HOBBIT

Collaborative Project

Holistic Benchmarking of Big Linked Data

Project Number: 688227 **Start Date of Project:** 2015/12/01 **Duration:** 36 months

Deliverable 7.1
First Preparations Report

Dissemination Level	Public
Due Date of Deliverable	Month 18, 26/05/2017
Actual Submission Date	Month 18, 26/05/2017
Work Package	WP7 - Evaluation Campaigns
Task	T7.1 & T7.2
Type	Report
Approval Status	Final
Version	1.0
Number of Pages	25

Abstract: This deliverable presents the state of the preparations for the organization of the first series of evaluation campaigns and the first series of workshops. The report also covers an overview of first evaluations of the platform for its use in the challenges.

Project funded by the European Commission's Horizon 2020 Program.

History

Version	Date	Reason	Revised by
0.1	28/04/2017	First draft created	Vassiliki Rentoumi & Grigorios Tzortzis
0.2	12/05/2017	Peer Reviewed	Mirko Spasic
0.3	19/05/2017	Content Revised	Vassiliki Rentoumi & Grigorios Tzortzis
1.0	25/05/2017	Final version created	Vassiliki Rentoumi & Grigorios Tzortzis

Author List

Organization	Name	Contact Information
NCSR-D	Vassiliki Rentoumi	vrentoumi@iit.demokritos.gr
NCSR-D	Grigorios Tzortzis	gtzortzi@iit.demokritos.gr
NCSR-D	Anastasia Krithara	akrithara@iit.demokritos.gr
NCSR-D	Dimitrios Vogiatzis	dimitrv@iit.demokritos.gr

Executive Summary

This deliverable presents the state of the preparations for the organization of the first series of evaluation campaigns and the first series of workshops. The report also covers an overview of first evaluations of the platform for its use in the challenges.

Hobbit is organizing five benchmarking challenges (i.e. evaluation campaigns) which will host the four HOBBIT benchmarks and their respective tasks (Section 2) and will obtain first results by Summer-Autumn 2017. Regarding the challenges, HOBBIT will be organizing the following challenges (Section 4):

- the DEBS Grand Challenge at the 11th ACM International Conference on Distributed and Event-Based Systems (DEBS) 2017
- the Mighty Storage Challenge (MOCHA) at the European Semantic Web Conference (ESWC) 2017
- the Open Knowledge Extraction (OKE) challenge at the European Semantic Web Conference (ESWC) 2017
- the Question Answering over Linked Data (QALD-7) challenge at the European Semantic Web Conference (ESWC) 2017
- the Instance Matching Benchmark for Spacial Data Challenge (OM2017) at the 16th International Semantic Web Conference (ISWC) 2017

In addition HOBBIT is going to organize the Workshop on Benchmarking Linked Data (BLINK) at ISWC 2017 where the participants will have the opportunity to run queries employing one of the HOBBIT benchmarks of their choice and evaluating the HOBBIT platform.

Contents

1 Introduction

The main objectives of WP7 - Evaluation Campaigns are (i) to organize benchmarking campaigns in order to measure the fitness of implemented systems for processing Big Linked Data and (ii) to organize workshops in order to present and promote these systems and the results of the evaluation campaigns. To achieve these goals, HOBBIT has defined four benchmarks and is currently organizing five benchmarking challenges (i.e. evaluation campaigns) at renowned conferences, where participants are invited to submit systems that tackle the HOBBIT benchmarks.

Specifically, benchmarks about (i) the generation and acquisition of data, (ii) the analysis and processing of data, (iii) storage and curation systems and (iv) the visualisation and service part of the big data value chain, have been devised. Each benchmark is divided into two tasks and HOBBIT is organizing the following challenges, that cover all the tasks of the HOBBIT benchmarks:

- the DEBS Grand Challenge at DEBS 2017
- the MOCHA challenge at ESWC 2017
- the OKE challenge at ESWC 2017
- the QALD-7 challenge at ESWC 2017
- the OM2017 challenge at ISWC 2017

In addition HOBBIT is going to organize the BLINK workshop at ISWC 2017, where participants will have the opportunity to run queries employing one of the HOBBIT benchmarks of their choice and evaluate the HOBBIT platform.

This deliverable reports on the state of the preparations for the organization of these challenges, while also covering an overview of the first evaluation of the HOBBIT platform for its use in the challenges. In particular, in Section 2 the four benchmarks and their tasks are presented. In Section 3 a report is presented that elaborates on the investigation we have conducted in order to locate the most appropriate event for hosting each task of the HOBBIT benchmarks. For each benchmark task, the most prominent events were identified and an analysis is provided regarding their suitability for hosting the task. The report concludes with the events that were finally chosen-booked to host the tasks. Section 4, describes the five HOBBIT challenges and the BLINK workshop. The motivation and an overview of each challenges is given, while the tasks comprising each challenge, on which systems will compete, are analyzed. Section 5 outlines the main actions that are required in order to organize and run the challenges and presents the current status of each challenge in terms of the actions completed. Finally, in Section 6 a first evaluation of the HOBBIT platform is provided.

2 Benchmarks Description

This section briefly describes the four HOBBIT benchmarks for which evaluation campaigns are organised in the context of HOBBIT. The four benchmarks are:

- **Benchmark I (INFAI) - Generation & Acquisition:** This benchmark concerns the generation and acquisition of data to test the performance of systems that implement approaches for obtaining RDF data from (i) semi-structured data streams such as sensor data (smart metering, geo-spatial information, etc.) and (ii) unstructured data streams (Twitter, RSS feeds, etc.). The benchmark will

test for the scalability as well as the accuracy of systems. The key performance indicators for the benchmark will include the runtime of the approaches, their precision, recall, F-measure (both micro and macro), as well as fine-grained evaluations w.r.t. the types of resources and relations to extract. The benchmark consists of two tasks:

  - : Data Extraction Benchmark for Sensor Data (INFAI)
  - Task B: Data Extraction Benchmark for Unstructured Data (INFAI)

- **Benchmark II (FORTH) - Analysis & Processing:** This benchmark concerns the linking and analysis of data KPIs, of the big data value chain, to test (i) the performance of OM2017 Challenge methods and tools for Linked Data and (ii) the performance of machine learning methods (supervised and unsupervised) for data analytics. The benchmark consists of two tasks:
  - Task A: Link Discovery Analysis Benchmark (FORTH)
  - Task B: Structured Machine Learning Benchmark (AGT)
- **Benchmark III (OPENLINK) - Storage & Curation:** This benchmark concerns the testing of storage and curation systems and extends the Social Network Benchmark by the Linked Data Council. It introduces important modifications in its synthetic data generator, its dataset and its queries, all while preserving their most relevant features, resulting in the first version of the Data Storage benchmark in HOBBIT. It utilizes a common SPC benchmark framework, which was also previously used by SPC-2. The benchmark consists of two tasks:
  - Task A: Data Storage Benchmark (OPENLINK)
  - Task B: Versioning Benchmark (FORTH)
- **Benchmark IV (FRAUNHOFER) - Visualisation & Services:** This benchmark concerns the visualization and service part of the big data value chain. Specifically, the focus of the benchmark is towards (i) question answering and (ii) faceted browsing. The project will not benchmark user interfaces themselves, but focus on providing performance and accuracy measurements for approaches used in interfaces. The benchmark consists of two tasks:
  - Task A: Question Answering Benchmark (FRAUNHOFER)
  - Task B: Faceted Browsing Benchmark (FRAUNHOFER)

More information on the HOBBIT benchmarks are available at the project website[1].

3 Preliminary Report

After conducting a thorough investigation regarding existing workshops and events that could potentially host the tasks of the HOBBIT benchmarks, described in Section 2, in the following paragraphs we present a preliminary report which contains the findings of this investigation. The initial choices regarding potential events are summarized in Table 5, while in Table 6 we present the finally chosen event for each task of each benchmark.

[1] https://project-hobbit.eu/

3.1 Benchmark I (INFAI)

The suggested workshops/events for both Task A - Data Extraction Benchmark for Sensor Data & Task B - Data Extraction Benchmark for Unstructured Data of Benchmark I are summarized in Table 1 and are further analyzed in Sections 3.1.1 and 3.1.2.

Table 1: Potential Events for Benchmark I

Benchmark I (INFAI)	Event
Data Extraction Benchmark for Sensor Data (INFAI)	**New challenge in ESWC**
	DEBS Grand Challenge
	Geo Linked Data workshop (*last edition: 2015*)
	RDF Stream Processing workshop (*last edition: 2015*)
Data Extraction Benchmark for Unstructured Data (INFAI)	**OKE:** Open Knowledge Extraction challenge
	Microposts
	DEBS Grand Challenge

3.1.1 Task A: Data Extraction Benchmark for Sensor Data (INFAI)

We suggested the organization of a new challenge at the European Semantic Web Conference (ESWC) in conjunction with the Data Storage Benchmark task from Benchmark III and the extension of the "Top-k Shortest Paths in large typed RDF Datasets Challenge" @ ESWC 2016 to also include a task related to the Data Extraction Benchmark for Sensor Data.

The DEBS Grand Challenge is held within the context of the ACM International Conference on Distributed and Event-Based Systems (DEBS). It comprises a dominant choice for the organization of the Data Extraction Benchmark for Sensor Data task (Benchmark I), since on its 2016 version it focused on the evaluation of event based systems for real time analytics over high volume data streams within the context of graph models. In particular, the DEBS 2016 Grand Challenge explicitly focused on processing streaming data and thus dynamic graphs.

However, the DEBS 2015 Grand Challenge was even more related with the Data Extraction Benchmark for the Sensor Data task as it focused on the evaluation of event based systems in the context of real time analytics over high volume geospatial data streams. Therefore, we needed to ask the organizing committee of DEBS Grand Challenge which subjects/tasks were going to be included in the DEBS 2017 Grand Challenge and if the Data Extraction Benchmark for Sensor Data task could comprise a part of it.

The Geo Linked Data workshop could comprise a third choice assuming it will be organized again in 2017, since the topics presented on its 2015 version (i.e. Interoperability and Integration: Geospatial Linked Data and Standards (Geo SPARQL, INSPIRE, W3C, OGC), Extraction/Transformation of Geospatial Linked Data from Conventional Sources, Integration (schema mapping, interlinking, fusion) Techniques for Geospatial RDF Data) are very similar to the ones demonstrated by this task. Further the RDF Stream

Processing (RSP) workshop has not been held in 2016. Therefore, we needed to ask the RSP@ESWC 2015 organizers if this workshop will be held in 2017.

3.1.2 Task B: Data Extraction Benchmark for Unstructured Data (INFAI)

Both Open Knowledge Extraction challenge (OKE) and Microposts (6th workshop on Making Sense of Microposts) were very predominant choices, since they both exhibit very similar topics with Task B of Benchmark I. OKE, which is held within the context of ESWC, is put first in our list since it comprises a pure challenge containing three tasks of similar content (Entity Recognition, Linking and Typing for Knowledge Base Population, Class Induction and Entity Typing for Vocabulary and Knowledge Base Enrichment, Web-Scale Knowledge Extraction by Exploiting Structured Annotation) with the ones exhibited in HOBBIT's Benchmark I Task B.

On the other hand, Microposts, held within the context of the World Wide Web Conference (WWW), is ranked second since it is mainly a workshop that also contains a challenge, which is however limited to one task, namely the Named Entity Recognition and Linking (NEEL) task. This task involves the automatic recognition of entities and their types from English microposts, and their linking to the corresponding English DBpedia 2014 resources (if the resources exist) or NIL identifiers.

Additionally, the DEBS Grand Challenge, held within the context of the ACM DEBS conference, could also comprise a choice for Task B of Benchmark I, since on its 2016 version it focuses on the evaluation of event based systems for real time analytics over high volume data streams within the context of graph models. All in all the DEBS 2016 version mostly focused on social network analysis and processing of streaming data in order to detect posts that trigger the greater activity and to identify the communities that are currently involved in a topic.

Although the DEBS 2015 Grand Challenge involved the analysis of social network graphs (e.g. Twitter), it did not focus on the extraction of graph-entities and their properties or the analysis of natural language streams (as it is asked in Task B of Benchmark I). We took the decision to contact the organisers to ask if the HOBBIT Data Extraction Benchmark for Unstructured Data task could be hosted within the context of the DEBS 2017 Grand Challenge.

3.2 Benchmark II (FORTH)

The suggested workshops/events for both Task A -Link Discovery Analysis Benchmark & Task B - Structured Machine Learning Benchmark of Benchmark II are summarized in Table 2 and are further analyzed in Sections 3.2.1 and 3.2.2.

3.2.1 Task A: Link Discovery Analysis Benchmark (FORTH)

Our first choice for this task is the Ontology Matching Workshop (OM) since it dealt with similar subjects as the Link Discovery Benchmark. In 2016 it was collocated with the International Semantic Web Conference (ISWC). We propose that the Link Discovery Benchmark will be part of the OM2017 Challenge at ISWC 2017. Through this Challenge HOBBIT will conduct an extensive and rigorous evaluation of ontology matching and instance matching (link discovery) approaches through the OAEI (Ontology Alignment Evaluation Initiative) 2017 campaign.

Know@LOD (organised in 2016 in conjuction with ESWC) and especially the Linked Data Mining Challenge (LDMC), hosted by the same workshop, comprises our first suggestion for the Link Discovery Analysis Benchmark task since it is the only event out of the six proposed ones that includes the organization

Table 2: Potential Events for Benchmark II

Benchmark II (FORTH)	Event
Link Discovery Analysis Benchmark (FORTH)	**ISWC:**
	LDMC: Linked Data Mining Challenge @Know@LOD
	Grades: Graph Data-management Experiences & Systems @SIGMOD
	SBD: The International Workshop on Semantic Big Data
	LDOW: Linked Data on the Web
	LWDM: International Workshop on Linked Web Data Management
Structured Machine Learning Benchmark (AGT)	**DEBS** Grand Challenge
	PKDD Discovery Challenge @ECML

of a challenge. Further, the Linked Data Mining Challenge (LDMC) held on 2016 consists of only one task, which is "the prediction of the review class of music albums". Although the topics communicated within the context of the Know@LOD 2016 workshop are very relevant to the Link Discovery Analysis Benchmark task, the task run by the Linked Data Mining Challenge is not very relevant. Therefore, we asked the organizing committee if the Linked Data Mining Challenge could incorporate Task A from Benchamrk II.

Grades, sponsored by the Linked Data Benchmark Council[2] and co-located with the 8th TUC meeting, comprises our second choice since it involves proposals for benchmarks for data integration tasks (i.e. instance matching and ETL techniques). On the other hand, Grades does not include any provision for the organization of challenges.

SBD, held in 2016 in conjunction with the ACM SIGMOD conference, accepts various types of papers (i.e. research/experiments/applications papers) focusing among others on topics such as "Ontology-based Approaches for Modeling, Mapping" and "Evolution and Real-world Ontologies in the Context of Semantic Big Data".

The LDOW workshop, held within the context of the WWW conference, is very relevant to Task A of Benchamrk II since one of its main topics refers to the Integration of Web Data from Large Numbers of Data Sources, while also including similar subtopics such as "Schema Matching and Clustering", and "Evaluation of Linking and Schema Matching Methods". LDOW does not include any provision for organizing challenges.

The LWDM workshop (co-located with the Querying Graph Structured Data (GraphQ 2016) workshop) covers a broad variety of related topics to Task A of Benchamrk II, such as "Linked Web Data Modeling: Linked Data Integration, Ontology mapping, merging, and alignment".

[2] ldbcouncil.org

3.2.2 Task B: Structured Machine Learning Benchmark (AGT)

The first choice for Task B of Benchmark II is the DEBS Grand Challenge, held within the context of the ACM International Conference on Distributed and Event-Based Systems (DEBS). The DEBS Grand Challenge held on 2016 referred to the evaluation of event-based systems for real-time analytics over high volume data streams in the context of graph models. In particular, its 2016 version targeted (1) the identification of the posts that currently trigger the most activity in the social network, and (2) the identification of large communities that are currently involved in a topic. Under these terms the DEBS Grand Challenge communicated tasks are very similar with the targets set by Task B of Benchmark II.

The second choice for our task is the PKDD Discovery Challenge held within the context of the ECML-PKDD conference. This event is the premier European machine learning and data mining conference and builds upon a very successful series of ECML and PKDD conferences which have been jointly organized for the past 15 years. The challenge organized in 2016 involved two tasks; the "Bank Card Usage Analysis" and the "SPHERE Challenge Activity Recognition with Multimodal Sensor Data" task. In particular, among the objectives of the "Sphere Challenge Activity Recognition with Multimodal Sensor Data" task is the use of sensor data in order to be able to learn patterns of behaviour, and to track the deterioration/progress of persons that suffer or recover from various medical conditions. To achieve these goals the use of machine learning techniques is necessary, primarily, in order to predict the Activities of Daily Living (ADL) (e.g. tasks such as meal preparation, watching television) and secondarily to predict the posture/ambulation of a sufferer (e.g. walking, sitting, transitioning). Under these terms the goals of the SPHERE challenge are very similar with the ones set by Task B of Benchmark II.

3.3 Benchmark III (OPENLINK)

The suggested workshops/events for both Task A - Data Storage Benchmark and Task B - Versioning Benchmark of Benchmark III are summarized in Table 3 and are further analyzed in Sections 3.3.1 and 3.3.2.

3.3.1 Task A: Data Storage Benchmark (OpenLink)

Despite the fact that all three suggested events are very relevant to the Data Storage Benchmark task, none, apart from the ESWC, has so far organized any challenges.

SBD, held in 2016 in conjunction with the ACM SIGMOD conference, accepts various types of papers (i.e. research/experiments/applications papers) on topics such as "Semantic Big Data Archives: Efficient Archiving and Preservation, Evolution Representation, Compression Approaches" that are very relevant to Task A of Benchmark III.

The BPOE workshop held in 2016 includes topics very relevant to the Data Storage Benchmark task. In particular, the topics are as follows: "Big data workload characterization and benchmarking", "Performance analysis of big data systems", and "Workload-optimized big data systems".

3.3.2 Task B: Versioning Benchmark (FORTH)

The Open RDF Archiving Challenge is held within the context of the workshop on Managing the Evolution and Preservation of the Data Web, which is an event co-located with the 13th ESWC conference. The Open RDF Archiving Challenge involves very similar subjects with the Versioning Benchmark task. In particular, it involves the showcasing of developments on the following topics: functionality over RDF archives, a

Table 3: Potential Events for Benchmark III

Benchmark III (OpenLink)	Event
Data Storage Benchmark (OpenLink)	New challenge in ESWC
	SBD: The International Workshop on Semantic Big Data
	BPOE: Big Data Benchmarks, Performance, Optimization and Emerging Hardware
Versioning Benchmark (FORTH)	**New challenge in ESWC OR Open RDF Archiving Challenge** @ Workshop on Managing the Evolution and Preservation of the Data Web
	Diachron: International Workshop on Preservation of Evolving Big Data
	LWDM: Linked Web Data Management
	LDOW: Linked Data on the Web

potential commercial application of RDF archives, tools to support/manage RDF archives at Web scale, and visual interfaces for evolving data/archives.

The Diachron workshops in 2015 and 2016, addressed a variety of topics related especially on preservation of linked data and among others it also focused on "Efficient archiving techniques for Big Data" and on "Benchmarking of Big Data archiving tools". Diachron provides a forum for Big Data researchers and practitioners to discuss exchange and disseminate their work but it has not included so far the organization of challenges.

The LWDM covers a broad variety of related topics to our task, such as "data management issues related to the Linked Data", "linked data modeling", and "engineering linked web systems".

The LDOW workshop has subjects close to Task B of Benchmark III, such as "Web Data Quality Assessment/Web Data Cleansing" and "Mining the Web of Data/Linked Data Applications".

3.4 Benchmark IV (FRAUNHOFER)

The suggested workshops/events for both Task A - Question Answering Benchmark and Task B - Faceted Browsing Benchmark" of Benchmark IV are summarized in Table 4 and are further analyzed in Sections 3.4.1 and 3.4.2.

Table 4: Potential Events for Benchmark IV

Benchmark IV (Fraunhofer)	Event
Question Answering Benchmark (Fraunhofer)	**QALD-7 :** Question Answering over Linked Data
	LWDM: Linked Web Data Management
	Diachron: International Workshop on Preservation of Evolving Big Data
Faceted Browsing Benchmark (Fraunhofer)	New challenge in ESWC
	SBD: The International Workshop on Semantic Big Data
	Diachron: International Workshop on Preservation of Evolving Big Data
	LWDM: Linked Web Data Management

3.4.1 Task A: Question Answering Benchmark (IAIS)

The most appropriate event to host the Question Answering Benchmark task is the 2017 QALD-7 challenge which will be held at ESWC. It comprises a series of evaluation campaigns on question answering over linked data, with a strong emphasis on multilingualism and hybrid approaches using information from both structured and unstructured data. QALD 2016 involved three tasks concerning the following topics:

- Task 1: Multilingual question answering over RDF data
- Task 2: Hybrid question answering over RDF data and free text
- Task 3: Statistical question answering over RDF data cubes.

All the tasks are very relevant to the HOBBIT Question Answering Benchmark task of Benchmark IV.

The GraphQ workshop, was co-located in 2016 with the 19th International Conference on Extending Database Technology / 19th International Conference on Database Theory (EDBT/ICDT 2016 Joint Conference) and held in conjunction with the 6th International Workshop on Linked Web Data Management (LWDM). It seeks to bring together many research groups working on graph-structured data, with a particular focus on the querying issues. The topics addressed in GraphQ are very relevant to the ones addressed in our task. Some of the topics covered include:

- Query models and languages for graph-structured data

- Querying RDF graphs, ontologies, or other knowledge representations
- Querying XML and semi-structured databases
- Querying graph data using big data technologies.

Although the topics addressed are very relevant with the subjects of Task A, it has not organized any challenges.

The Diachron workshop, both in 2015 and 2016 included topics such as "User interfaces for evolving knowledge presentation" and "Querying techniques for Big Data archives". Diachron provides a forum for Big Data researchers and practitioners to discuss, exchange and disseminate their work but it has not yet organized any challenges.

3.4.2 Task B: Faceted Browsing Benchmark (IAIS)

We suggested to liaise with the ESWC organizers to investigate the probability of organizing a new challenge at ESWC or to extend the "Top-k Shortest Paths in large typed RDF Datasets Challenge".

SBD is also accepting papers with a variety of topics which also include a "statistics and visualization" topic.

The Diachron Workshop both in 2015 and 2016, addressed a variety of topics related especially to the preservation of linked data and among others it also focused on "Visualization and exploration analysis of evolving big data" and "Visualizing trends, changes and paradigm shifts".

The LWDM covers a broad variety of related topics to task b such as "Browsing and visualization of Linked Data on the Web".

All the workshops proposed above, although, in their majority, contain very relevant topics to the Faceted Browsing Benchmark task, they mainly provide fora for Big Data researchers and practitioners to discuss, exchange and disseminate their work, rather than providing the opportunity of organizing challenges. We asked the organizers of the above events if they would welcome the organization of challenges on topics addressed within the context of the Faceted Browsing Benchmark task.

3.5 Summary of Suggested Events for each Benchmark/Task

Based on the investigation we conducted, an initial choice of events to host the benchmarks' tasks was made. Table 5 lists the chosen event for each task.

3.6 Chosen Events for each Benchmark/Task

The final list of appropriate events to host HOBBIT benchmarks I-IV and their respective tasks are presented in Table 6.

For Task A - Data Extraction Benchmark for Sensor Data pertaining to Benchmark I, for Task A - Data Storage Benchmark and Task B - Versioning Benchmark pertaining to Benchmark III as well as for Task B - Faceted Browsing Benchmark pertaining to Benchmark IV we decided to propose and organize a new challenge, called Mighty Storage Challenge (MOCHA), at ESWC 2017. For Task B - Data Extraction Benchmark for Unstructured Data pertaining to Benchmark I we have selected the Open Knowledge Extraction Challenge (OKE) which will be held at ESWC 2017.

Table 5: Initial List of Events/Benchmarks

Benchmark I (INFAI)	**Event**
Data Extraction Benchmark for Sensor Data (INFAI)	**ESWC:** a new challenge
Data Extraction Benchmark for unstructured data (INFAI)	**OKE:** Open Knowledge Extraction challenge (@ESWC)
Benchmark II (FORTH)	**Event**
Link Discovery Analysis Benchmark (FORTH)	**ISWC:**OM 2017 Challenge
hline Structured Machine Learning Benchmark (AGT)	**DEBS** Grand Challenge
Benchmark III (OpenLink)	**Event**
Data Storage Benchmark (OpenLink)	**ESWC:** a new challenge
Versioning Benchmark (FORTH)	**ESWC:** a new challenge
Benchmark IV (Fraunhofer)	**Event**
Question Answering Benchmark (Fraunhofer)	**QALD-7 :** Question Answering over Linked Data (@ESWC)
Faceted Browsing Benchmark (Fraunhofer)	**ESWC:** a new challenge

For Task A - Link Discovery Analysis Benchmark pertaining to Benchmark II, the OM challenge at ISWC 2017 has been selected.

For Task B - Structured Machine Learning Benchmark pertaining to Benchmark II, the 2017 DEBS Grand Challenge has been selected, which will be held at DEBS 2017. Finally, for Task A - Question Answering Benchmark pertaining to Benchmark IV, the Question Answering over Linked Data Challenge (QALD-7) has been selected, which will be held at ESWC 2017.

Regarding each of the selected events we submitted challenge proposals which described the scope and the content of the proposed challenges. All five challenge proposals were accepted by the respective conferences.

In Section 4 we present the content of each of the five proposed and booked challenges.

4 Description of Proposed and Booked Campaigns/Challenges

After the successful submission of challenge proposals at the ESWC 2017, DEBS 2017 and ISWC 2017 conferences, HOBBIT is going to organize in 2017 the first evaluation campaigns. These are: the Mighty Storage Challenge (MOCHA), the Open Knowledge Extraction Challenge (OKE) and the Question Answering over Linked Data Challenge (QALD-7) which will be held in conjunction with ESWC 2017, the DEBS Grand Challenge at DEBS 2017 and the OM2017 Challenge at ISWC 2017.

Further, the 2016 Benchmarking Linked Data workshop (BLINK) was held at ISWC and organized by HOBBIT. BLINK provided a forum where topics related to the evaluation of Linked Data technologies for different steps of the Big Linked Data Chain (e.g., expressive power, usability and performance) were discussed and elaborated upon. For 2017, the BLINK workshop will be organized by HOBBIT and we

Table 6: Final List of Events/Benchmarks

Event	Task	Benchmark
MOCHA @ESWC 2017 May 28th to June 1st, 2017	Task A: Data Extraction Benchmark for Sensor Data (INFAI)	I (INFAI)
	Task A: Data Storage Benchmark (OpenLink)	III (OpenLink)
	Task B: Versioning Benchmark (FORTH)	III (OpenLink)
	Task B: Faceted Browsing Benchmark (IAIS)	IV (IAIS)
OKE @ESWC 2017 May 28th to June 1st, 2017	Task B: Data Extraction Benchmark for unstructured data (INFAI)	I (INFAI)
QALD-7 @ESWC 2017 May 28th to June 1st, 2017	Task A: Question Answering Benchmark, (IAIS)	IV (IAIS)
Grand Challenge @DEBS 2017 June 19th to June 23rd, 2017	Task B: Structured Machine Learning Benchmark (AGT)	II (FORTH)
OM2017 @ISWC 2017 October 21st to October 25th, 2017	Task A: Link Discovery Analysis Benchmark (FORTH)	II (FORTH)

suggested and booked to host it again at ISWC. For the 2017 BLINK we suggested to accept contributions presenting experiences with benchmarking Linked Data technologies as well as technical contributions regarding the development of benchmarks for all aspects of the Linked Data/Big Data lifecycle. We further suggested that the interested parties will be able to run additional queries for the HOBBIT benchmarks of their choice using the HOBBIT platform. All in all, BLINK aims at giving the opportunity to participants to employ and evaluate all benchmarks provided by HOBBIT (see Section 2 for the HOBBIT benchmarks).

In the following paragraphs we present the detailed descriptions of all evaluation campaigns, including the BLINK 2017 workshop, together with the HOBBIT benchmarks/tasks that each campaign hosts, against which competing systems will be evaluated. More details can be found at the project's website[3].

4.1 MOCHA Challenge @ ESWC 2017

4.1.1 Challenge Motivation

Triple stores are the backbone of most applications based on Linked Data. Hence, devising systems that achieve an acceptable performance on real datasets and real loads is of central importance for the practical applicability of Semantic Web technologies. So far, it is only partly known whether we have already passed this cap. With this challenge, we aim to provide objective measures for how well current systems perform on real tasks of industrial relevance and detect bottlenecks of existing systems to further push their development towards practical usage.

[3] https://project-hobbit.eu/challenges/

4.1.2 Challenge Overview

The aim of the MOCHA challenge is to test the performance of solutions for SPARQL processing in aspects that are relevant for modern applications. These include ingesting data, answering queries on large datasets and serving as backend for applications driven by Linked Data. The challenge will test the systems against data derived from real applications and with realistic loads. An emphasis will be put on dealing with changing data in form of streams or updates.

The MOCHA challenge is accepted at ESWC 2017, which takes places from May 28th to June 1st, 2017, in Portoroz, Slovenia.

4.1.3 Challenge Tasks

In the context of the MOCHA challenge four benchmarking tasks are organized, pertaining to Task A (Data Extraction Benchmark for Sensor Data) from Benchmark I, Task A (Data Storage Benchmark) and Task B (Versioning Benchmark) from Benchmark III and Task B (Faceted Browsing Benchmark) from Benchmark IV (see Section 2 for details on benchmarks and their tasks). These tasks are:

- **RDF Data Ingestion:** The constant growth of the Linked Data Web in velocity and volume has increased the need for triple stores to ingest streams of data and perform queries on this data efficiently. The aim of this task is to measure the performance of SPARQL query processing systems when faced with streams of data from industrial machinery in terms of efficiency and completeness. The experimental setup will hence be as follows: we will increase the size and velocity of RDF data used in our benchmarks to evaluate how well can a system store streaming RDF data obtained from industry. The data will be generated from one or multiple resources in parallel and will be inserted using SPARQL INSERT queries. This facet of triple stores has (to the best of our knowledge) never been benchmarked before. SPARQL SELECT queries will be used to test the system's ingestion performance and storage abilities.
- **Data Storage Benchmark:** The aim of this task is to measure how data stores perform with different types of queries. Therefore, we will develop an RDF benchmark that measures how datastores perform with interactive, simple, read, SPARQL queries as well as complex, business intelligence (BI) queries. Running the queries will be accompanied with high insert rate of the data (SPARQL UPDATE queries), in order to mimic the real use cases where READ and WRITE operations are bundled together. Typical bulk loading scenarios will also be supported. The queries and query mixes will be designed to stress the system under test in different choke-point areas, while being credible and realistic.
- **Versioning RDF Data:** The evolution of datasets often requires storing different versions of the same dataset, so that interlinked datasets can refer to older versions of an evolving dataset and upgrade at their own pace, if at all. Supporting the functionality of accessing and querying past versions of an evolving dataset is the main challenge for archiving/versioning systems. The aim of this task is to test the ability of versioning systems to efficiently manage evolving datasets and queries, evaluated across the multiple versions of such datasets.
- **Faceted Browsing Benchmark:** Faceted browsing stands for a session-based (state-dependent) interactive method for query formulation over a multidimensional information space. It provides a user with an effective way for exploration through a search space. After having defined the initial search space, i.e., the set of resources of interest to the user, a browsing scenario consists of applying

(or removing) filter restrictions of object-valued properties or of changing the range of a number-valued properties. Using such operations, aimed at selecting resources with desired properties, the user browses from state to state, where a state consists of the currently chosen facets and facet values and the current set of instances satisfying all chosen constraints. The aim of this task is to test the capability of systems in enabling faceted browsing through large-scale RDF datasets, that is, it analyses their efficiency in navigating through large datasets, where the navigation is driven by intelligent iterative restrictions.

More information on the MOCHA tasks are available on the challenge website[4].

4.2 OKE Challenge @ ESWC 2017

4.2.1 Challenge Motivation

Most of the Web content consists of natural language text, e.g., websites, news, blogs, micro-posts, etc., hence a main challenge is to extract as much relevant knowledge as possible from this content and publish it in the form of Semantic Web triples. There is huge work on knowledge extraction and knowledge discovery contributing to address this problem. In fact, results of knowledge extraction systems are usually evaluated against tasks that do not focus on specific Semantic Web goals. For example, tasks such as named entity recognition, named entity disambiguation and relation extraction are certainly of importance for the Semantic Web, but in most cases such tasks are designed without considering the output design and formalization in the form of Linked Data and OWL ontologies. This makes results of existing methods often not directly reusable for populating the Semantic Web, until a translation from linguistic semantics to formal semantics is performed.

4.2.2 Challenge Overview

The goal of the OKE challenge is to test the performance of knowledge extraction systems with respect to the Semantic Web. The OKE challenge has the ambition to provide a reference framework for research on "Knowledge Extraction from text for the Semantic Web" by redefining a number of tasks (typically from information and knowledge extraction) by taking into account specific Semantic Web requirements.

The OKE challenge is accepted at ESWC 2017, which takes places from May 28th to June 1st, 2017, in Portoroz, Slovenia.

4.2.3 Challenge Tasks

In the context of the OKE challenge four benchmarking tasks are organized, pertaining to Task B (Data Extraction Benchmark for Unstructured Data) from Benchmark I (see Section 2 for details on benchmarks and their tasks). These tasks are:

- **Focused NE Identification and Linking:** The task comprises the identification of named entities in sentences and the disambiguation of the identified entities to the DBpedia knowledge base. The task is limited to three DBpedia ontology classes (Person, Place, Organisation) and their associated subclasses. A competing system is expected to identify elements in a given text by its start and end

[4] https://project-hobbit.eu/challenges/mighty-storage-challenge/msc-tasks/

index, further to generate an RDF formalizing the linking of the identified entities to the DBpedia knowledge base.

- **Broader NE Identification and Linking:** This task extends the above task in terms of the entity classes considered from DBpedia. Beside the three types in the first task, a competing system has to identify additional types of entities in this case.
- **Focused Musical NE Recognition and Linking:** The task is composed of two sub-tasks focused on musical NE recognition and linking, where MusicBrainz as Linked Data (MBL)[5] is used as the underlying knowledge base. The first sub-task, Focused Musical NE Recognition, consists of the identification and classification of named entities over a subset of the entities types in MBL which are defined according to the Music Ontology[6]. A competing system is expected to identify entity mentions in a given text by their start and end index, and to further assign them to one of the considered entity types. In the second sub-task, Musical NE Linking, a system has to link the recognized entities of the previous sub-task to the corresponding resource in MBL. Competing systems have to fulfill both sub-tasks.
- **Knowledge Extraction:** The task's goal is to extract knowledge from a given text and to formalize the knowledge in RDF triples. DBpedia is the used knowledge base in this task. The quality of competing systems is measured in terms of precision, recall and F1-score, which are calculated based on the comparison of the received graph, i.e., the triples that have been received from the competing system, and the expected graph. The evaluation module will try to match both graphs to each other in a way that the number of matching nodes and edges are as high as possible. Matching triples are counted as true positives. False positives are triples that have been received from the competing system but are not available in the expected graph. Finally, false negatives are triples that have been expected but are not received from the system.

More information on the OKE tasks are available on the challenge website[7].

4.3 QALD-7 Challenge @ ESWC 2017

4.3.1 Challenge Motivation

The past years have seen a growing amount of research on question answering (QA) over Semantic Web data, shaping an interaction paradigm that allows end users to profit from the expressive power of Semantic Web standards while at the same time hiding their complexity behind an intuitive and easy-to-use interface. At the same time the growing amount of data has led to a heterogeneous data landscape where QA systems struggle to keep up with the volume, variety and veracity of the underlying knowledge. The QALD-7 challenge aims at providing an up-to-date benchmark for assessing and comparing state-of-the-art-systems that mediate between a user, expressing his or her information need in natural language, and RDF data. It thus targets all researchers and practitioners working on querying Linked Data, natural language processing for question answering, multilingual information retrieval and related topics. The main goal is to gain insights into the strengths and shortcomings of different approaches and into possible solutions for coping with the large, heterogeneous and distributed nature of Semantic Web data. QALD has already a 6-year history of developing a benchmark that is increasingly being used as standard evaluation benchmark for question answering over Linked Data.

[5] https://drive.google.com/file/d/0B-oq_x72w8NUSDZfck53dXN0TEU
[6] http://musicontology.com/
[7] https://project-hobbit.eu/challenges/oke2017-challenge-eswc-2017/oke2017-challenge-tasks/

4.3.2 Challenge Overview

The key challenge for Question Answering (QA) over Linked Data is to translate a user's information need into a form such that it can be evaluated using standard Semantic Web query processing and inferencing techniques. The main task of QALD-7 therefore is the following: Given one or several RDF dataset(s) as well as additional knowledge sources and natural language questions or keywords, return the correct answers or a SPARQL query that retrieves these answers.

The QALD-7 challenge is accepted at ESWC 2017, which takes places from May 28th to June 1st, 2017, in Portoroz, Slovenia.

4.3.3 Challenge Tasks

In the context of the QALD-7 challenge four benchmarking tasks are organized, pertaining to Task A (Question Answering Benchmark) from Benchmark IV (see Section 2 for details on benchmarks and their tasks). These tasks are:

- **Multilingual question answering over DBpedia:** Given the diversity of languages used on the web, there is an impeding need to facilitate multilingual access to semantic data. The core of this task is to retrieve answers from an RDF data repository given an information need expressed in a variety of natural languages. The underlying RDF dataset will be DBpedia 2016-04. The task consists of more than 500 questions compiled and curated from previous challenges. The questions will be available in eight different languages (English, Spanish, German, Italian, French, Dutch, Romanian, and Farsi), possibly with the addition of three further languages (Korean, Hindi and Brazilian Portuguese). Those questions are general, open-domain factual questions, and vary with respect to their complexity. Each question is annotated with a manually specified SPARQL query and answers.
- **Hybrid question answering:** A lot of information is still available only in textual form, both on the web and in the form of labels and abstracts in Linked Data sources. Hence, approaches are needed that can not only deal with the specific character of structured data but also with finding information in several sources, processing both structured and unstructured information, and combining such gathered information into one answer. Therefore, this task asks competing systems to retrieve answers for questions that require the integration of data both from RDF and from textual sources. The task builds on DBpedia 2016-04 as RDF knowledge base, together with the English Wikipedia as textual data source, and consists of more than 100 English questions from past challenges. The questions are annotated with answers as well as a pseudo query that indicates which information can be obtained from RDF data and which from free text. The pseudo query is like an RDF query but can contain free text as subject, property, or object of a triple.
- **Large-Scale Question answering over RDF:** The focus of this task is towards the inclusion of a large-scale question set. Successful approaches should be able to scale up to a big data volume, handle a vast amount of questions and speed up the question answering process by parallelization, such that the highest number of questions can be answered, as accurately as possible in the shortest possible time. The task utilizes DBpedia 2016-04 as RDF knowledge base and questions are annotated with SPARQL queries and answers. Participating systems will be evaluated with respect to both the number of correct answers and the time needed.
- **English question answering over Wikidata:** In this task, questions originally formulated for DBpedia should be answered using Wikidata[8]. Thus, competing systems will have to deal with a different

[8] https://www.wikidata.org/

data representation structure. This task helps in evaluating how generic an approach is and how easy it is to adapt to a new data source. It comprises of 100 questions taken from last year's iteration of QALD-7 .

More information on the QALD-7 tasks are available on the challenge website[9].

4.4 Grand Challenge @ DEBS 2017

4.4.1 Challenge Motivation

The 2017 ACM DEBS Grand Challenge is the seventh in a series of challenges which seek to provide a common ground and uniform evaluation criteria for a competition aimed at both research and industrial event-based systems. The goal of this year's competition is to evaluate event-based systems for real-time analytics over high velocity and high volume data streams.

4.4.2 Challenge Overview

The focus of the 2017 Grand Challenge is on the analysis of the RDF streaming data generated by digital and analogue sensors embedded within manufacturing equipment, in order to detect anomalies in the behaviour of such manufacturing equipment.

The 2017 edition of the Grand Challenge is part of DEBS 2017, which takes places from June 19th to June 23rd, 2017, in Barcelona, Spain and is co-organized by ACM DEBS conference and HOBBIT.

4.4.3 Challenge Tasks

In the context of the 2017 Grand Challenge two benchmarking tasks are organized, pertaining to Task B (Structured Machine Learning Benchmark) from Benchmark II (see Section 2 for details on benchmarks and their tasks).

The overall goal of both tasks is to detect abnormal behaviour of a manufacturing machine based on the observation of the stream of measurements provided by such a machine. Data comes from two types of machines: injection molding machines and assembly machines. The data produced by each sensor is clustered and the state transitions between the observed clusters are modelled as a Markov chain. Based on this classification, anomalies are detected as sequences of transitions that happen with a probability lower than a given threshold. The difference between the first and the second task is that in the first task the number of machines to observe is fixed, while in the second task new machines dynamically join and leave the set of observable machines.

More information on the Grand Challenge tasks are available on the challenge website[10].

[9] https://project-hobbit.eu/challenges/qald2017/qald2017-challenge-tasks/
[10] https://project-hobbit.eu/challenges/debs-grand-challenge/

4.5 Instance Matching Benchmark for Spatial Data Challenge (OM2017 Challenge) at ISWC 2017

4.5.1 Challenge Motivation

The number of datasets published in the Web of Data as part of the Linked Data Cloud is constantly increasing. The Linked Data paradigm is based on the unconstrained publication of information by different publishers, and the interlinking of Web resources across knowledge bases. In most cases, the cross-dataset links are not explicit in the dataset and must be automatically determined using Instance Matching (IM) tools. The large variety of techniques requires their comparative evaluation to determine which one is best suited for a given context. Performing such an assessment generally requires well-defined and widely accepted benchmarks to determine the weak and strong points of the proposed techniques and/or tools.

4.5.2 Challenge Overview

Although a fairly large number of benchmarks has been proposed for evaluating Instance Matching Systems, only a limited number of link discovery benchmarks target the problem of linking geo-spatial entities. However, some of the largest knowledge bases on the Linked Open Data Web are geo-spatial knowledge bases (e.g., LinkedGeoData with more than 30 billion triples) and the systems for linking spatial resources requires techniques that differ from the classical mostly string-based approaches since the topology of the spatial resources and the topological relations between them is of central importance to systems driven by spatial data. We believe that due to the large amount of available geo-spatial datasets employed in Linked Data and in several domains, it is critical that benchmarks for geo-spatial link discovery are developed. In this challenge, a benchmark will be used which is based upon the Lance[11] scalable, schema-agnostic benchmark generator extended with appropriate transformations to tackle geo-spatial link discovery tasks.

The Instance Matching Benchmark for Spatial Data Challenge (OM2017 Challenge) is accepted at the OM workshop at ISWC 2017, which takes places from October 21st to October 25th, 2017, in Vienna, Austria.

4.5.3 Challenge Tasks

In the context of the OM2017 Challenge, benchmarking tasks pertaining to Task A (Link Discovery Analysis Benchmark) from Benchmark II will be organized (see Section 2 for details on benchmarks and their tasks). The exact details of these tasks have not been finalized yet, since the challenge will run in Autumn 2017, but the tasks shall focus on the different types of spatial object representations and will be provided with different severity levels for the applied transformations. In these transformations, objects may keep their representation, they may change their geometry, type or attributes, merge with other objects, or can completely disappear. This is a scenario that stems from the heterogeneous datasets (in structure and semantics) used to describe geo-spatial entities. The produced tasks will be used by IM tools that implement string-based as well as topological approaches for identifying matching entities. The IM competing systems will be evaluated for both accuracy (precision, recall and f-measure) and scalability.

[11] T. Saveta, E. Daskalaki, G. Flouris, I. Fundulaki, and A. Ngonga-Ngomo. LANCE: Piercing to the Heart of Instance Matching Tools. In ISWC, 2015.

4.6 BLINK Workshop @ ISWC 2017

4.6.1 Workshop Motivation

Scalable Linked Data processing frameworks are of key importance for the use of RDF, OWL and other Semantic Web technologies in real-world applications. However, the cost and effort required for introducing this technology in a value chain is significant. It is thus of utmost importance to devise Linked Data solutions that can be demonstrated to scale up to the requirements of practical users. Throughout the history of computer science, benchmarks and benchmarking platforms have been demonstrated to push the development of effective and efficient solutions. With this workshop, we aim to bundle the newest developments centered around (1) the methodology of benchmarking, (2) the benchmarks (including their strengths and weaknesses) and (3) the performance of current solutions across the Linked Data lifecycle. The first BLINK workshop at ISWC 2016 has demonstrated clearly that their is a need for a dedicated forum for these purposes.

4.6.2 Workshop Overview

BLINK provides a forum where topics related to the evaluation (included, but not limited to, expressive power, usability and performance) of Linked Data Technologies for different steps of the Big Linked Data Value Chain can be discussed and elaborated upon. The objectives of this workshop are to i) create a discussion forum where researchers and industrials can meet and discuss topics related to the performance of Linked Data systems and ii) expose and initiate discussions on best practices, different application needs and scenarios related to Linked Data management. Potential participants will be invited to submit papers for presentation at the workshop. In addition to presentation of submitted work, we plan to invite a leading researcher in this area to deliver a keynote. Our first choices as invited keynote speakers are Chris Bizer, Soren Auer, Orri Erling , Jerome Euzenat, Ihab Ilyas, and Philippe Cudre-Mauroux. Two paper sessions will take place. A final discussion session will close the event. This will seek to identify specific actions necessary to advance future research and development.

BLINK 2017 is part of ISWC 2017, which takes places from October 21st to October 25th, 2017, in Vienna, Austria.

4.6.3 Relation to HOBBIT Evaluation Campaigns

Although BLINK is a workshop and not an evaluation campaign like the aforementioned ones, in its 2017 edition interested parties will be invited to submit papers where they describe 5 queries per task for the HOBBIT benchmarks of their choice that would stress existing Linked Data technologies. The interested parties will be able to run those additional queries using the HOBBIT platform and the results of these experiments will be presented at a special session of the BLINK workshop during which authors will be able to discuss with the workshop organisers their experience with the results of the HOBBIT project. Hence, BLINK will give the opportunity to participants to employ and evaluate all benchmarks provided by HOBBIT.

More information on BLINK are available on the workshop website[12].

[12] https://project-hobbit.eu/events/workshop-on-benchmarking-linked-data-blink-2017/

5 Milestones for the Organization of Challenges

In order to reassure the smooth preparation and organization of the selected events/workshops we undertook several actions:

1. Preparation & Publication of Call for Participation (CFP): We structured the call for participation for each challenge and we published it to WikiCFP[13] and EasyChair Smart CFP[14], as well as to several mailing lists related to the semantic web (e.g. semantic-web@w3.org). Also the CFPs were sent to thousands of individuals who had participated in the targeted challenges the previous years.
2. Task Definition: We finalized the benchmark tasks descriptions that participate in each challenge.
3. Datasets & HOBBIT Platform: We provided the detailed description of the training & test data sets. We also ensured the timely preparation and delivery of training & test data pertaining to each task and of the HOBBIT platform for evaluation/testing. The detailed descriptions of data sets participating in each challenge task can be found on the website of each challenge under the tab "Tasks & Training Data".
4. Specification of Program and Organizing Committees: We defined the organizing and the program commitees for each challenge. Towards this action we sent several thousand email invitations to specific persons and to lists of persons with scientific expertise relevant to the tasks pertaining to each challenge. This action has been performed in cooperation with FORTH.
5. Website Preparation: We proceeded to the website preparation pertaining to each challenge in cooperation with the challenge organizers and FORTH. The full details for each challenge can be found on the HOBBIT website: https://project-hobbit.eu/challenges/.
6. For each of the booked challenges we undertook the following actions with regards to various deadlines:
   (a) Publication of Call for Papers
   (b) Publication of training data for each challenge task
   (c) Submission of challenge papers by participants
   (d) Reviewing of challenge papers
   (e) Notification of acceptance/rejection of challenge papers
   (f) Publication of test data for each challenge task
   (g) Submission of camera-ready papers by participants
   (h) Submission of systems by participants
   (i) Release of evaluation results for participating systems
   (j) Proclamation of winners
   (k) Proceedings preparation

   All the deadlines foreseen for points a-k for each of the upcoming challenges have been published in the corresponding challenges' websites, under the tab "important dates".

[13] http://www.wikicfp.com/cfp/
[14] https://easychair.org/cfp/

All the above actions (1-6) have been accomplished in cooperation with the HOBBIT partners responsible for the organization of each challenge (i.e. the HOBBIT Benchmark/Task Leaders) as shown in the following list:

- Benchmark I: INFAI
- Benchmark II: FORTH, AGT
- Benchmark III: OPENLINK, FORTH
- Benchmark IV: FRAUNHOFER

5.1 Current Status of Challenges

The three challenges that are hosted by ESWC 2017 from May 28th to June 1st, 2017, namely MOCHA, OKE and QALD-7, are at an advanced stage of preparation. Actions 1-5 have been completed, while for Action 6, the deadlines for points a-h have already passed and these points have been successfully completed. The MOCHA and OKE challenges have received 2 participating systems-papers each, while the QALD-7 challenge has received 3 systems-papers. The camera-ready papers describing the participating systems will be included in the challenges' post proceedings, which will be published by Springer as a Lecture Notes in Computer Science (LNCS) volume after the ESWC conference.

The Grand Challenge that is hosted by DEBS 2017 from June 19th to June 23rd, 2017, is also at an advanced stage of preparation. Actions 1-5 have been completed, while for Action 6, the deadlines for points a-h have already passed and these points have been successfully completed. The Grand Challenge has received 7 participating systems-papers. The Grand Challenge proceedings will be published by ACM.

Finally, the OM2017 Challenge and the BLINK workshop which are hosted by ISWC 2017 from October 21st to October 25th, 2017 are at the very beginning of preparation[15]. Currently, the website and the call for participation are being prepared, while immediately after the program committee and the benchmarking tasks will be finalized.

6 First Evaluation of the HOBBIT Platform

In this section, we present an overview of the first evaluation of the HOBBIT platform[16] to confirm its suitability for the challenges. The evaluation of the platform is focused on its usage in a common environment. Its goal is to show that it is possible to use the open-source platform even without a server environment. The scenario focuses on the sending of SPARQL queries and result sets during the evaluation phase of a SPARQL endpoint. Messages, that are transferred:

- from data generator to task generator: SPARQL query + result set
- from task generator to system: SPARQL query
- from task generator to evaluation storage: result set
- from system to evaluation storage: result set

[15] The notification of acceptance from ISWC 2017 was received on the 7th of April.
[16] This evaluation report was originally reported in D2.2.1 - First Version of the HOBBIT Platform.

From the Linked SPARQL Queries Dataset[17] it is known that the average size of a SPARQL query is 545.45 characters. A result set comprises 122.45 single results on average. Unfortunately, we do not know the average size of a single result. However, from several example queries we gathered the assumption that on average a result has 100 to 120 characters.

The platform was deployed on a small machine with an Intel Core i5 processor (2 cores with 2.5 GHz) and 2 GB RAM for the sake of the evaluation. The single benchmark runs are listed in Table 7. We executed the benchmark with three different numbers of queries that were generated with an expected response by two data generators. Our results show that the platform can run even within this minimalistic environment, hence proving, that it is highly flexible. The results also clearly indicate the necessity to deploy the platform in a large-scale environment to test some of the Big Linked Data systems. In particular, our results indicate that the average runtime per query grows as well as the standard deviation of the runtime grows with an increase of the amount of queries. This clearly shows that generating too many query can lead to a traffic jam in the queues of the message bus. Still, we can conclude (1) that the platform can be executed even on systems with limited resources. However, (2) more powerful hardware is needed for benchmarking Big Linked Data systems.

Table 7: Results of the Platform Benchmark on a Single Machine

Experiments	Experiment 1	Experiment 2	Experiment 3
Data generators	2	2	2
Task generators	1	1	1
Queries	1,000	2,000	5,000
Avg. query runtime (in ms)	10,414	39,789	60,009
Query runtime std. dev.	2,086	2,536	7,832
Overall runtime (in s)	17.9	54.0	140.9
Queries per second	55.9	37.0	35.5

[17] M. Saleem, I. Ali, A. Hogan, Q. Mehmood, and A. Ngonga-Ngomo. LSQ: The Linked SPARQL Queries Dataset. In ISWC, 2015