

# *Unified Benchmarking of Big Data Platforms*

## *The HOBBIT Platform*

Axel-Cyrille Ngonga Ngomo

Horizon 2020  
GA No 688227  
01/12/2016–30/11/2018



Apache Big Data  
Sevilla, Spain  
November 11, 2016

## 40 ZETTABYTES

(40 TRILLION GIGABYTES)  
of data will be created by 2020, an increase of 300 times from 2005

6 BILLION PEOPLE have cell phones

WORLD POPULATION: 7 BILLION

## Volume SCALE OF DATA

It's estimated that **2.5 QUINTILLION BYTES** (2.5 TRILLION GIGABYTES) of data are created each day

Most companies in the U.S. have at least **100 TERABYTES** (100,000 GIGABYTES) of data stored

The New York Stock Exchange captures

**1 TB OF TRADE INFORMATION** during each trading session



## Velocity ANALYSIS OF STREAMING DATA

By 2016, it is projected there will be

**18.9 BILLION NETWORK CONNECTIONS**

—almost 2.5 connections per person on earth



Modern cars have close to **100 SENSORS** that monitor items such as fuel level and tire pressure



# The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015 **4.4 MILLION IT JOBS** will be created globally to support big data, with 1.9 million in the United States



As of 2011, the global size of data in healthcare was estimated to be

**150 EXABYTES** (150 TRILLION GIGABYTES)



**30 BILLION PIECES OF CONTENT** are shared on Facebook every month



## Variety DIFFERENT FORMS OF DATA

By 2014, it's anticipated there will be

**420 MILLION WEARABLE, WIRELESS HEALTH MONITORS**

**4 BILLION+ HOURS OF VIDEO** are watched on YouTube each month



**400 MILLION TWEETS** are sent per day by about 200 million monthly active users



**1 IN 3 BUSINESS LEADERS** don't trust the information they use to make decisions



Poor data quality costs the US economy around

**\$3.1 TRILLION A YEAR**



**27% OF RESPONDENTS**

## Veracity UNCERTAINTY OF DATA

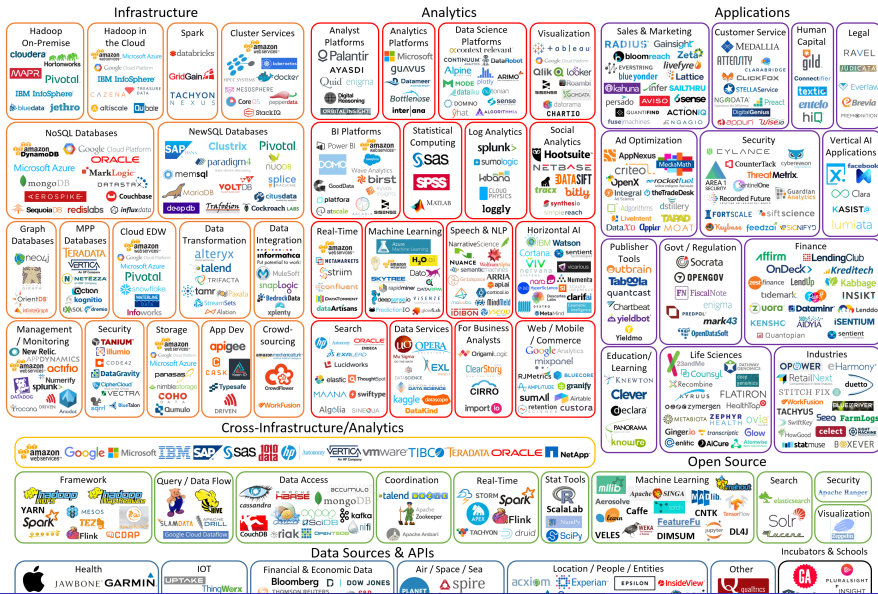
in one survey were unsure of how much of their data was inaccurate

## The Dataflog Open Source Landscape 2.0



<sup>2</sup><https://cloudrablings.me/>

## Big Data Landscape 2016 (Version 3.0)





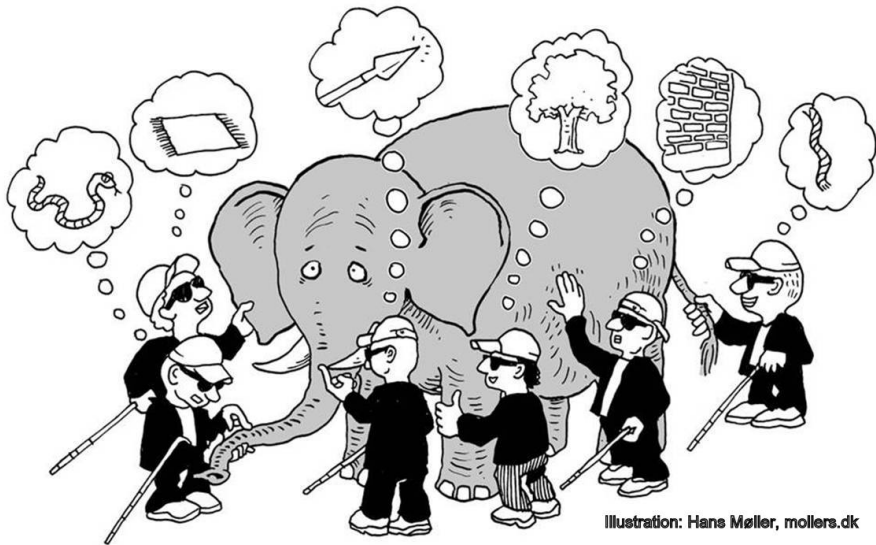


Illustration: Hans Møller, mollers.dk

<sup>4</sup>[https://steemit.com/philosophy/@10k1/  
subjectivity-and-truth-how-blockchains-model-consensus-building](https://steemit.com/philosophy/@10k1/subjectivity-and-truth-how-blockchains-model-consensus-building)

- Developers: How good is my tool?
- Vendors: Who is my tool good for?
- Users: Which tool(s) should I use for my application?



- Where are the current bottlenecks?
- Which steps of the data lifecycle are critical?
- Which solutions are available?
- Which key performance indicators are relevant?
- How well do or should tools perform?
- How do existing solutions perform w.r.t. relevant indicators?



# bench·mark

/ˈben(t)SHmärk/

*noun*

1. a standard or point of reference against which things may be compared or assessed.  
"a benchmark case"  
*synonyms:* **standard**, point of reference, **gauge**, **guide**, **guideline**, guiding principle, **norm**, **touchstone**, **yardstick**, **barometer**, **indicator**, **measure**, **model**, **exemplar**, **pattern**, **criterion**, **specification**, **convention**  
"the settlement became the benchmark for all future negotiations"
2. a surveyor's mark cut in a wall, pillar, or building and used as a reference point in measuring altitudes.

*verb*

### Components

- Dataset(s), e.g., Twitter stream, sensor data
- Task(s), i.e., NER, NEL, ingestion
- Key Performance Indicators, e.g., precision, recall

# bench·mark

/ˈben(t)SHmärk/

*noun*

1. a standard or point of reference against which things may be compared or assessed.  
"a benchmark case"  
*synonyms:* **standard**, point of reference, **gauge**, **guide**, **guideline**, guiding principle, **norm**, **touchstone**, **yardstick**, **barometer**, **indicator**, **measure**, **model**, **exemplar**, **pattern**, **criterion**, **specification**, **convention**  
"the settlement became the benchmark for all future negotiations"
2. a surveyor's mark cut in a wall, pillar, or building and used as a reference point in measuring altitudes.

*verb*

# Challenges

## Dataset Mismatch

	Year	ACE	Wiki	AQUAINT	MSNBC	IITB	Meij	AIDA/CoNLL	N <sup>3</sup> collection	KORE 50	Wiki-Disamb30	Wiki-Annot30	Spotlight Corpus	SemEval-2013 task 12	SemEval-2007 task 7	SemEval-2007 task 17	Senseval-3	NIF-based corpus	Microposts2014	Software available?	Webservice available?	
Cucerzan	2007				✓																	
Wikipedia Miner	2008			✓*																		✓
Illinois Wikifier	2011	✓	✓	✓*	✓																✓	✓
Spotlight	2011												✓								✓	✓
AIDA	2011							✓													✓	✓
TagMe 2	2012										✓	✓									✓	✓
Dexter	2013																				✓	✓
KEA	2013																				✓	✓
WAT	2013																				✓	✓
AGDISTIS	2014			✓	✓	✓	✓	✓	✓	✓			✓								✓	✓
Babelfy	2014							✓		✓				✓	✓	✓	✓				✓	✓
NERD-ML	2014							✓											✓		✓	✓
BAT-Framework	2013	✓	✓	✓	✓	✓	✓	✓*													✓	
NERD Framework	2014						✓	✓											✓		✓	✓
GERBIL	2014	✓	✓	✓	✓	✓	✓	✓*	✓	✓			✓					✓	✓		✓	✓

# Challenges

## *Unclear KPI Semantics*



### Example

- Which time do we measure?
  - First or last result?
  - With or without network delay?





On August 20, 2005, in a private ceremony, Thompson's ashes were fired from a cannon a top a tower of his own design (in the shape of a double-thumbed fist clutching a peyote button) to the tune of Norman Greenbaum's "Spirit in the Sky" and Bob Dylan's Mr. Tambourine Man. Red, white, blue, and green fireworks were launched along with his ashes. As the city of Aspen would not allow the cannon to remain for more than a month, the cannon has been dismantled and put into storage until a suitable permanent location can be found. According to his widow Anita, Thompson's funeral was financed by actor Johnny Depp, a close friend of Thompson. Depp told the Associated Press, "All I'm doing is trying to make sure his last wish comes true. I just want to send my pal out the way he wants to go out." Other famous attendees at the funeral included U.S. Senator John Kerry and former U.S. Senator George McGovern; 60 Minutes correspondent Ed Bradley and Charlie Rose; actors Jack Nicholson, Bill Murray, Benicio del Toro, Sean Penn, and Josh Hartnett; singers Lyle Lovett, John Oates and numerous other friends. An estimated 280 people attended the funeral.

The plans for this monument were initially drawn by Thompson and Ralph Steadman and were shown as part of an Omnibus program on the BBC entitled Fear and Loathing in GonzoVision (1978). It is included as a special feature on the second disc of the 2003 Criterion Collection DVD release of Fear and Loathing in Las Vegas (labeled on the DVD as "Fear and Loathing on the Road to Hollywood"). The video footage of Steadman and Thompson drawing the plans and outdoor footage showing where he wanted the cannon constructed were planned prior to the unveiling of his cannon at the funeral.

### Example

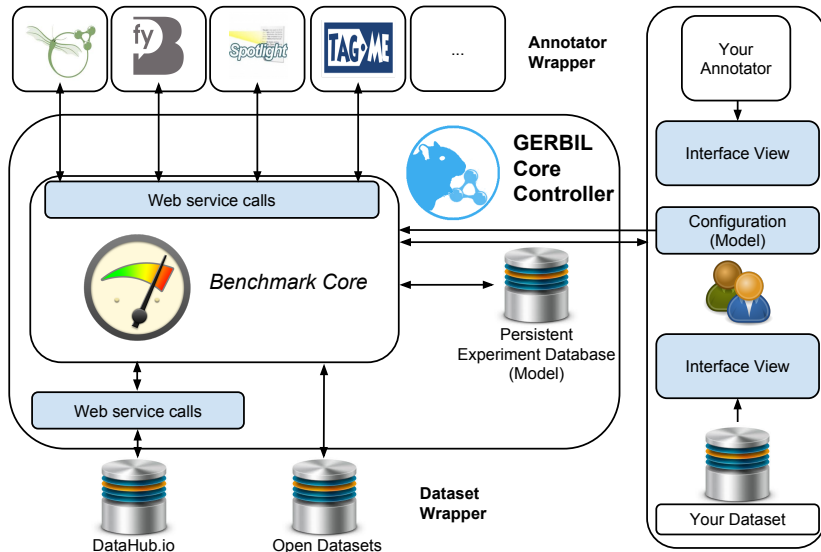
- When is an annotation correct?
  - Weak or strong annotation?
  - Semantically equivalent or exact URI?

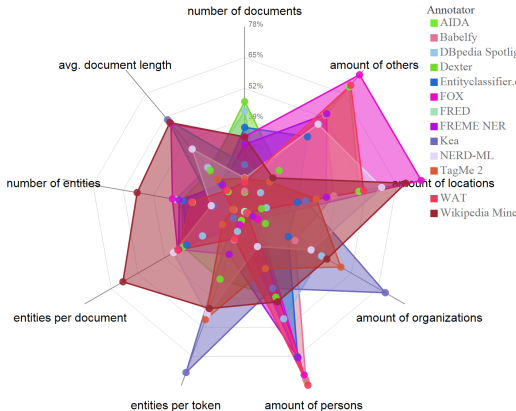
On August 20, 2005, in a private ceremony, Thompson's ashes were fired from a cannon a top a tower of his own design (in the shape of a double-thumbed fist clutching a peyote button) to the tune of Norman Greenbaum's "Spirit in the Sky" and Bob Dylan's Mr. Tambourine Man. Red, white, blue, and green fireworks were launched along with his ashes. As the city of Aspen would not allow the cannon to remain for more than a month, the cannon has been dismantled and put into storage until a suitable permanent location can be found. According to his widow Anita, Thompson's funeral was financed by actor Johnny Depp, a close friend of Thompson. Depp told the Associated Press, "All I'm doing is trying to make sure his last wish comes true. I just want to send my pal out the way he wants to go out." Other famous attendees at the funeral included U.S. Senator John Kerry and former U.S. Senator George McGovern; 60 Minutes correspondent Ed Bradley and Charlie Rose; actors Jack Nicholson, Bill Murray, Benicio del Toro, Sean Penn, and Josh Hartnett; singers Lyle Lovett, John Oates and numerous other friends. An estimated 280 people attended the funeral.

The plans for this monument were initially drawn by Thompson and Ralph Steadman and were shown as part of an Omnibus program on the BBC entitled Fear and Loathing in Gonzo vision (1978). It is included as a special feature on the second disc of the 2003 Criterion Collection DVD release of Fear and Loathing in Las Vegas (labeled on the DVD as "Fear and Loathing on the Road to Hollywood"). The video footage of Steadman and Thompson drawing the plans and outdoor footage showing where he wanted the cannon constructed were planned prior to the unveiling of his cannon at the funeral.

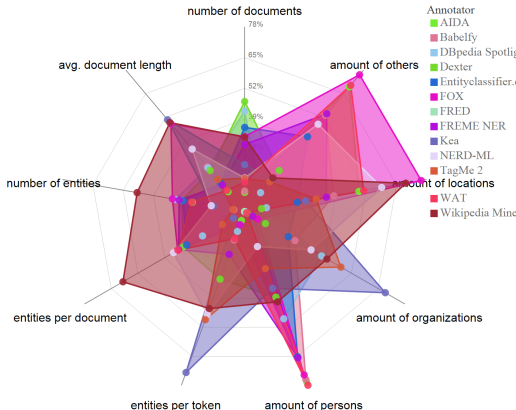
# Solution

## Unified Benchmarking Framework

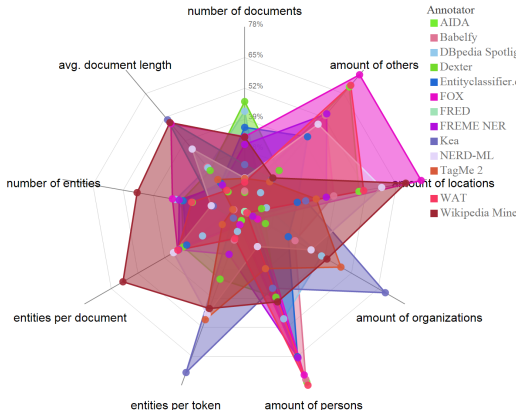




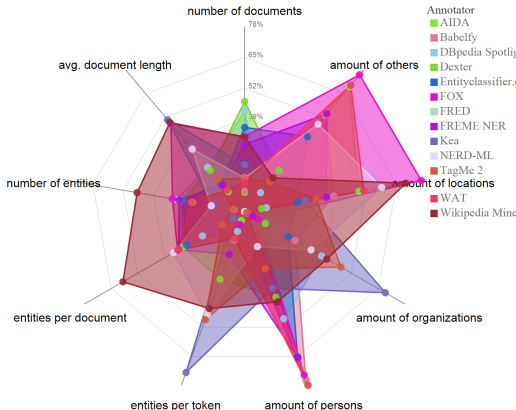
- Evaluation platform for NER/NEL
- 18 reference annotation systems
- 32 reference datasets
- Benchmarking 10× faster



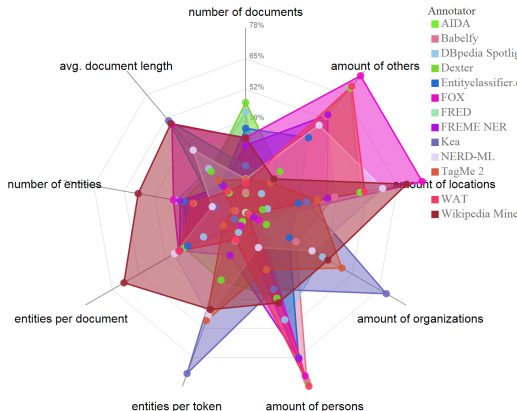
- Evaluation platform for NER/NEL
- 18 reference annotation systems
- 32 reference datasets
- Benchmarking 10× faster
- Archiving of results
- Citeable URIs
- Additional analysis



- Evaluation platform for NER/NEL
- 18 reference annotation systems
- 32 reference datasets
- Benchmarking 10× faster
- Archiving of results
- Citeable URIs
- Additional analysis
- Open-source project
- Local deployment
- Normalized implementation of KPIs
- Online instance
- Feedback for developers and users



Annotator	Tasks
NIF-based Annotators	2519
Babelfy	958
DBpedia Spotlight	922
TagMe 2	811
WAT	787
Kea	763
Wikipedia Miner	714
NERD-ML	639
Dexter	587
AGDISTIS	443
Entityclassifier.eu NER	410
FOX	352
Cetus	1
Overall	24.3K exps 50+ papers





## *Rationale*



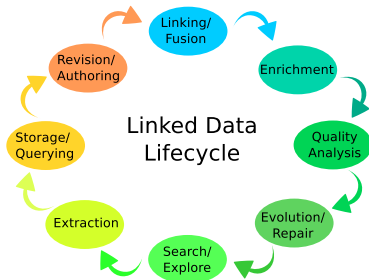
# HOBBIT

A community-driven benchmarking framework for the community

*Rationale*

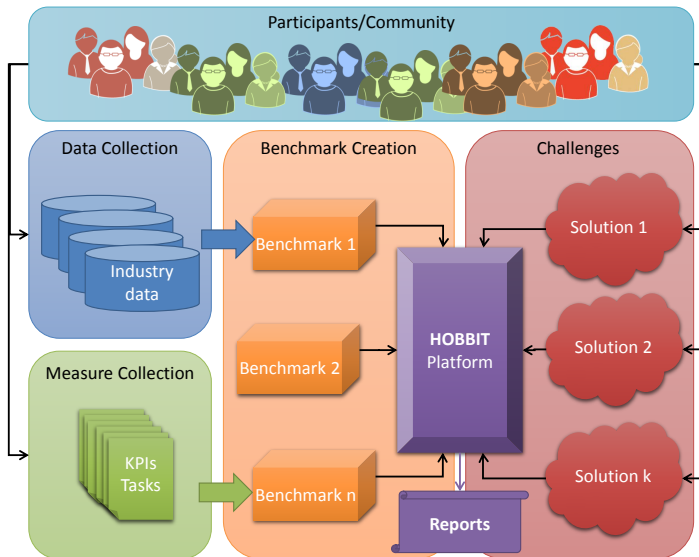
A community-driven benchmarking framework for the community

- Focus on Big (Linked) Data
- Build upon 24.3K experiments performed with GERBIL
- Cover all steps of the Linked Data lifecycle
  - Used by a growing number of companies
  - Mature and maturing technologies
- Open benchmarks based on industrial data and use cases



- ① Gather real requirements
  - Performance indicators
  - Performance thresholds
- ② Develop benchmarks based on real data
- ③ Provide universal benchmarking platform
  - Standardized hardware
  - Comparable results
- ④ Periodic benchmarking challenges
- ⑤ Periodic reporting
- ⑥ Found independent HOBBIT association





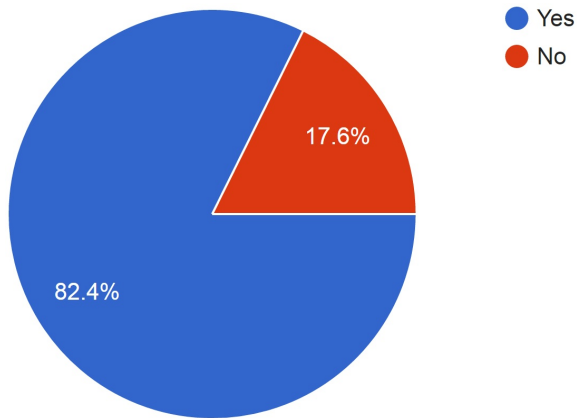
## Questions

- In what areas are organizations active?
- What do people expect from benchmarks?
- How are benchmarks being used?



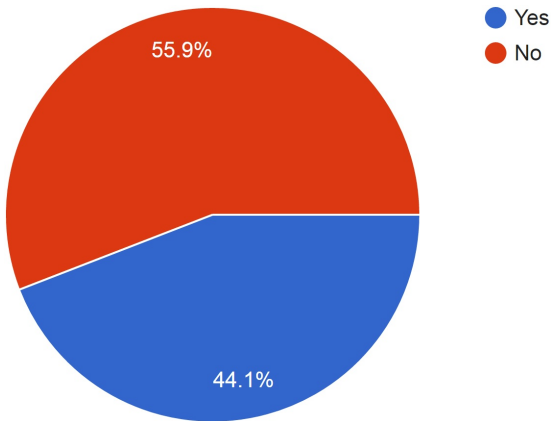
Profile	Count
Solution providers	56
Technology users	67
Scientific community	65

*Can your solution be benchmarked?*



*Do you benchmark your solution?*

- Own datasets and settings in many cases
- Own implementations of measures
- Results not comparable





Health



Food & Agriculture



Energy



Transport



Climate



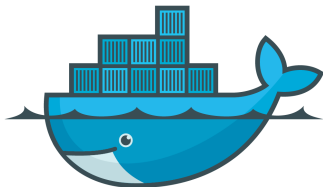
Social Sciences



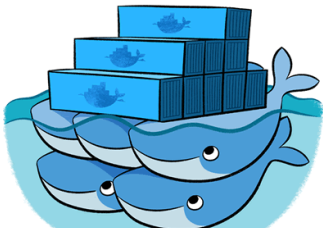
Security



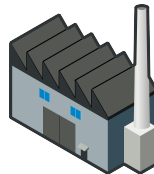
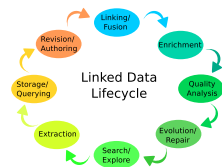
- Uses established deployment technologies (Docker)
  - Decoupled components
  - Benchmark and Systems can be written in different languages
- Uses scalable message queues for communication
- Open-source implementation
- Supports distributed benchmarks and systems
- Online instance on server cluster



 RabbitMQ



- Addresses all steps of the Linked Data Lifecycle
- Benchmarks derived from industry use cases
- Real data under the benchmarks
- Scalable size of benchmarks
- Open-source implementation



- Streaming and static deterministic benchmarks
- Realistic benchmarks
- Controlled volume and velocity

### Generation and Acquisition

- Conversion of XML into RDF
- Entity recognition and linking
- Relation extraction

### Analysis and Processing

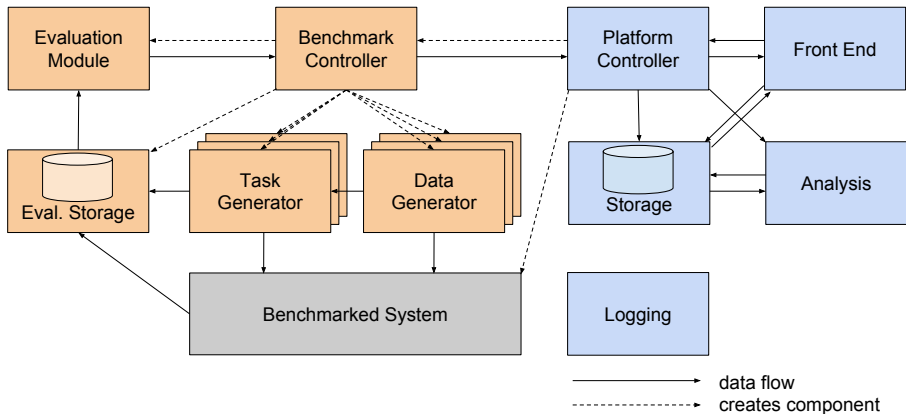
- Link Discovery
- Machine Learning
- Supervised and unsupervised

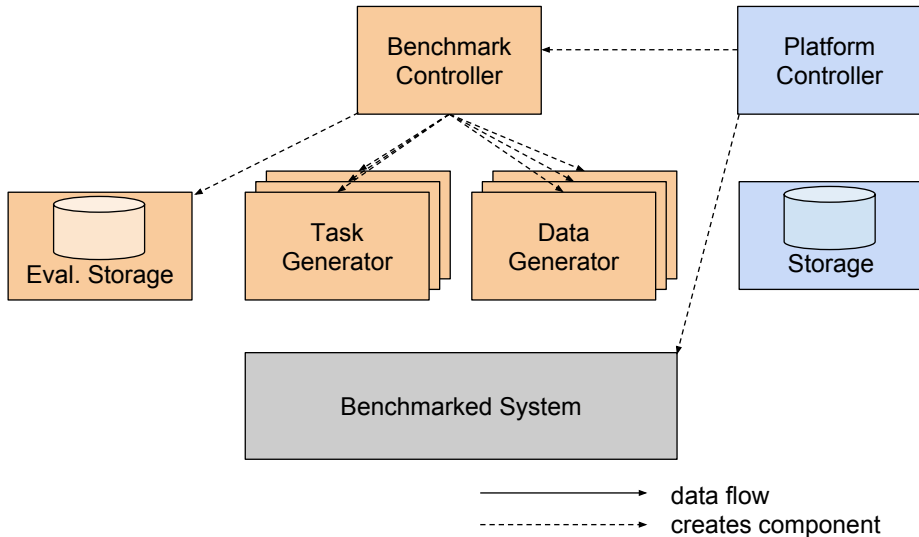
### Storage and Curation

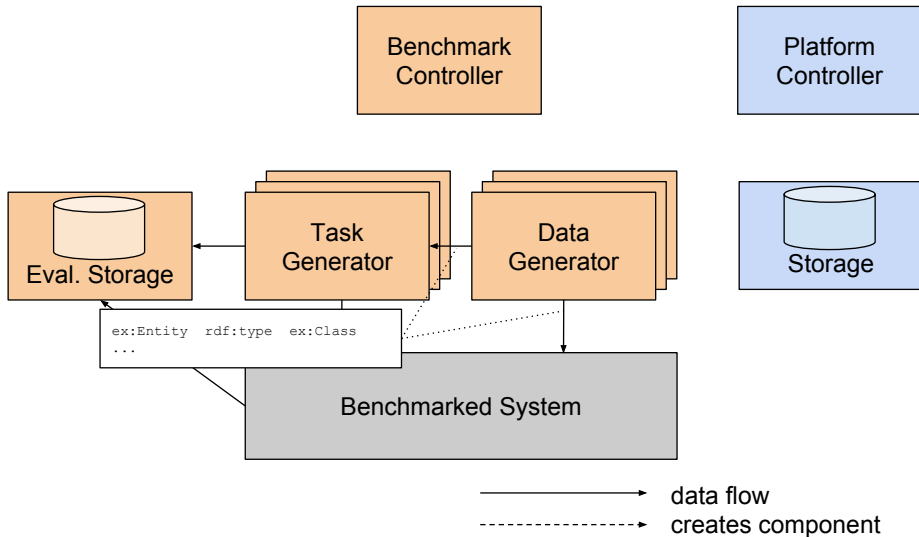
- Triple stores
- Versioning
- Incl. updates

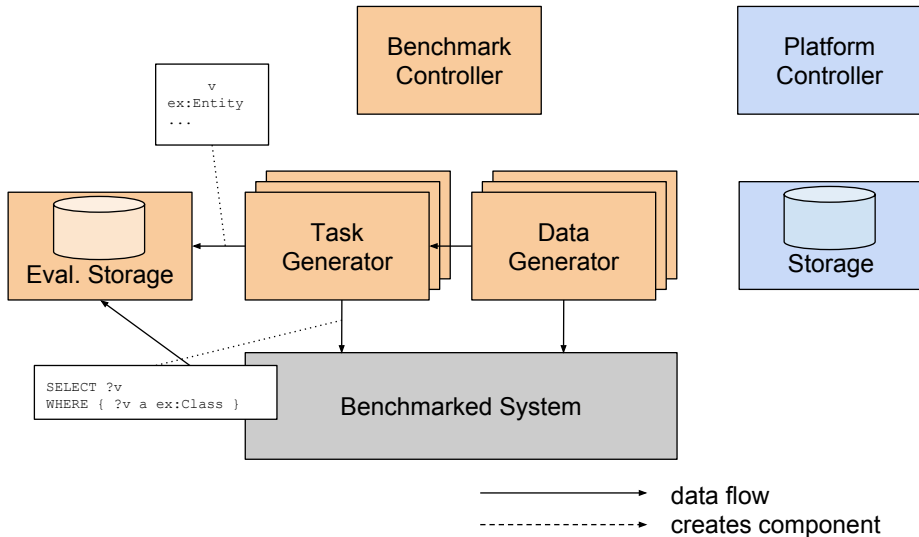
### Visualization and Services

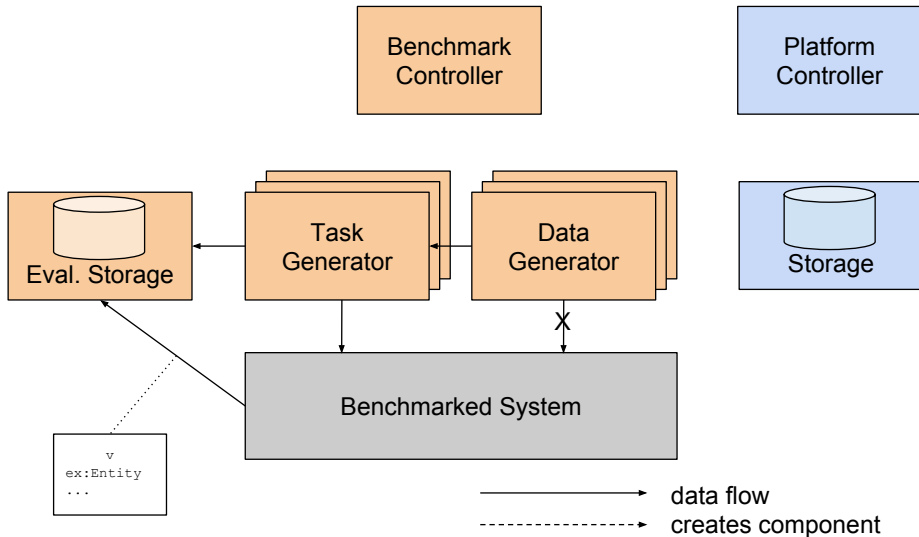
- Question Answering
- Faceted Browsing
- Usage-based benchmarks



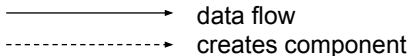
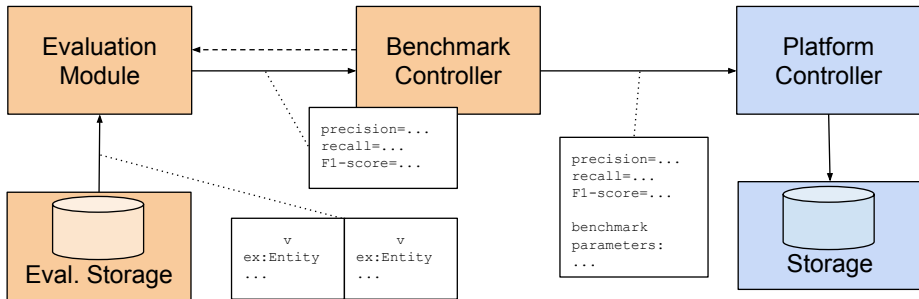




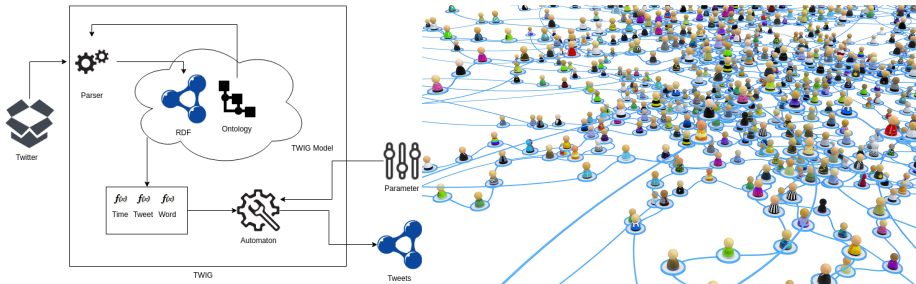




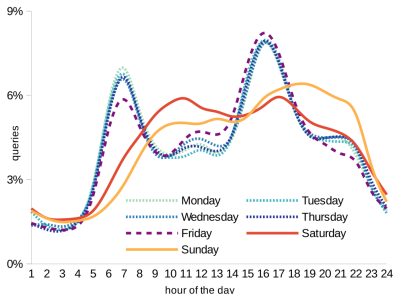




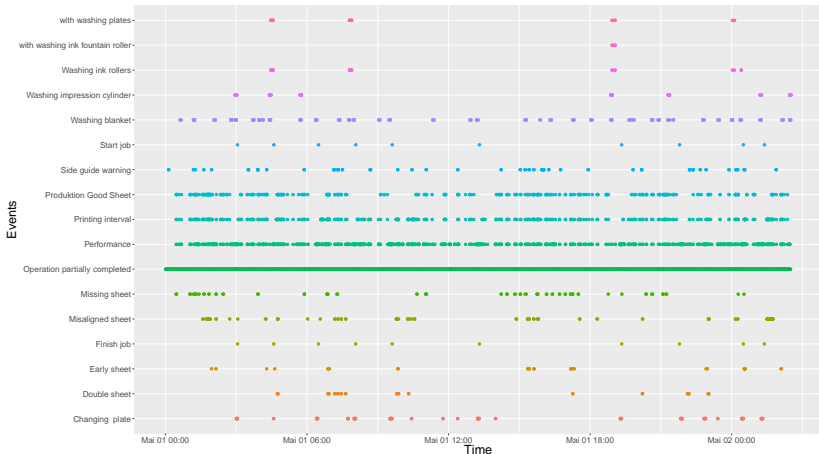
- Goal: Simulate real Twitter Firehose
- Relies on 476 million tweets as training data
- Mimicking algorithm based on
  - Distribution of character frequencies
  - Distribution of transportation frequency
  - Network topology



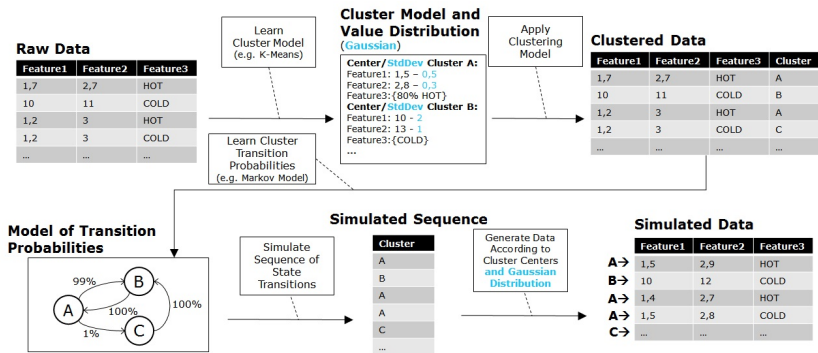
- Goal: Simulate real transport network
- Real transportation data from Belgium for training
- Mimicking algorithm based on
  - Observed correlation between population density and transportation
  - Distribution of transportation frequency
  - Network topology



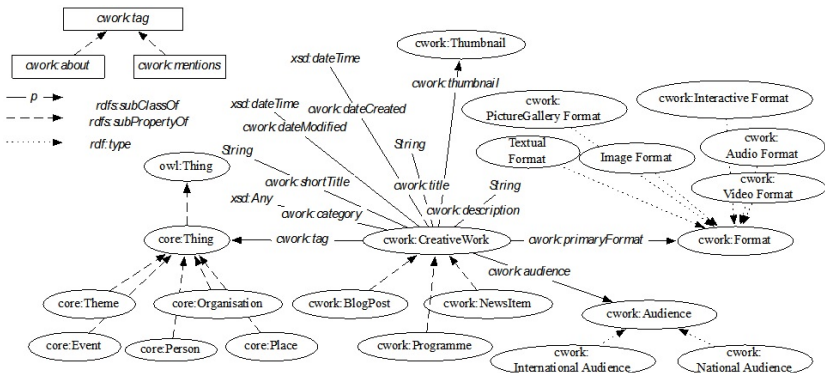
- Goal: Simulate events from printing machinery
- Mimicking algorithm using event correlations and distributions



- Goal: Simulate events from injection molding machinery
- Mimicking algorithm using event correlations and distributions



- Goal: Simulate data from the BBC
- Generator based on manually configurable set of correlations



- Join the HOBBIT community
- Provide KPIs
- Provide datasets
- Join the platform development
- Follow us on Twitter



[https://twitter.com/hobbit\\_project](https://twitter.com/hobbit_project)

Be  
part  
of it.

- Streaming and static deterministic benchmarks
- Realistic benchmarks
- Controlled volume and velocity

## Generation and Acquisition

- Conversion of XML into RDF
- Entity recognition and linking
- Relation extraction

## Analysis and Processing

- Link Discovery
- Machine Learning
- Supervised and unsupervised

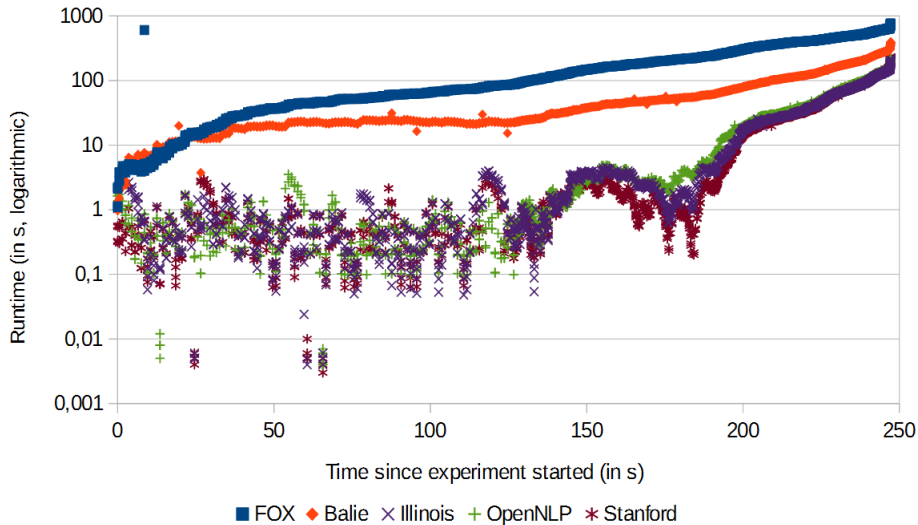
## Storage and Curation

- Triple stores
- Versioning
- Incl. updates

## Visualization and Services

- Question Answering
- Faceted Browsing
- Usage-based benchmarks





System	Precision	Recall	F1-measure
FOX	<b>0.515</b>	<b>0.310</b>	<b>0.351</b>
Balie	0.369	0.230	0.249
Illinois	0.500	0.288	0.327
OpenNLP	0.442	0.241	0.285
Stanford	0.486	0.303	0.335

A2KB, weak annotation match, Micro F1-measure

System	AIDA/CoNLL-Comp.	IITB	KORE50	MSNBC	Microp.2014-Train	N3-Reuters-128
AIDA	0.668	0.141	0.625	0.622	0.363	0.391
Babelfy	0.448	0.129	0.564	0.423	0.311	0.289
DBpedia Spotlight	0.545	0.262	0.341	0.457	0.448	0.320
FOX	0.512	0.100	0.268	0.127	0.309	0.518
FREME NER	0.358	0.074	0.160	0.208	0.254	0.263
WAT	0.673	0.137	0.543	0.631	0.403	0.480
xLisa	0.363	0.233	0.352	0.365	0.322	0.274



<http://project-hobbit.eu/get-involved/>



<http://goo.gl/forms/1iRIoG4Xpb>



[https://twitter.com/hobbit\\_project](https://twitter.com/hobbit_project)



# HOBBIT

Holistic Benchmarking  
of Big Linked Data

This work was supported by grants from the EU H2020 Framework Programme provided for the project HOBBIT (GA no. 688227).