Created: Jul 29, 2014 10:14 AM

Total Clicks: **8**

#### 5. LDBC blog post: Why is SNB not like Facebook's Linkbench

Created: Jul 31, 2014 9:10 AM

Total Clicks: **6**

#### 6. *Sparsity technologies* blog post: Synthetic graph data generators

Created: Jul 31, 2014 9:32 AM

Total Clicks: **6**

**7. LDBC Official repository:** [https://github.com/ldbc/ldbc\\_snb\\_datagen](https://github.com/ldbc/ldbc_snb_datagen)

33. Created: Jul 29, 2014 8:39 AM

Total Clicks: **6****Facebook**

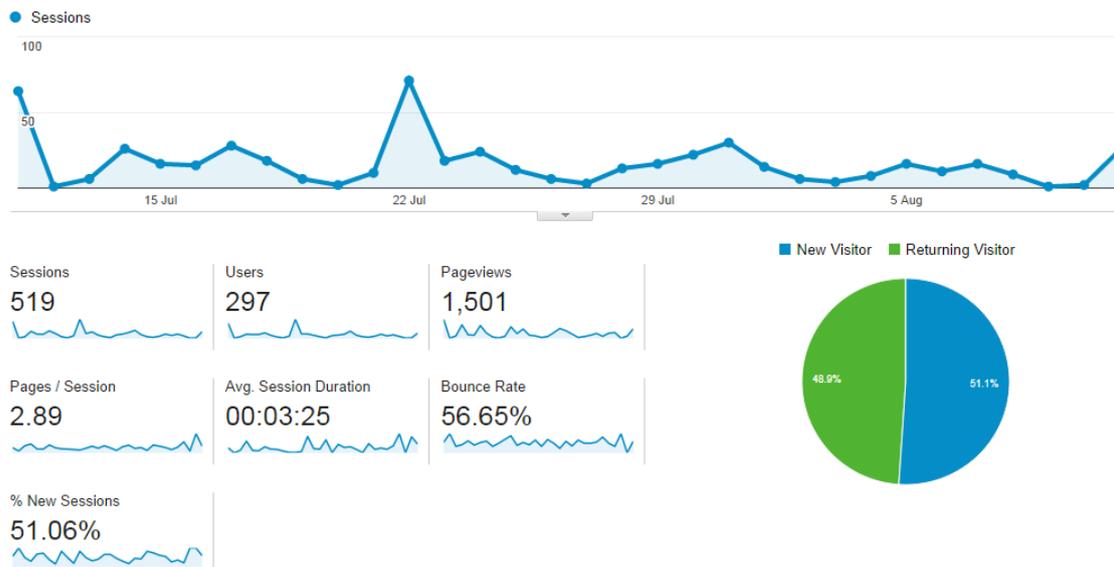
As it has been stated, Facebook is not a priority for the scientific dissemination in the field of database benchmarking, but it is another tool that we have to help and support our actions in other social media channels like LDBC's blog. This is why we didn't expect a lot of activity as it has actually been. Here is a summary of the activity during the first month of dissemination:

**Overview**TOTAL LIKES: **78**LIKES GROWTH: **6.8%**PAGE VIEWS: **50**WEEKLY REACH: **29****Engagement**POST LIKES: **5**POST COMMENTS: **5**LIKES BY REGION (top 5): **Greece (21), USA (11), UK (7), Austria (6), Spain (6).****LinkedIn discussion group**

Unlike Facebook, LinkedIn certainly is an interesting social network for our purposes. Our first decision was to create a debate page, so that people could interact more actively with the topics proposed in our blog posts, but as the time goes by, we see that the fact that we are the only ones posting debates could be against us. That's why depending on the stats of the next month we could consider changing the debate page for a corporate page.

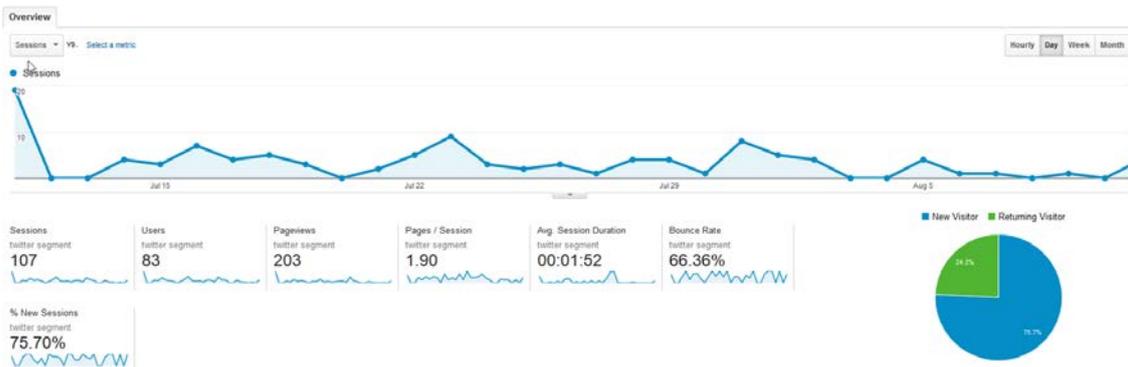
**Overview**PARTICIPANTS: **55** NEW DEBATES (blog posts shared): **3****Website****Overview**

The number of sessions during this month has been 519, with 297 individual users visiting the website. The average session duration is 3:25 minutes, which is a good mark to start with. The rate of returning visitors is nearly the same as the new visitors, which is something one could expect from a new website that is still in the process of being disclosed.



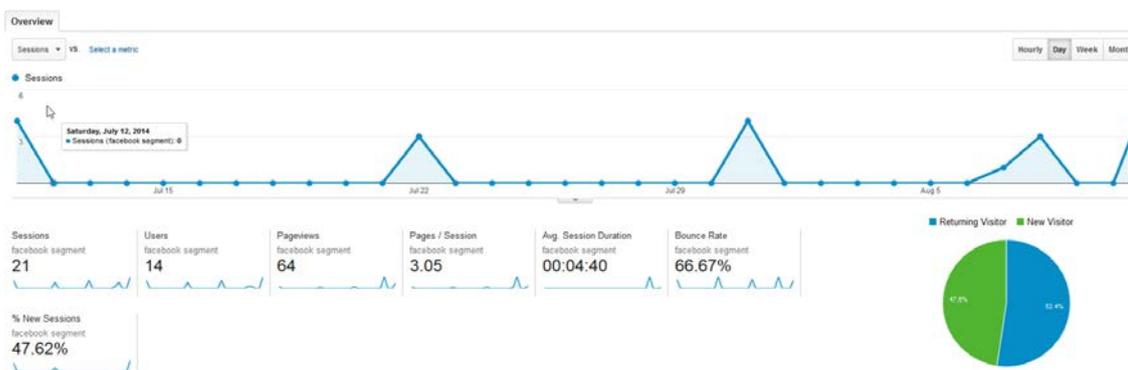
### Twitter traffic

Twitter is responsible for a big part of the traffic to ldbcouncil.org – 20 %. It is clear that it brings a lot of new visitors, but the bounce rate is also relatively high. During this month there are 83 users which were directed to the site from twitter.



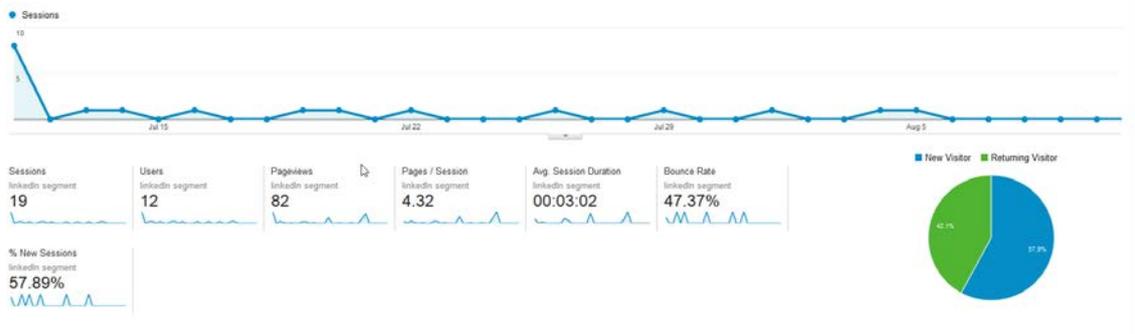
### Facebook

The traffic generated from Facebook is considerably smaller than from Twitter. This is probably because there are less posts on Facebook and also there is less interaction there. It is still positive that towards the end of the period the traffic seems to grow.



### LinkedIn Traffic

Linked in has attracted 12 different users during this month. The amount of traffic is rather small, even smaller than from Facebook, but the bounce rate is less than 50 % and average pages per session are 4.32, which shows that users attracted through LinkedIn have a more serious interest towards the LDBC content. It is therefore necessary to focus more on posting information on LinkedIn.



### Segmentation by referrals

It is clear that apart from direct traffic, Twitter generates the biggest part of the traffic. Traffic from search is relatively low which means we should optimize our content to be more easily accessible from direct search.

Source / Medium	Acquisition			Behavior		
	Sessions	% New Sessions	New Users	Bounce Rate	Pages / Session	Avg. Session Duration
	519 <small>% of Total: 100.00% (519)</small>	51.06% <small>Site Avg: 51.06% (0.00%)</small>	265 <small>% of Total: 100.00% (265)</small>	56.65% <small>Site Avg: 56.65% (0.00%)</small>	2.89 <small>Site Avg: 2.89 (0.00%)</small>	00:03:25 <small>Site Avg: 00:03:25 (0.00%)</small>
1. (direct) / (none)	222 (42.77%)	43.24%	96 (36.23%)	58.56%	3.02	00:04:05
2. t.co / referral	77 (14.84%)	66.23%	51 (19.25%)	54.55%	2.23	00:02:31
3. ldbccouncil.sti2.at / referral	51 (9.83%)	1.96%	1 (0.38%)	27.45%	4.35	00:05:08
4. ldbccouncil.org / referral	33 (6.36%)	57.58%	19 (7.17%)	60.61%	2.61	00:03:03
5. semalt.semalt.com / referral	20 (3.85%)	100.00%	20 (7.55%)	100.00%	1.00	00:00:00
6. ldbc.eu / referral	19 (3.66%)	68.42%	13 (4.91%)	31.58%	3.05	00:03:19
7. facebook.com / referral	17 (3.28%)	35.29%	6 (2.26%)	58.82%	3.53	00:05:45
8. linkedin.com / referral	16 (3.08%)	62.50%	10 (3.77%)	43.75%	4.75	00:03:32
9. google / organic	15 (2.89%)	66.67%	10 (3.77%)	46.67%	3.00	00:01:18
10. databasearchitects.blogspot.de / referral	10 (1.93%)	70.00%	7 (2.64%)	80.00%	1.50	00:00:06



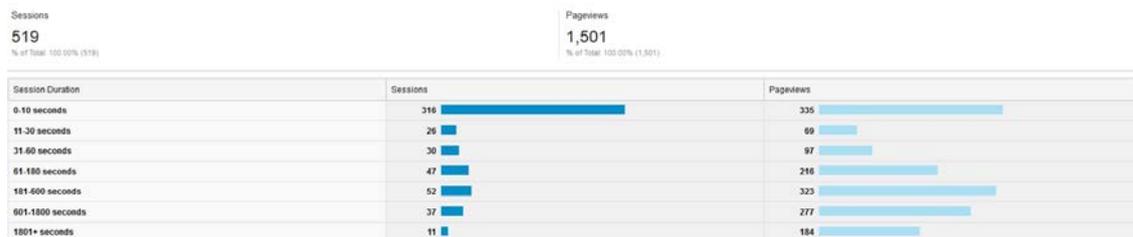
### Segmentation by location

As it happens in the Twitter account, the majority of the users visiting our site are from the United States. The rest of the list is fairly similar than the Twitter followers by location list.

Country / Territory	Acquisition			Behavior		
	Sessions	% New Sessions	New Users	Bounce Rate	Pages / Session	Avg. Session Duration
	519 <small>% of Total: 100.00% (519)</small>	51.06% <small>Site Avg: 51.06% (0.00%)</small>	265 <small>% of Total: 100.00% (265)</small>	56.65% <small>Site Avg: 56.65% (0.00%)</small>	2.89 <small>Site Avg: 2.89 (0.00%)</small>	00:03:25 <small>Site Avg: 00:03:25 (0.00%)</small>
1. United States	113 (21.77%)	86.73%	98 (36.98%)	76.99%	1.81	00:01:00
2. Germany	85 (16.38%)	20.00%	17 (6.42%)	38.82%	4.51	00:05:40
3. Spain	65 (12.52%)	18.46%	12 (4.53%)	32.31%	3.05	00:05:12
4. Austria	29 (5.59%)	20.69%	6 (2.26%)	37.93%	5.52	00:07:12
5. Switzerland	28 (5.39%)	3.57%	1 (0.38%)	42.86%	3.21	00:05:01
6. Netherlands	24 (4.62%)	20.83%	5 (1.89%)	50.00%	3.75	00:05:19
7. Brazil	16 (3.08%)	100.00%	16 (6.04%)	100.00%	1.00	00:00:00
8. United Kingdom	16 (3.08%)	62.50%	10 (3.77%)	43.75%	3.00	00:04:10
9. France	15 (2.89%)	86.67%	13 (4.91%)	73.33%	1.87	00:01:02
10. Greece	14 (2.70%)	50.00%	7 (2.64%)	64.29%	3.57	00:06:53

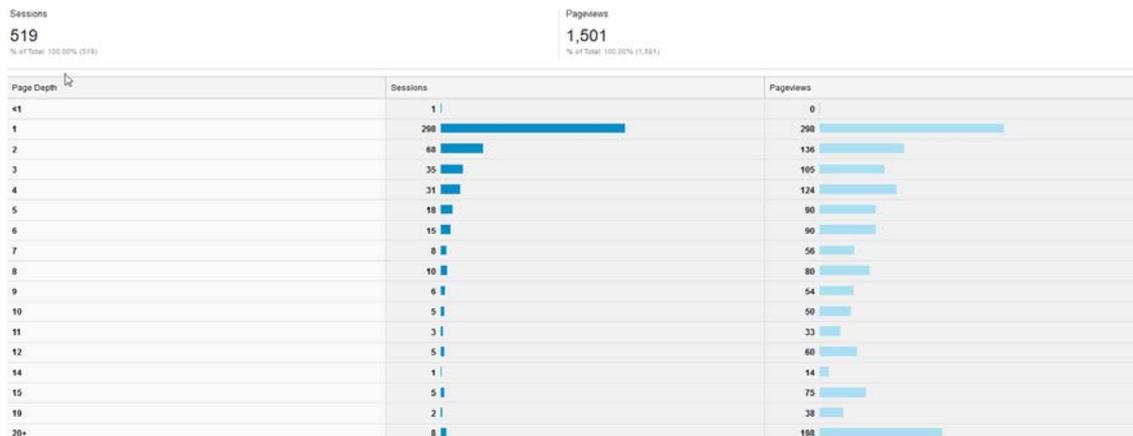
### Segmentation by session duration

In the table below we can see that the number of sessions with a duration greater than 10 seconds (203 sessions) is about a 2/3 of the number of sessions of less than 10 seconds (316 sessions).



### Segmentation by page depth

The table below presents the sessions and page views by depth segmentation, meaning the number of pages visited by the users.



### Segmentation by page

In the table below we can see how the most visited content in our website is the home page and the blog articles, which means that our content is interesting for the users and that the promotion of the posts on the social networks is working.

Page	Pageviews	Unique Pageviews	Avg. Time on Page	Entrances	Bounce Rate	% Exit
	1,501 % of Total: 100.00% (1,501)	1,053 % of Total: 100.00% (1,053)	00:01:48 Site Avg: 00:01:48 (0.00%)	518 % of Total: 100.00% (518)	56.65% Site Avg: 56.93% (0.00%)	34.51% Site Avg: 34.51% (0.00%)
1. /	522 (34.78%)	375 (35.61%)	00:02:11	335 (64.67%)	54.33%	46.36%
2. /blog/choke-point-based-benchmark-design	113 (7.52%)	71 (6.74%)	00:01:37	37 (7.14%)	70.27%	38.94%
3. /blog/snb-facebooks-linkbench	73 (4.86%)	49 (4.65%)	00:02:21	21 (4.05%)	71.43%	38.36%
4. /benchmarks/snb	65 (4.33%)	49 (4.65%)	00:01:18	15 (2.90%)	73.33%	36.92%
5. /benchmarks	62 (4.13%)	40 (3.80%)	00:00:40	1 (0.19%)	0.00%	9.68%
6. /public	54 (3.60%)	29 (2.75%)	00:01:13	5 (0.97%)	60.00%	18.52%
7. /blog/new-website-online-ldbc-benchmarks-reach-public-draft	48 (3.20%)	38 (3.61%)	00:01:40	25 (4.82%)	64.00%	41.67%
8. /developer	48 (3.20%)	32 (3.04%)	00:01:22	4 (0.77%)	0.00%	12.50%
9. /developer/snb	47 (3.13%)	33 (3.13%)	00:01:04	11 (2.12%)	58.33%	44.68%
10. /industry	38 (2.53%)	27 (2.58%)	00:00:28	2 (0.39%)	0.00%	5.26%

**A.4.4 Monthly report 08/11/2014 – 09/11/2014**

**Twitter**

We have been proceeding with the actions commented in the previous monthly report, that is a maintenance plan based in building a balanced followers/followed ratio. The outcome seems to be positive given that we have improved our monthly growth (from 15.1% to 16.8%).

**Overview:**

**FOLLOWERS: 208      FOLLOWING: 178**

**FOLLOWER’S GROWTH: 16.85%**

**Engagement:**

The number of retweets has decreased from 51 to 21 and the number of mentions from 20 to 5. This could be explained by the negative effect of summer vacations on the Twitter activity of industry and research-related. It also could have affected the fact that there has been no discussions during the last month regarding our blog posts or tweets.

**RETWEETS: 20      MENTIONS: 20**



*Approximate graphic by twittercounter.com from 08/12/14 to 09/11/14*

**Most Popular Links shared on Twitter**

The number of clicks per shared link has increased significantly. During the first month of dissemination we had a maximum of 15 clicks in one link, and this month the number has been doubled (30 clicks for the most clicked link). We also note that the links regarding SPB blog posts are more appealing for the users than the ones referred to SNB. Anyway, this is was not true during the first month of dissemination

actions, when the most clicked posts were referred to information regarding the Social Network Benchmark.

**1. Blog post: Getting Started with the Semantic Publishing Benchmark**

Shared: Aug 15, 2014 7:44 AM and Aug 18, 2014 5:50 AM

Total clicks: **16 + 14**

**2. Blog post: Getting Started with the Semantic Publishing Benchmark**

Shared: Aug 19, 2014 4:39 PM

Total clicks: **18**

**3. Blog post: Introducing SNB Interactive, the LDBC social network online workload**

Shared: Sep 04, 2014 11:13 AM and Sep 04, 2014 11:13 AM

Total clicks: **8 + 9**

**4. Using LDBC-SPB to Find OWLIM Performance Issues**

Shared: Aug 29, 2014 5:52 AM

Total clicks: **9**

**5. SPB distribution (GitHub)**

Shared: Aug 07, 2014 7:14 AM

Total clicks: **3**

**6. Blog post: DATAGEN: data generation for the Social Network Benchmark**

Shared: Aug 19, 2014 11:10 AM

Total clicks: **2**

**7. Datagen (GitHub)**

Shared: Aug 25, 2014 9:29 AM

Total clicks: **2**

**8. Blog post: Getting started with SNB**

Shared: Aug 22, 2014 9:14 AM

Total clicks: **0**

## Facebook

The page views have slightly decreased (50 to 43) and we gained 2 followers but we are reaching less people weekly (from 29 to 18.5).

### Overview

TOTAL LIKES: **80**                      LIKES GROWTH: **2.56%**

PAGE VIEWS: **43**

WEEKLY REACH: **18.5**

### Engagement

POST LIKES: **4**                                      POST COMMENTS: **0**

LIKES BY REGION (top 5): **Greece (20), USA (11), UK (7), Austria (7), Spain (6).**

### LinkedIn

From 55 to 57 group members. Our debates have a low impact, we have some views and clicks but they don't generate engagement (no recommendations nor comments). For this reason plus what was suggested on the previous report, the CMs recommend to open a corporate LinkedIn to avoid making a use of the group that could be perceived as spam for some users.

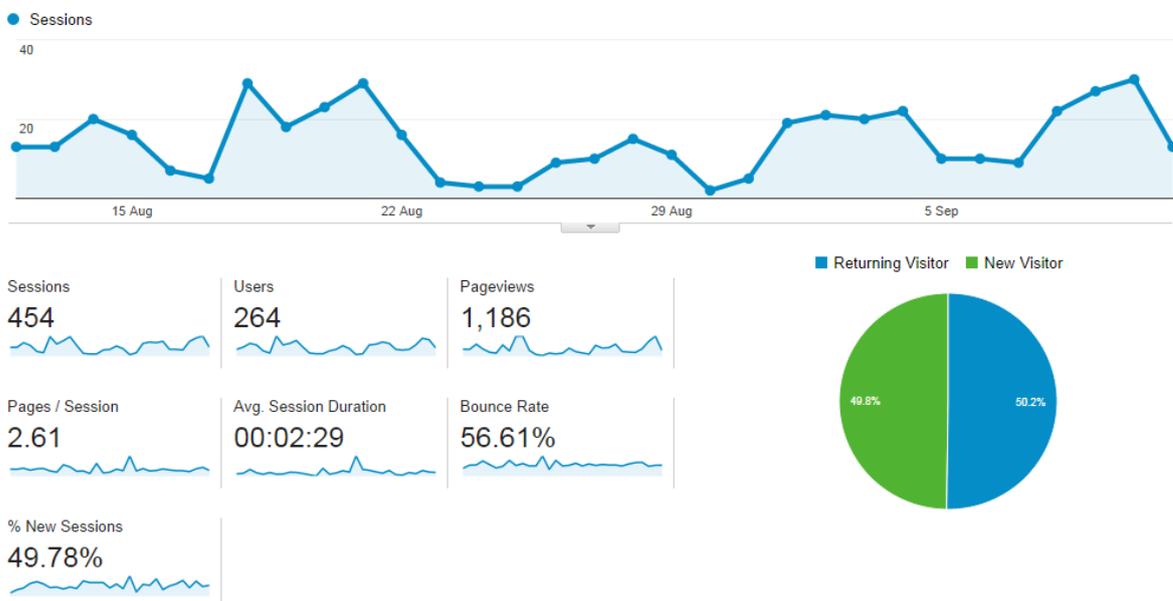
### Overview

PARTICIPANTS: **57** NEW DEBATES (blog posts shared): **3**

### Website

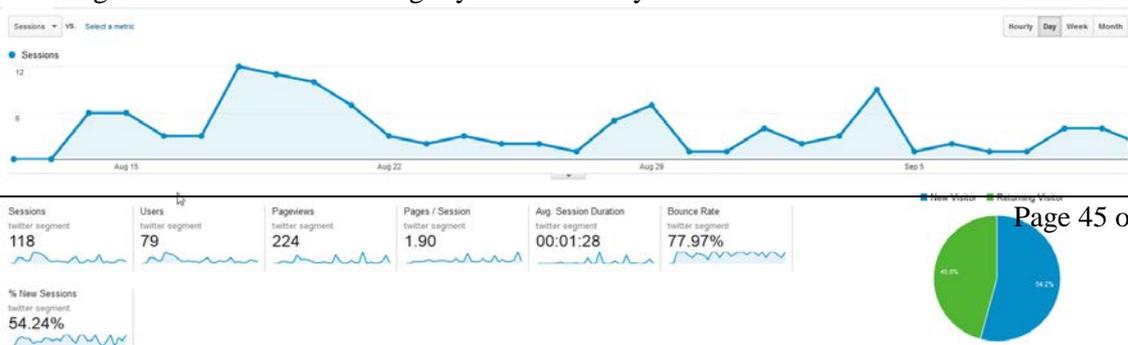
#### Overview

The number of sessions during this month has been 454, with 264 individual users visiting the website, this is slightly lower than the previous month (519 and 297 respectively). The average session duration has decreased from 3:25 to 2:29 minutes, which is still a notable duration.



#### Twitter traffic

The report for twitter for the second month shows that we have managed to establish a user base of returning visitors. The users are slightly fewer but they have more sessions.



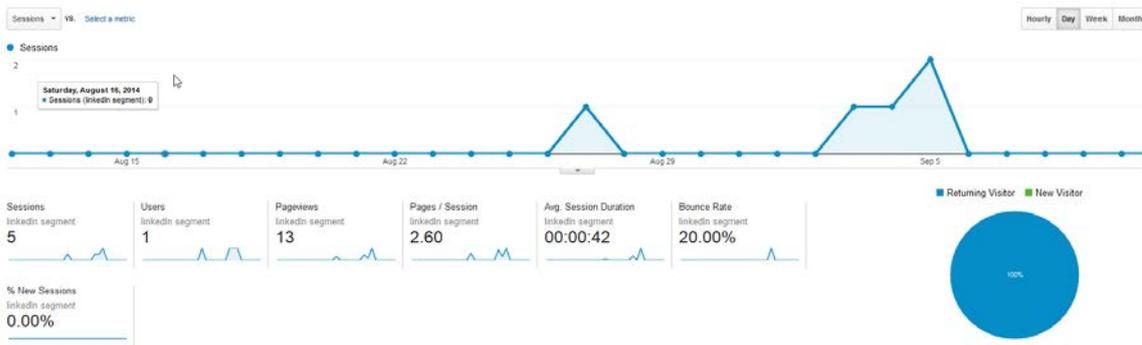
### Facebook traffic

The traffic generated from Facebook is less than the previous month. This can be attributed to the fact that August is generally a holiday season, but shows also that our efforts should be greater.



### LinkedIn traffic

We haven't been much active on Linked in during the month of August which is also reflected in the visitor's number.



### Segmentation by referrals

Twitter still generated the biggest part of referral traffic. There is a slight decrease from the previous month but a positive signal is that there are more users from google search landing to our page.

Source / Medium	Acquisition			Behavior		
	Sessions	% New Sessions	New Users	Bounce Rate	Pages / Session	Avg. Session Duration
	454 <small>% of Total: 100.00% (454)</small>	49.78% <small>Site Avg: 49.78% (0.00%)</small>	226 <small>% of Total: 100.00% (226)</small>	56.61% <small>Site Avg: 56.61% (0.00%)</small>	2.61 <small>Site Avg: 2.61 (0.00%)</small>	00:02:29 <small>Site Avg: 00:02:29 (0.00%)</small>
1. (direct) / (none)	163 (35.90%)	41.72%	68 (30.09%)	47.85%	3.33	00:03:19
2. t.co / referral	74 (16.30%)	27.03%	20 (8.85%)	66.22%	2.42	00:02:19
3. 213.239.209.240 / referral	57 (12.56%)	70.18%	40 (17.70%)	50.88%	2.18	00:01:59
4. ldbc.eu / referral	35 (7.71%)	57.14%	20 (8.85%)	48.57%	3.09	00:03:41
5. ldbcouncil.sti2.at / referral	27 (5.95%)	25.93%	7 (3.10%)	48.15%	2.81	00:01:44
6. google / organic	21 (4.63%)	57.14%	12 (5.31%)	61.90%	1.57	00:01:43
7. facebook.com / referral	7 (1.54%)	42.86%	3 (1.33%)	57.14%	1.43	00:00:11
8. linkedin.com / referral	5 (1.10%)	0.00%	0 (0.00%)	20.00%	2.60	00:00:42
9. semalt.semalt.com / referral	4 (0.88%)	100.00%	4 (1.77%)	100.00%	1.00	00:00:00
10. 10.10.10.1 / referral	2 (0.44%)	0.00%	0 (0.00%)	50.00%	1.50	00:00:10

### Segmentation by location

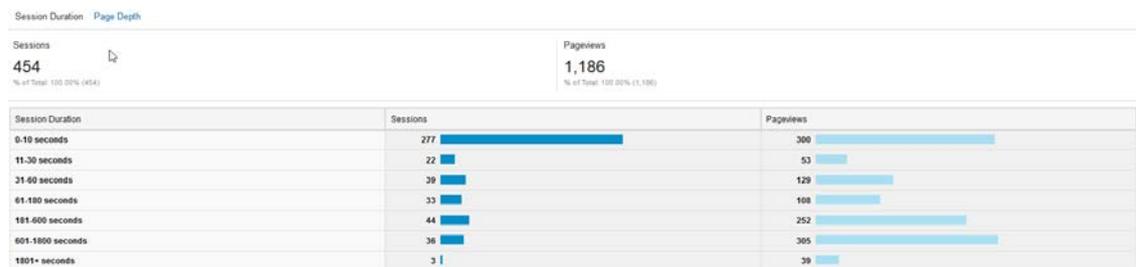
We have lost some of the traffic from the top three countries from previous month, but the traffic for the rest of the table has increased, which shows a more evenly spread interest.

Page	Pageviews	Unique Pageviews	Avg. Time on Page	Entrances	Bounce Rate	% Exit
	1,186 % of Total: 100.00% (1,186)	869 % of Total: 100.00% (869)	00:01:32 Site Avg: 00:01:32 (0.00%)	453 % of Total: 100.00% (453)	56.61% Site Avg: 56.61% (0.00%)	38.20% Site Avg: 38.20% (0.00%)
1. /	429 (36.17%)	326 (37.51%)	00:01:43	306 (67.55%)	55.23%	49.65%
2. /blog/getting-started-semantic-publishing-benchmark	100 (8.43%)	73 (8.40%)	00:01:24	53 (11.70%)	73.58%	56.00%
3. /developer/community	41 (3.46%)	25 (2.88%)	00:01:58	8 (1.77%)	50.00%	26.83%
4. /benchmarks/snb	39 (3.29%)	32 (3.68%)	00:01:50	4 (0.88%)	25.00%	33.33%
5. /user	39 (3.29%)	17 (1.96%)	00:00:36	5 (1.10%)	20.00%	5.13%
6. /blog/using-ldbc-spb-find-owim-performance-issues	33 (2.78%)	24 (2.76%)	00:01:03	13 (2.87%)	69.23%	39.39%
7. /industry/members	33 (2.78%)	18 (2.07%)	00:02:34	4 (0.88%)	50.00%	33.33%
8. /benchmarks/spb	32 (2.70%)	29 (3.34%)	00:01:24	8 (1.77%)	87.50%	34.38%
9. /blog	32 (2.70%)	17 (1.96%)	00:01:41	0 (0.00%)	0.00%	21.88%
10. /blog/datagen-data-generation-social-network-benchmark	28 (2.36%)	21 (2.42%)	00:01:37	4 (0.88%)	50.00%	39.29%

Country / Territory	Acquisition			Behavior		
	Sessions	% New Sessions	New Users	Bounce Rate	Pages / Session	Avg. Session Duration
	454 % of Total: 100.00% (454)	49.78% Site Avg: 49.78% (0.00%)	226 % of Total: 100.00% (226)	56.61% Site Avg: 56.61% (0.00%)	2.61 Site Avg: 2.61 (0.00%)	00:02:29 Site Avg: 00:02:29 (0.00%)
1. United States	77 (16.96%)	53.25%	41 (18.14%)	57.14%	2.17	00:01:47
2. Austria	46 (10.13%)	28.26%	13 (5.75%)	58.70%	2.41	00:03:46
3. Spain	46 (10.13%)	26.09%	12 (5.31%)	34.78%	3.93	00:05:19
4. Germany	35 (7.71%)	42.86%	15 (6.64%)	45.71%	2.89	00:01:35
5. Brazil	32 (7.05%)	100.00%	32 (14.16%)	100.00%	1.00	00:00:00
6. Switzerland	32 (7.05%)	28.12%	9 (3.98%)	59.38%	3.03	00:05:01
7. Netherlands	25 (5.51%)	20.00%	5 (2.21%)	32.00%	4.68	00:02:07
8. United Kingdom	23 (5.07%)	52.17%	12 (5.31%)	65.22%	1.96	00:01:38
9. Japan	21 (4.63%)	28.57%	6 (2.65%)	52.38%	1.76	00:02:08
10. Greece	16 (3.52%)	62.50%	10 (4.42%)	43.75%	2.31	00:03:56

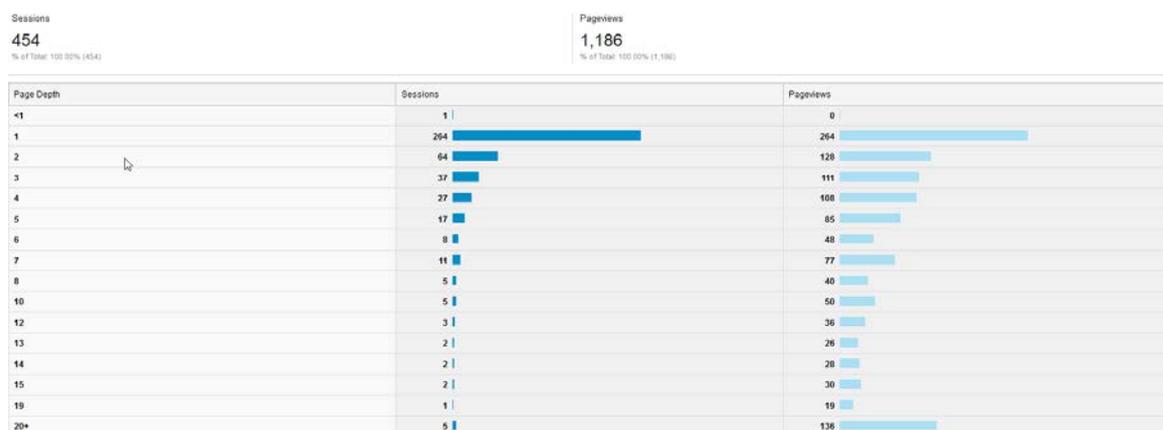
### Segmentation by session duration

It is similar to previous months taking into account the smaller number of users.



### Segmentation by page depth

The table below presents the sessions and page views by depth segmentation, meaning the number of pages visited by the users.



### Segmentation by page

In the table below we can see how the most visited content in our website is the home page and the blog articles, which means that our content is interesting for the users and that the promotion of the posts on the social networks is working.

### Segmentation by channel

A new interesting metric is the segmentation by channel. There is increase in the referrals and organic search which is positive. The decrease in social media referrals shows that our activity has been lower and it should be increased.

For 11.Jul- 11. Aug

Default Channel Grouping	Acquisition			Behavior		
	Sessions ? ↓	% New Sessions ?	New Users ?	Bounce Rate ?	Pages / Session ?	Avg. Session Duration ?
	519 <small>% of Total: 100.00% (519)</small>	51.06% <small>Site Avg: 51.06% (0.00%)</small>	265 <small>% of Total: 100.00% (265)</small>	56.65% <small>Site Avg: 56.65% (0.00%)</small>	2.89 <small>Site Avg: 2.89 (0.00%)</small>	00:03:25 <small>Site Avg: 00:03:25 (0.00%)</small>
1. Direct	222 (42.77%)	43.24%	96 (36.23%)	58.56%	3.02	00:04:05
2. Referral	150 (28.90%)	48.67%	73 (27.55%)	54.67%	2.97	00:03:13
3. Social	132 (25.43%)	65.15%	86 (32.45%)	56.82%	2.58	00:02:44
4. Organic Search	15 (2.89%)	66.67%	10 (3.77%)	46.67%	3.00	00:01:18

For 12.Aug-11.Sept

Default Channel Grouping	Acquisition			Behavior		
	Sessions ? ↓	% New Sessions ?	New Users ?	Bounce Rate ?	Pages / Session ?	Avg. Session Duration ?
	454 <small>% of Total: 100.00% (454)</small>	49.78% <small>Site Avg: 49.78% (0.00%)</small>	226 <small>% of Total: 100.00% (226)</small>	56.61% <small>Site Avg: 56.61% (0.00%)</small>	2.61 <small>Site Avg: 2.61 (0.00%)</small>	00:02:29 <small>Site Avg: 00:02:29 (0.00%)</small>
1. Referral	179 (39.43%)	67.04%	120 (53.10%)	62.57%	2.19	00:01:51
2. Direct	163 (35.90%)	41.72%	68 (30.09%)	47.85%	3.33	00:03:19
3. Social	91 (20.04%)	28.57%	26 (11.50%)	59.34%	2.40	00:02:25
4. Organic Search	21 (4.63%)	57.14%	12 (5.31%)	61.90%	1.57	00:01:43