HOBBIT
A Brief Overview

Michael Röder, Axel-Cyrille Ngonga Ngomo

DICE research group
Institute of applied Informatics, Leipzig, Germany
University Paderborn, Germany

(Horizon 2020, GA No 688227)

Presentation ZB MED
Paderborn, September 12th, 2019

Overview

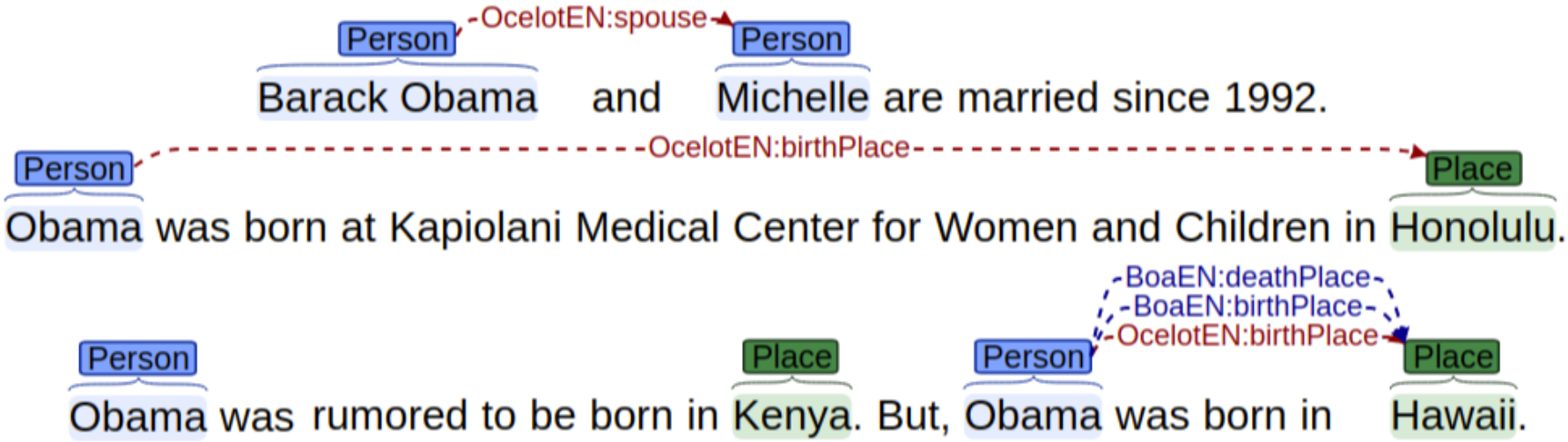
1. Fundamentals
2. Introduction
3. Project Highlights
4. Benchmarking Maching Learning
5. Future Directions

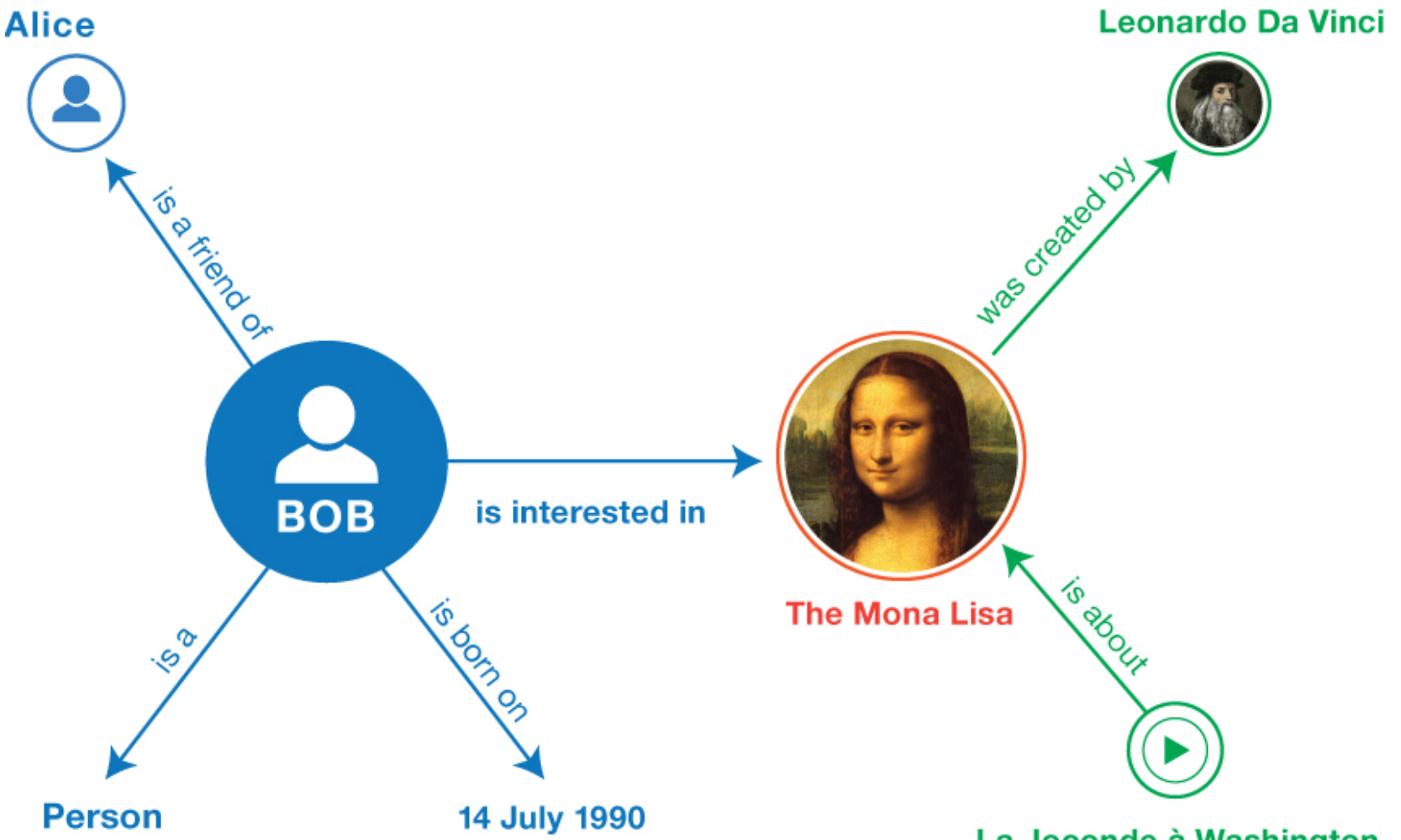
Section 1

Fundamentals

Barack Obama and Michelle are married since 1992. Obama was born at Kapiolani Medical Center for Women and Children in Honolulu. Obama was rumored to be born in Kenya. But, Obama was born in Hawaii.

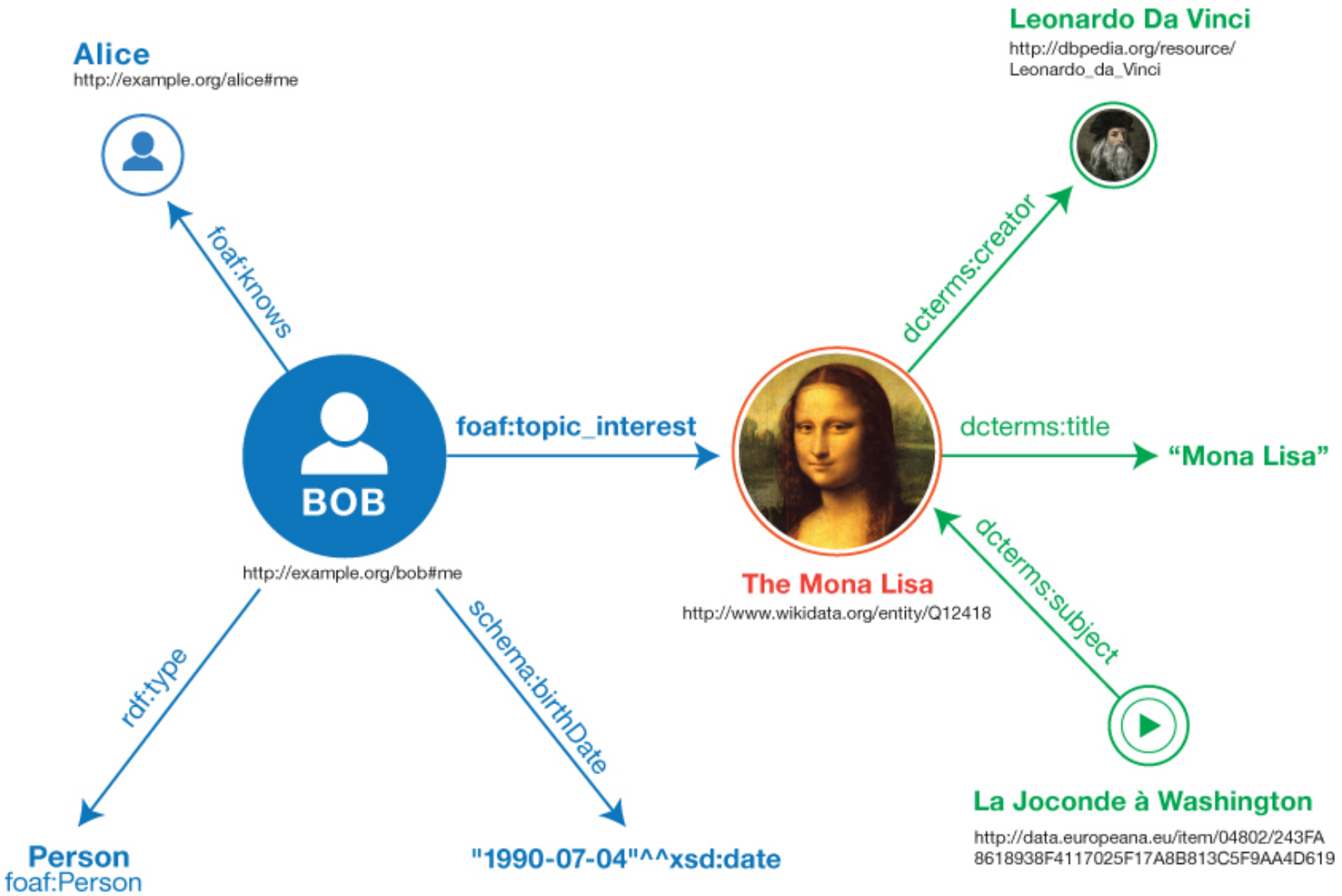
Barack Obama and Michelle are married since 1992. Obama was born at Kapiolani Medical Center for Women and Children in Honolulu. Obama was rumored to be born in Kenya. But, Obama was born in Hawaii.





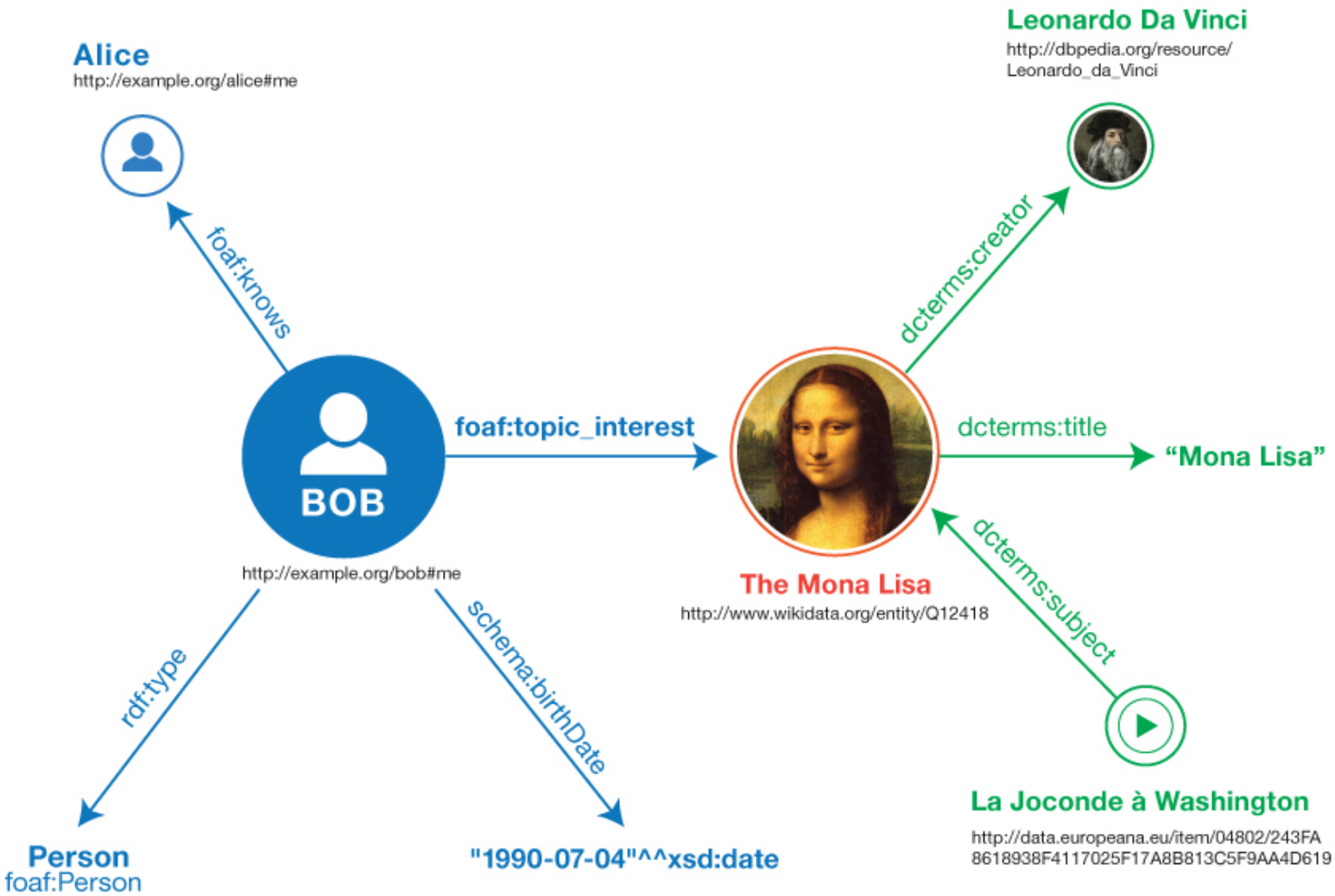
[1]

[1] https://www.w3.org/TR/rdf11-primer/

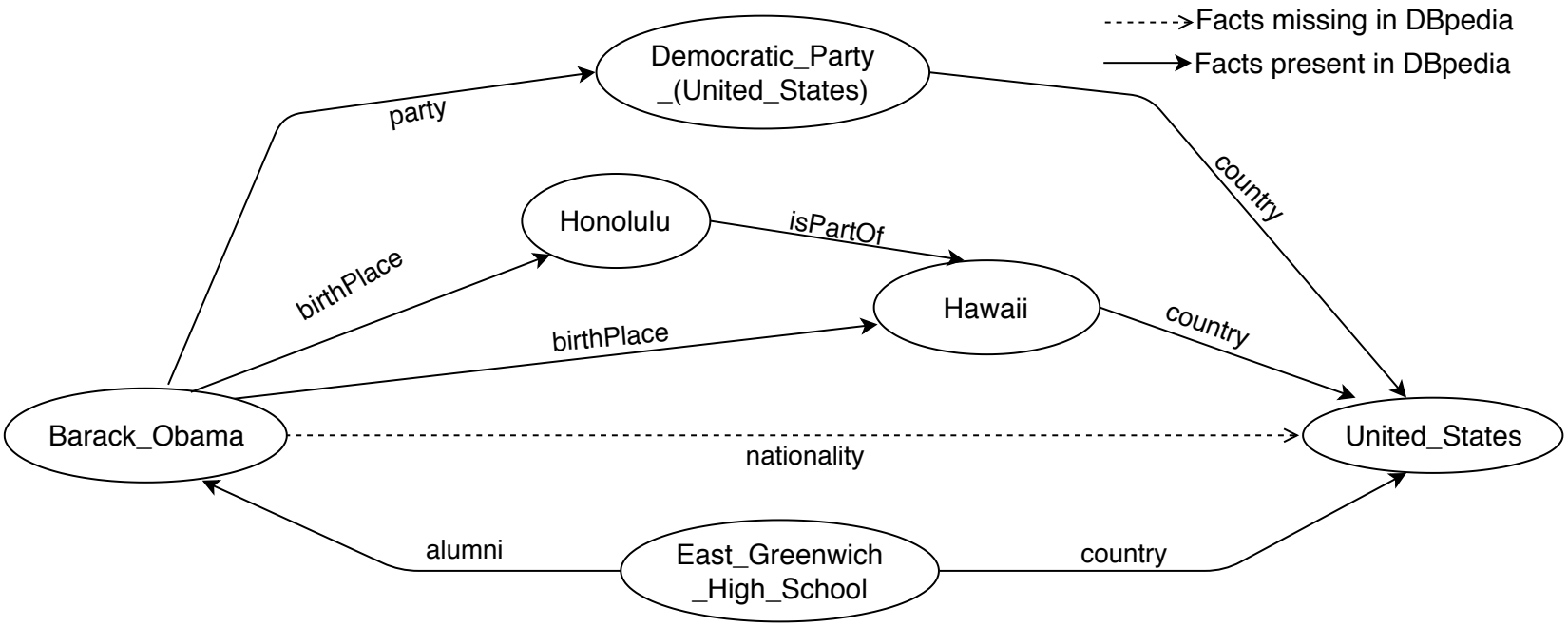


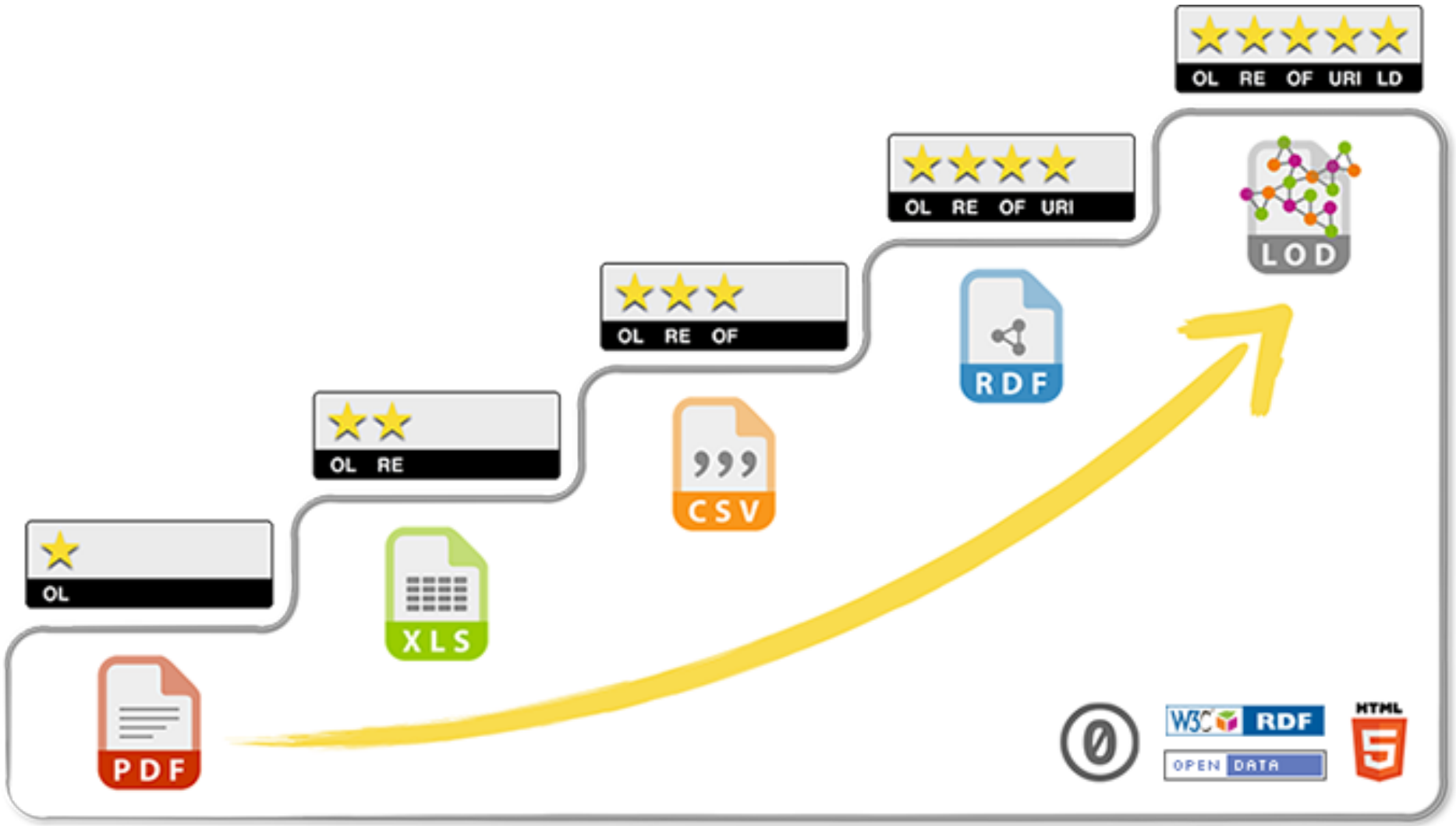
[2]

[2] https://www.w3.org/TR/rdf11-primer/




```
<http://example.org/bob#me> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://xmlns.com/foaf/0.1/Person> .
<http://example.org/bob#me> <http://xmlns.com/foaf/0.1/knows> <http://example.org/alice#me> .
<http://example.org/bob#me> <http://schema.org/birthDate> "1990-07-04"^^<http://www.w3.org/2001/XMLSchema#date> .
...
```

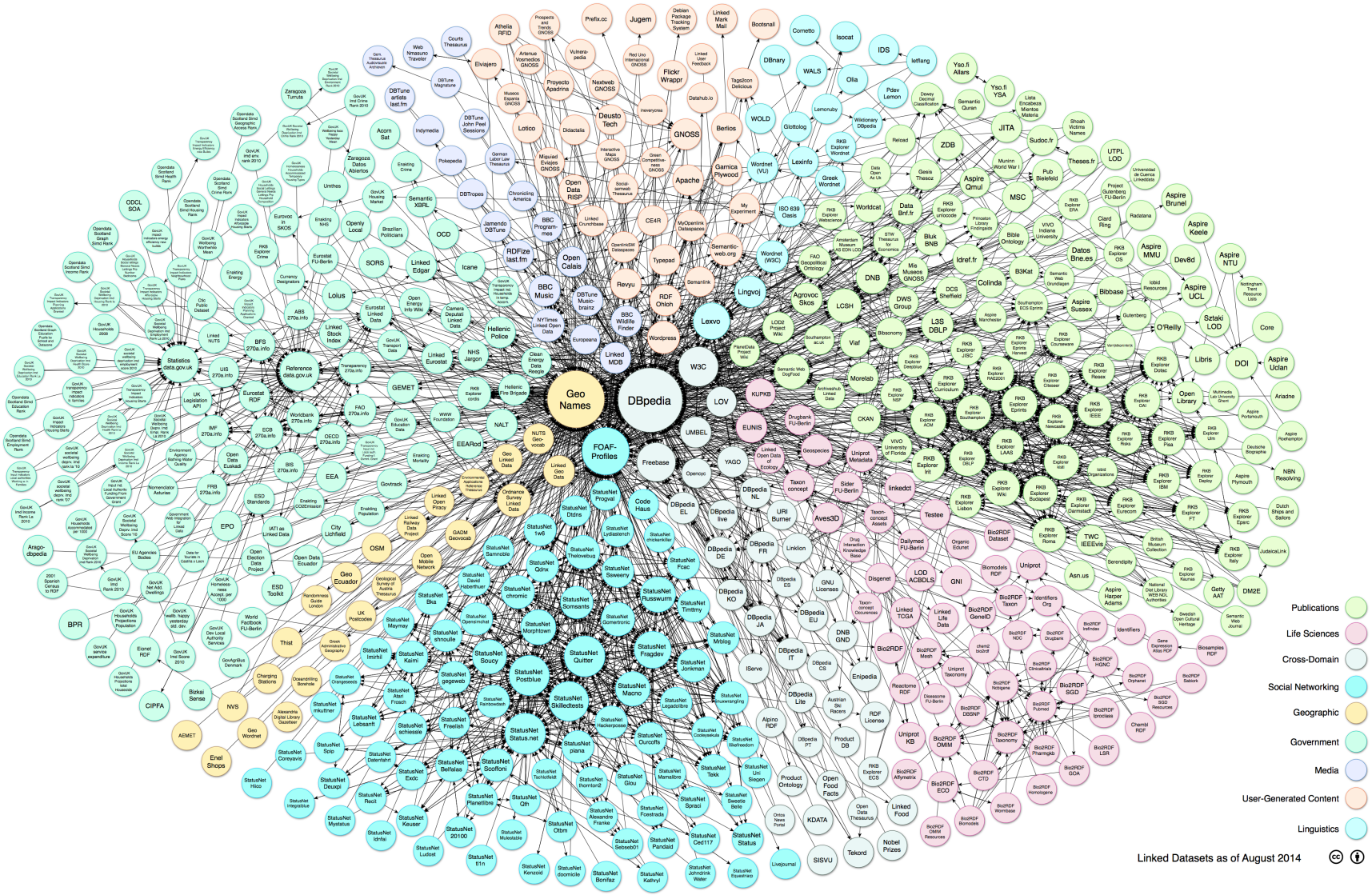

Facts missing in DBpedia
Facts present in DBpedia
Democratic_Party
_(United_States)
party
country
Honolulu
isPartOf
birthPlace
Hawaii
country
birthPlace
Barack_Obama
United_States
nationality
alumni
East_Greenwich
_High_School
country


OL
OL RE
OL RE OF
OL RE OF URI
OL RE OF URI LD
PDF
XLS
CSV
RDF
LOD
W3C RDF
OPEN DATA
HTML 5

Fundamentals

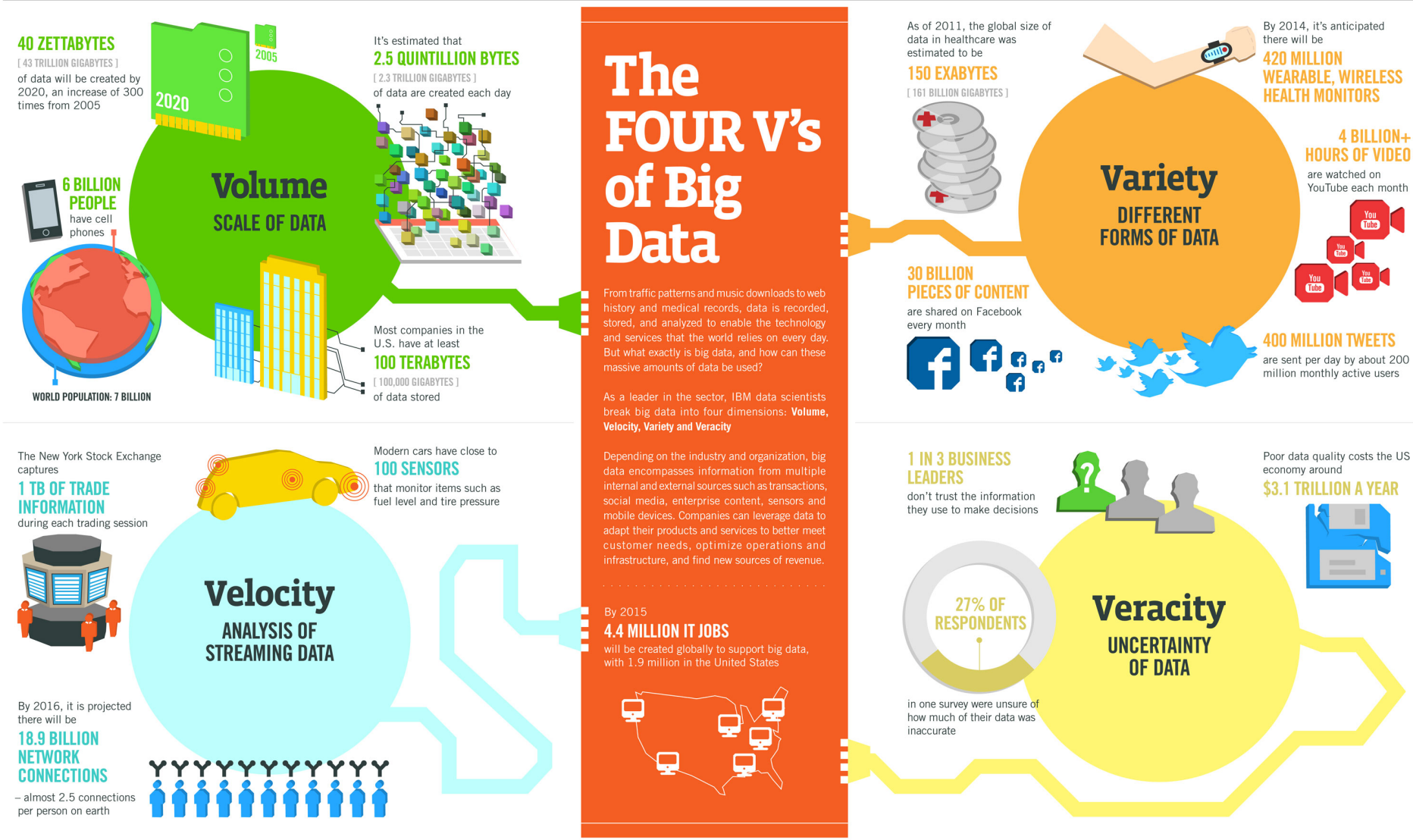
Linked Data

HOBBIT
Holistic Benchmarking of Big Linked Data



Section 2

Introduction



[3]

[3] http://www.ibmbigdatahub.com/infographic/four-vs-big-data

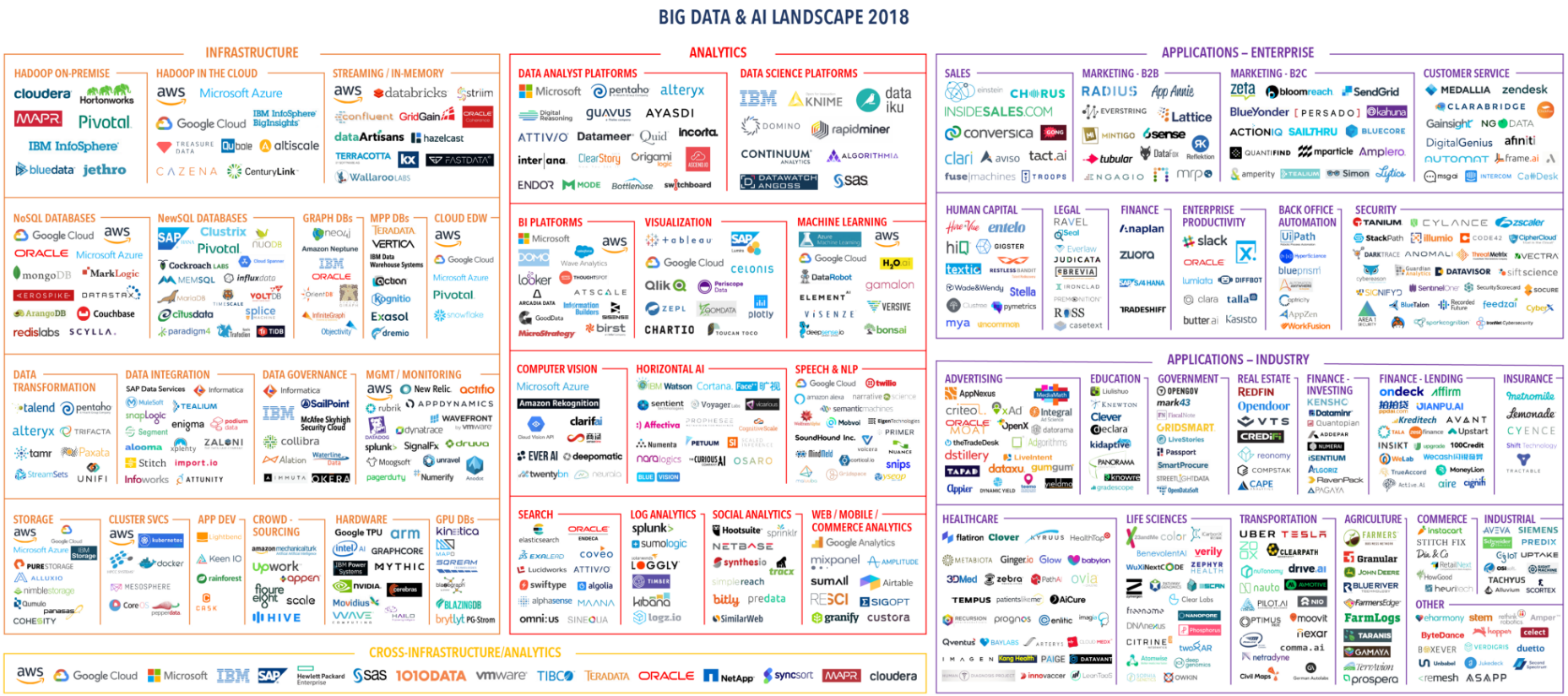
[4]

[4] https://cdn.datafloq.com/cms/os_big_data_open_source_tools-v2.png

Introduction

A Lot of Tools

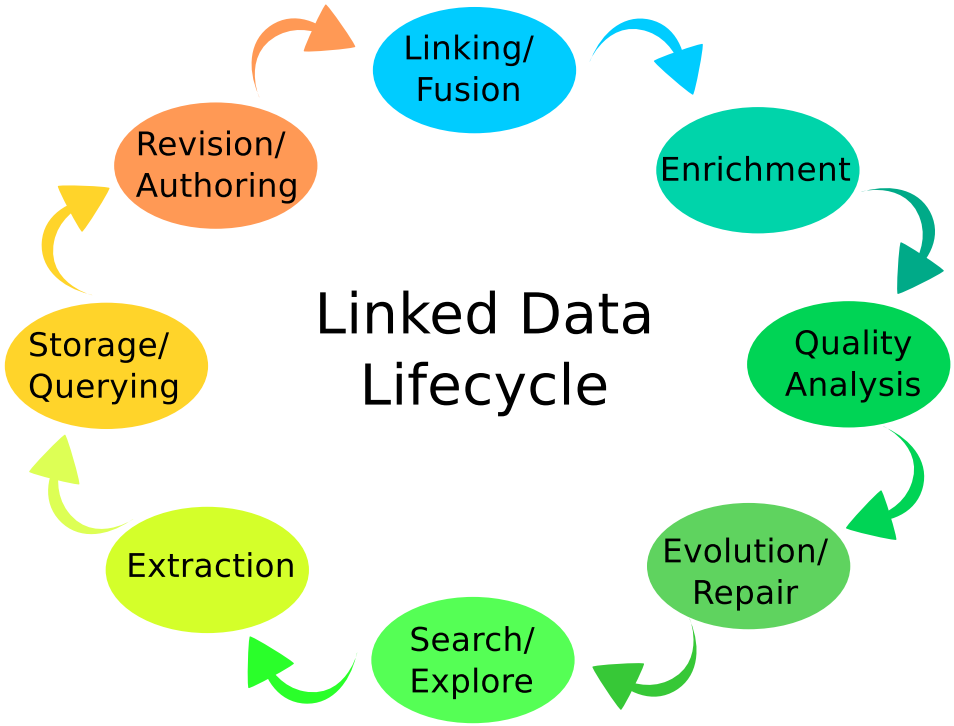
HOBBIT
Holistic Benchmarking of Big Linked Data



Which tool(s) should I use for my use case?

- Which tool(s) should I use for my use case?
- Which key performance indicators are relevant?
- How do existing solutions perform w.r.t. relevant indicators?
- Where are the current bottlenecks?
- Which steps of the data lifecycle are critical?
- ...

- Research project from 2015 – 2018 (Horizon 2020, GA No 688227)
- Focus on Big Linked Data
- Cover the business-critical steps of the Linked Data lifecycle
- Used by a growing number of companies
- Mature and maturing technologies



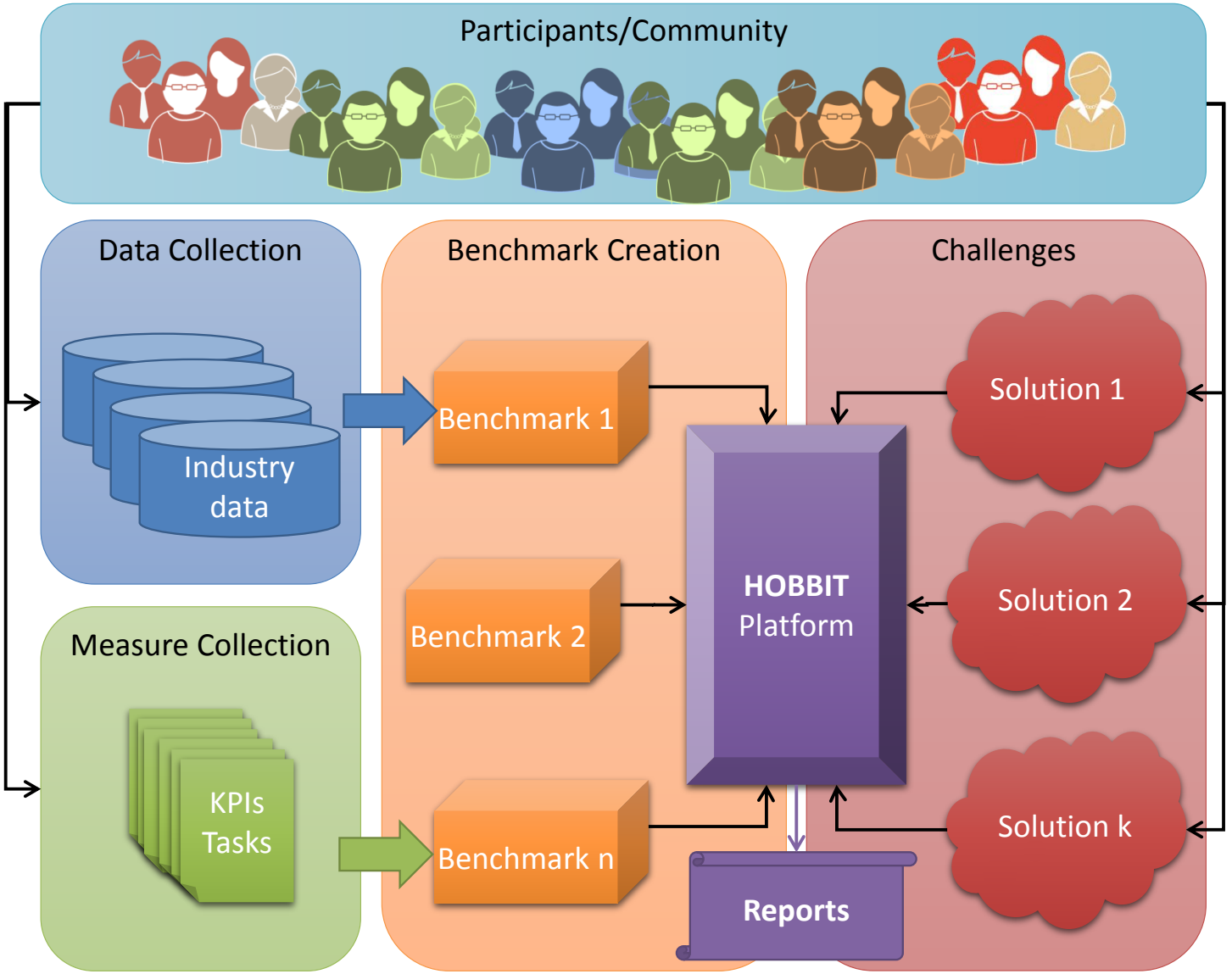
1. Gathered real requirements
   - Focussed on industrial requirements
   - Gathered relevant performance indicators
   - Gathered relevant performance thresholds
   - Gathered real datasets

1. Gathered real requirements
   - Focussed on industrial requirements
   - Gathered relevant performance indicators
   - Gathered relevant performance thresholds
   - Gathered real datasets
2. Developed benchmarks based on real data

1. Gathered real requirements
   - Focussed on industrial requirements
   - Gathered relevant performance indicators
   - Gathered relevant performance thresholds
   - Gathered real datasets
2. Developed benchmarks based on real data
3. Provided universal benchmarking platform
   - Comparable results
   - Hosted as a free-to-use online instance

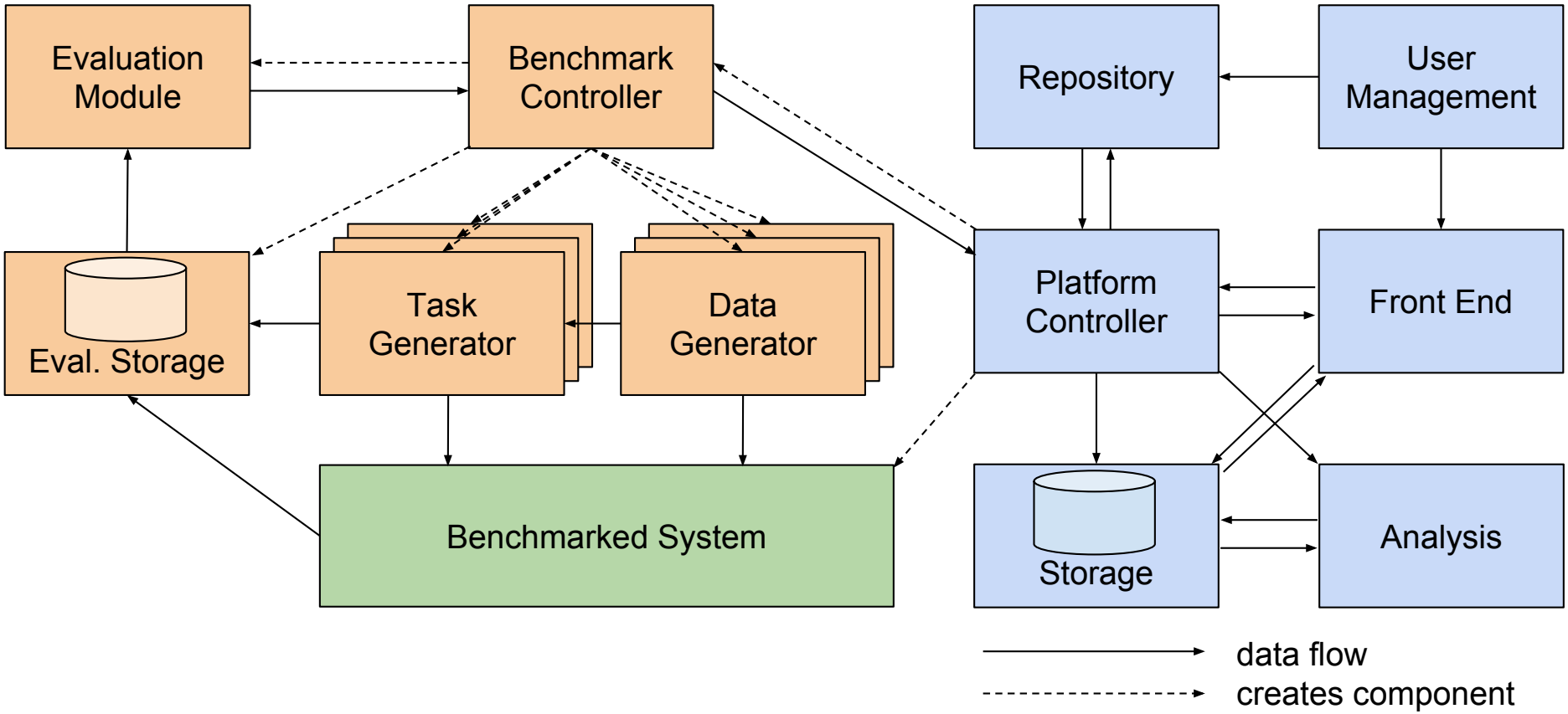
1. Gathered real requirements
   - Focussed on industrial requirements
   - Gathered relevant performance indicators
   - Gathered relevant performance thresholds
   - Gathered real datasets
2. Developed benchmarks based on real data
3. Provided universal benchmarking platform
   - Comparable results
   - Hosted as a free-to-use online instance
4. Periodic benchmarking challenges and reporting

1. Gathered real requirements
   - Focussed on industrial requirements
   - Gathered relevant performance indicators
   - Gathered relevant performance thresholds
   - Gathered real datasets
2. Developed benchmarks based on real data
3. Provided universal benchmarking platform
   - Comparable results
   - Hosted as a free-to-use online instance
4. Periodic benchmarking challenges and reporting
5. Created an association (Special Group 7 of Task Force 6)


Participants/Community
Data Collection
Industry data
Benchmark Creation
Benchmark 1
Benchmark 2
Benchmark n
Challenges
Solution 1
Solution 2
Solution k
HOBBIT Platform
Reports
Measure Collection
KPIs Tasks

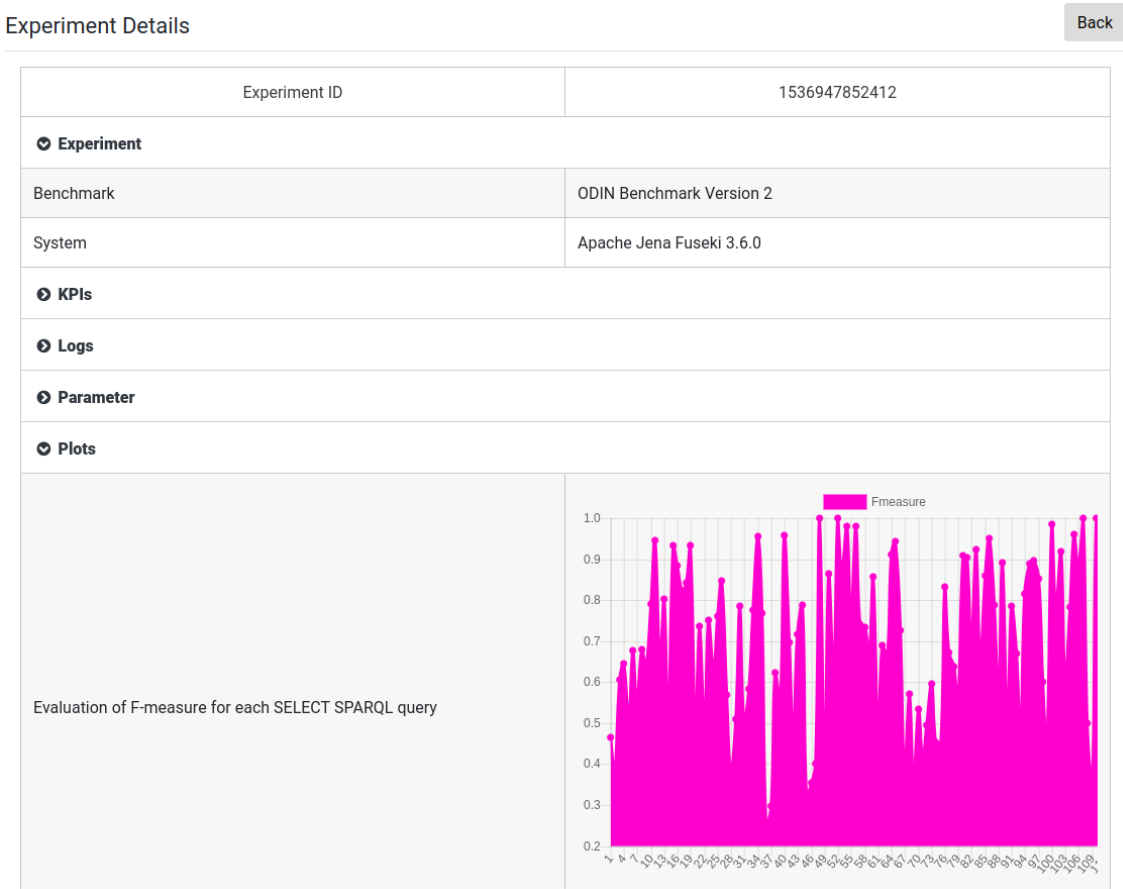
Section 3

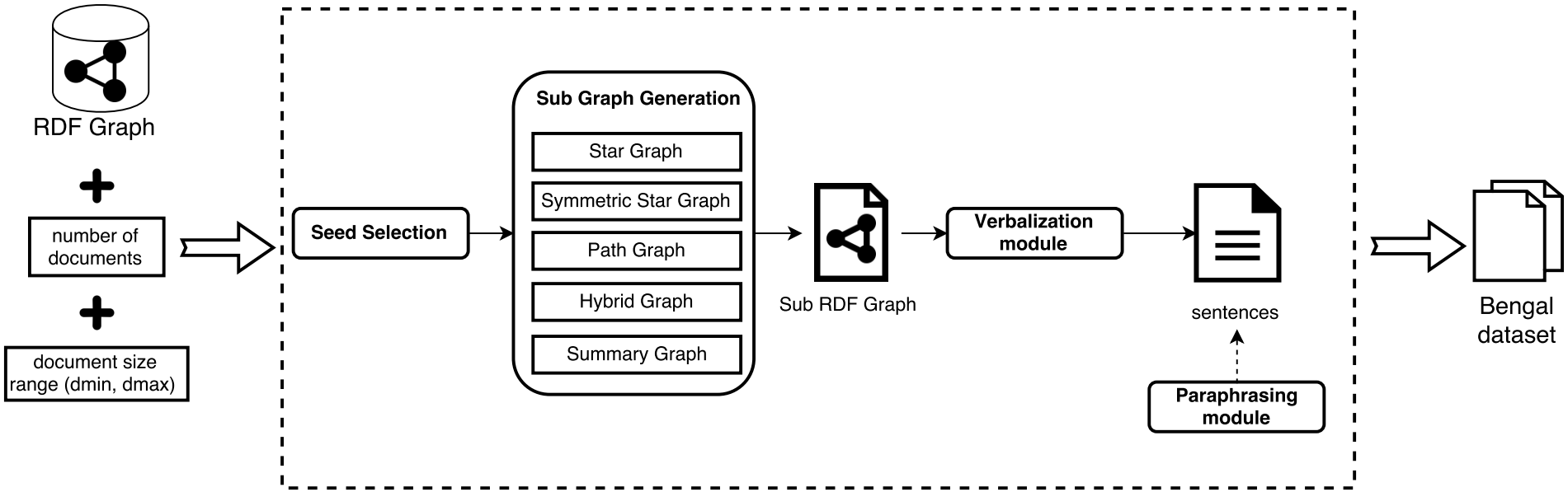
Project Highlights



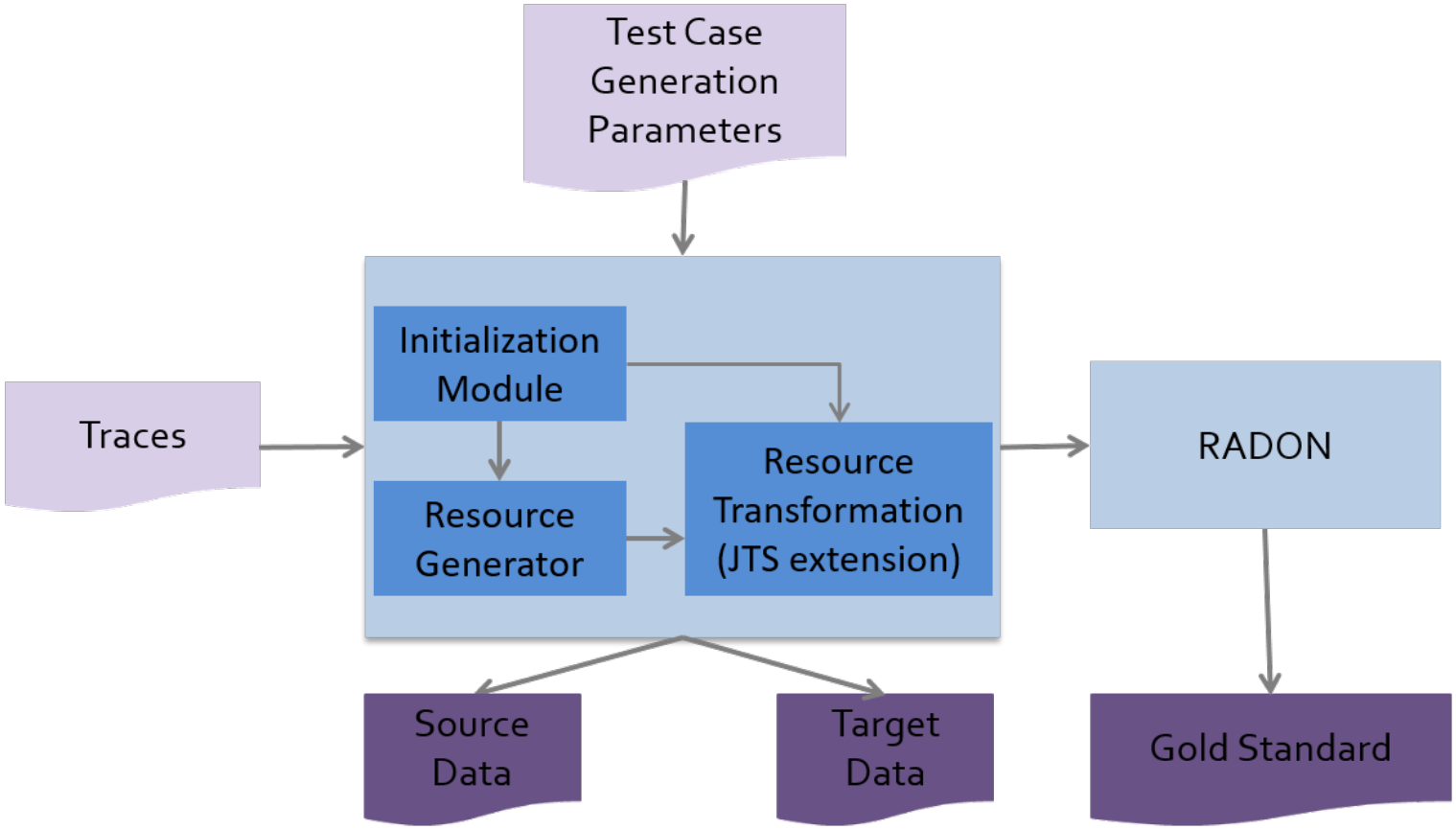
- Scalable open-source benchmarking platform
- Local, distributed and remote (AWS) deployment
- First FAIR platform for benchmarking Big Linked Data in a holistic manner

- ★ 5 mimicking algorithms
- ★ 52 benchmarks
- ★ 200+ systems
- ★ 14 challenges
  DEBS GC 2017 and 2018
- ★ 300+ users
- ★ 13K+ experiments

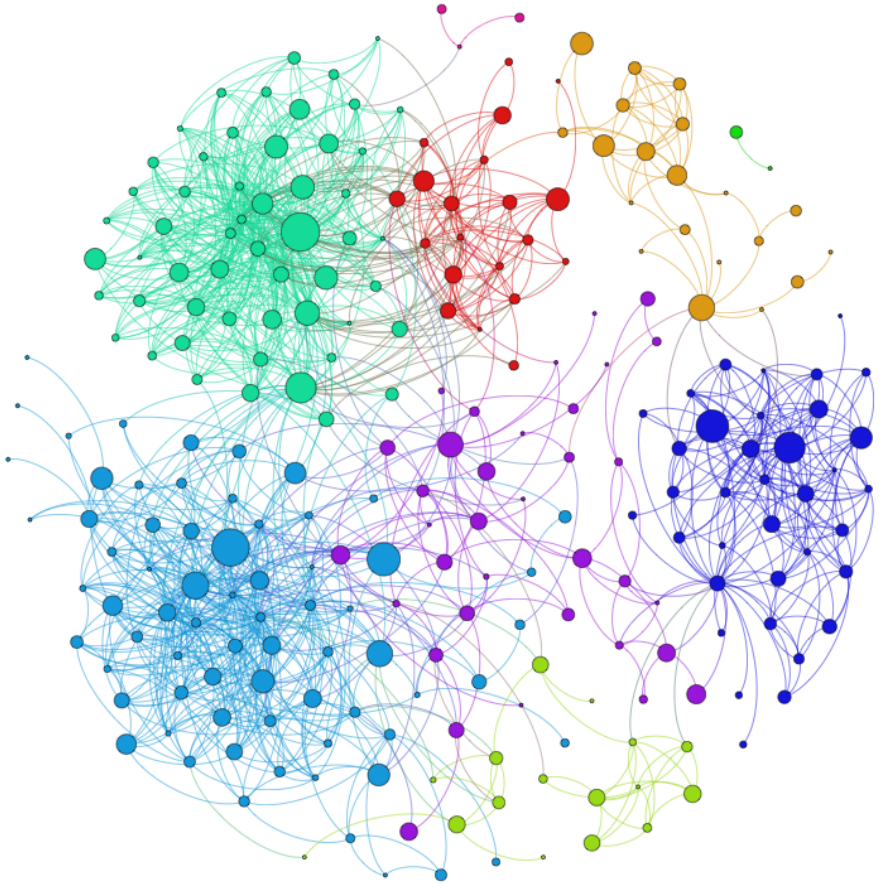




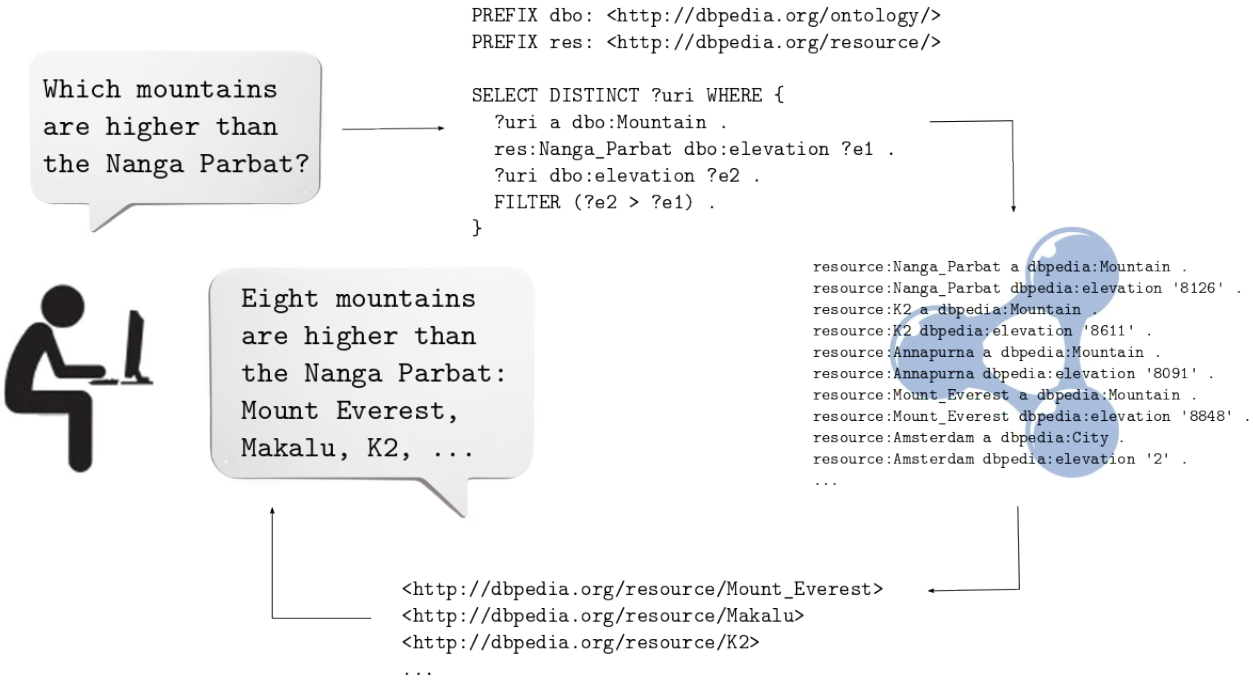
- ⋆ 21 systems benchmarked
- ⋆ 2 challenges organized
- ⋆ 2 mimicking algorithms (BENGAL and TWIG)
- ⋆ Integration with GERBIL



- ⋆ 7 systems benchmarked
- ⋆ 2 challenges (2× OAEI)
- ⋆ Collaboration with OAEI

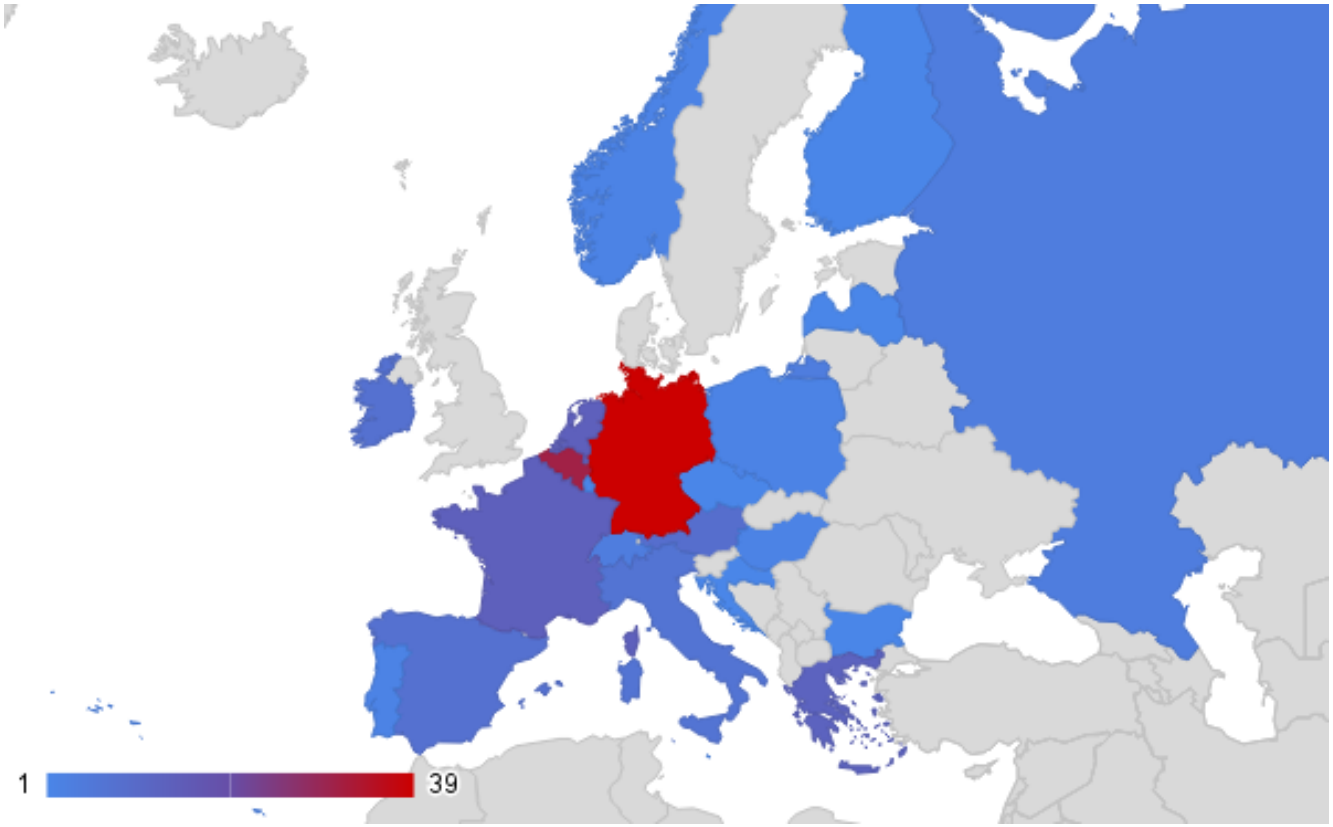


- ⋆ 9 systems benchmarked
- ⋆ 3 challenges organized
- ⋆ Improvement of world's most scalable triple store



- ★ 7 systems benchmarked
- ★ 3 challenges organized
- ★ More realistic benchmarks unveiled current systems' drawbacks

- ⋆ 14 challenges over 8 benchmarks organized (ESWC, ISWC and DEBS)
- ⋆ 77 systems benchmarked overall



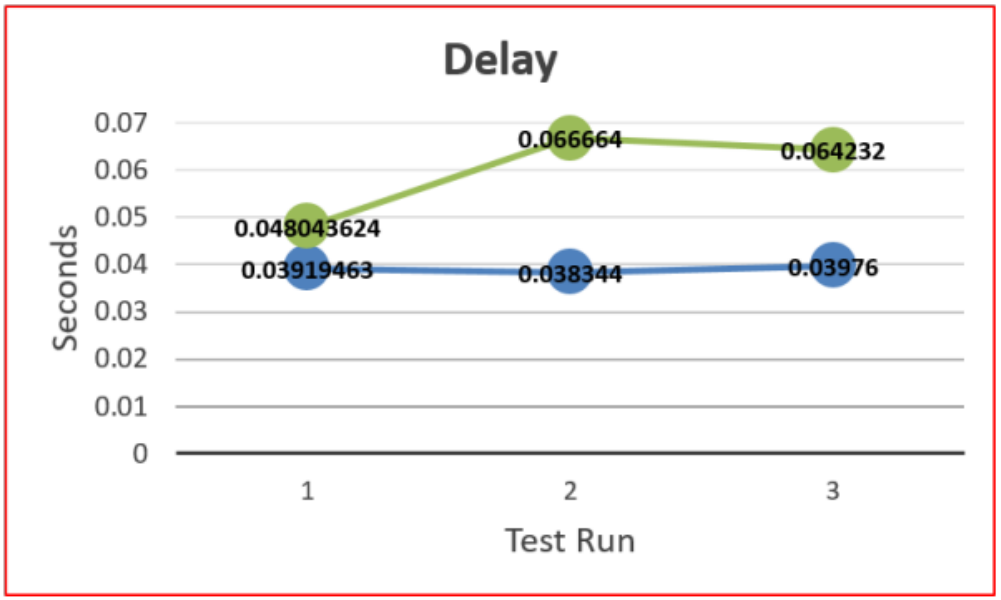
- ⋆ Association created as Special Group 7 of Task Force 6 of BDVA
- ⋆ 300 members in industry-focused contact list
- ⋆ Participation in 40+ events with 10K+ participants
- ⋆ European focus with members from outside Europe

Section 4

Benchmarking Machine Learning

- The task: find anomalies in molding machine sensor data to predict maintenance intervals (predictive maintenance).
- Mimicking algorithm based on real data
- Data was streamed as in the real world
- Participants had to use Markov Models to identify anomalies
- 14 Participants, 7 made it into the last round


Delay
2.5
2
1.5
1
0.5
0
Seconds
1
2
3
Test Run


Delay
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
Seconds
0.066664
0.064232
0.048043624
0.03919463
0.038344
0.03976
1
2
3
Test Run

- The task: predictions about ship routes based on AIS data
  - Spatio-temporal streaming data
  - Predictions for vessels' destinations and arrival times

BIG DATA OCEAN

Team	Earlyness rate	A	Working time (sec)	B	Total Q1
University of Iasi	0.685	1	99	2	1.25
University of Illinois	0.672	2	86	1	1.75
Jean Monnet University	0.668	3	149	5	3.5
Chungnam National University	0.653	4	102	3	3.75
University of Iasi (2nd)	0.647	5	157	6	5.25
Israel Institute of Technology	0.5	6	129	4	5.5
Dresden University of Technology	-	-	-	-	-
Insight Centre	-	-	-	-	-
University of Carthage	-	-	-	-	-

Team	Mean Absolute Error (min.)	A	Working time (sec)	B	Total Q2
University of Iasi	959.839	1	100	2	1.25
Jean Monnet University	1099	2	145	4	2.5
Chungnam National University	1251.15	3	100	2	2.75
Israel Institute of Technology	1493.18	4	133	3	3.75
University of Illinois	5425.53	5	86	1	4.75
University of Iasi (2nd)	1705.35	6	164	5	5
Dresden University of Technology	-	-	-	-	-
Insight Centre	-	-	-	-	-
University of Carthage	-	-	-	-	-

Team	**Q1**	**Q2**	**Total Score**
University of Iasi	1.25	1.25	2.5
Jean Monnet University	3.5	2.5	6
Chungnam National University	3.75	2.75	6.5
University of Illinois	1.75	4.75	6.5
Israel Institute of Technology	5.5	3.75	9.25
University of Iasi (2nd)	5.25	5	10.25
Dresden University of Technology	-	-	-
Insight Centre	-	-	-
University of Carthage	-	-	-

Section 5

Future Directions

- KnowGraphs (Innovative Training Networks (ITN))
  - 4 years, starting in October 2019
  - 15 Early-Stage Researchers (ESRs)
  - HOBBIT will be used as central benchmarking platform
  - Further datasets will be integrated (e.g., UICML datasets)
- RAKI (BMWi project)
  - 3 years, starting in September 2019
  - HOBBIT will be used for evaluation
- More projects pending

Future Directions

- KnowGraphs (Innovative Training Networks (ITN))
  - 4 years, starting in October 2019
  - 15 Early-Stage Researchers (ESRs)
  - HOBBIT will be used as central benchmarking platform
  - Further datasets will be integrated (e.g., UICML datasets)
- RAKI (BMWi project)
  - 3 years, starting in September 2019
  - HOBBIT will be used for evaluation
- More projects pending

→ Further development of the HOBBIT platform

- HOBBIT is open for the community! Benchmarks, systems, datasets can be added
- Not limited to linked data

Thank you

HOBBIT offers

- Scalable benchmarking
- Based on real world data in an
- Extendable,
- Open source platform
- Following the FAIR data principles

`http://project-hobbit.eu/`
`https://dice-research.org/about/`