



# HOBBIT

Newsletter  
September 2017

**HOBBIT**  
Holistic Benchmarking of  
Big Linked Data

H2020 Research and  
Innovation Grant Agreement  
No. 688227

Contact  
Dr. Axel-Cyrille Ngonga  
Ngomo  
Institute for Applied  
Informatics  
Hainstraße 11, 04109  
Leipzig Germany  
Phone: +49 341 97 32362  
ngonga@informatik.uni-  
leipzig.de



Dear Hobbit friends and Community members,

We are happy to announce that HOBBIT platform is running and ready to execute several [benchmarks](#) for Question Answering, Data Storage, Sensor Data, etc, that we have developed during this project. The benchmarks have been used in open challenges on major Semantic Web conferences, such as: [Mighty Storage \(MOCHA\)](#), [Query Answering over Linked Data \(QALD\)](#), [Open Knowledge Extraction \(OKE\)](#) and [DEBS Grand Challenge](#).

## The HOBBIT PLATFORM

The HOBBIT evaluation platform is a distributed FAIR benchmarking platform for the Linked Data lifecycle. We offer an open source evaluation platform that can be downloaded and executed locally. Secondly, we offer an online instance of the platform for a) running public challenges and b) making sure that even people without the required infrastructure are able to run the benchmarks they are interested in.

The online instance of the HOBBIT benchmarking platform is accessible at [master.project-hobbit.eu](http://master.project-hobbit.eu). More information about its features and how to use it can be found in our deliverable [D2.2.1](#) and our [platform wiki](#). The code of the platform and the benchmarks are accessible at <https://github.com/hobbit-project>. All benchmarks are also available on our CKAN: <https://ckan.project-hobbit.eu/dataset>, along with their source code and related publications.

## BENCHMARKS

### Benchmark I - Generation & Acquisition

Hobbit provides benchmarks to measure the performance of SPARQL query processing systems when faced with streams of data from industrial machinery in terms of efficiency and completeness. We aim to reflect real loads on triple stores used in real applications. We hence use the following datasets:

- Public Transport Data
- Social network data from Twitter (Twitter Data)
- Car traffic data gathered from sensor data gathered by TomTom (Transportation Data)
- Sensor data from plastic injection moulding industrial plants of Weidmüller (Molding Machines Data)

Hobbit also provides benchmarks to measure the performance of extraction systems for unstructured streams of natural-language data. Extraction tasks comprise the recognition of known and unknown entities inside the data as well as the linking to different knowledge bases. For the benchmarks the following datasets are integrated:

- curated unstructured datasets by experts from real sources

- produced unstructured data streams by Bengal, a generic data generator

Read more information here: <https://project-hobbit.eu/outcomes/benchmark-i-generation-acquisition/>

#### [Benchmark II - Analysis & Processing](#)

HOBBIT provides benchmarks for the linking and analysis of Linked data. The linking benchmark tests the performance of instance matching methods and tools for Linked Data while the analytics benchmark tests the performance of machine learning methods (supervised and unsupervised) for data analytics.

- Linking: Linking spatial resources requires techniques that differ from the classical mostly string-based approaches. We propose two benchmarks that deal with link discovery for spatial data:
  - the *Linking Benchmark* that is based on SPIMBENCH, to test the performance of Instance Matching tools that implement string-based approaches for identifying matching entities
  - the *Spatial Benchmark* that can be used to test the performance of systems that deal with topological relations proposed in the state of the art DE-9IM (Dimensionally Extended nine-Intersection Model) model.
- Machine Learning on Structured Data: Given graph data by USU and AGT, we investigate supervised and unsupervised machine learning methods.

Read more information here: <https://project-hobbit.eu/outcomes/benchmark-ii-analysis-processing/>

#### [Benchmark III – Storage & Curation](#)

HOBBIT provides benchmarks for storage and curation systems. The Data Storage benchmark focuses on the typical challenges faced by the storage systems. The Versioning Benchmark aims to test the ability of versioning systems to efficiently manage evolving Linked Data datasets and queries evaluated across multiple versions of these datasets.

Read more information here: <https://project-hobbit.eu/outcomes/benchmark-iii-storage-curation/>

#### [Benchmark IV - Visualisation & Services](#)

HOBBIT provides benchmarks for the Visualization and Service part of the Big Data Value Chain. The focus is on benchmarks with well defined metrics that do not involve users. More specifically, we are working on benchmarks regarding a) question answering and b) faceted browsing. The project does not benchmark user interfaces themselves but focus on providing performance and accuracy measurements for approaches used in interfaces.

Read more information here: <https://project-hobbit.eu/outcomes/benchmark-iv-visualisation-services/>

#### [OPEN CALL FOR BENCHMARKS](#)

To drive the Semantic Web and Big Linked Data community towards transparent performance evaluation using standardized hardware (i.e. by running on HOBBIT platform), we ask you to add your benchmarks to the HOBBIT CKAN instance. Just follow the steps described in our Open Call for Benchmarks (<https://project-hobbit.eu/open-call-for-benchmarks>).

Join our community to remain informed at [project-hobbit.eu/get-involved](https://project-hobbit.eu/get-involved). If you have any questions, do not hesitate to contact us as

explained at [project-hobbit.eu/contacts](http://project-hobbit.eu/contacts). We are looking forward to benchmarking your system soon. Stay tuned!

Best,  
HOBBIT team.

---



Share



Tweet



Forward