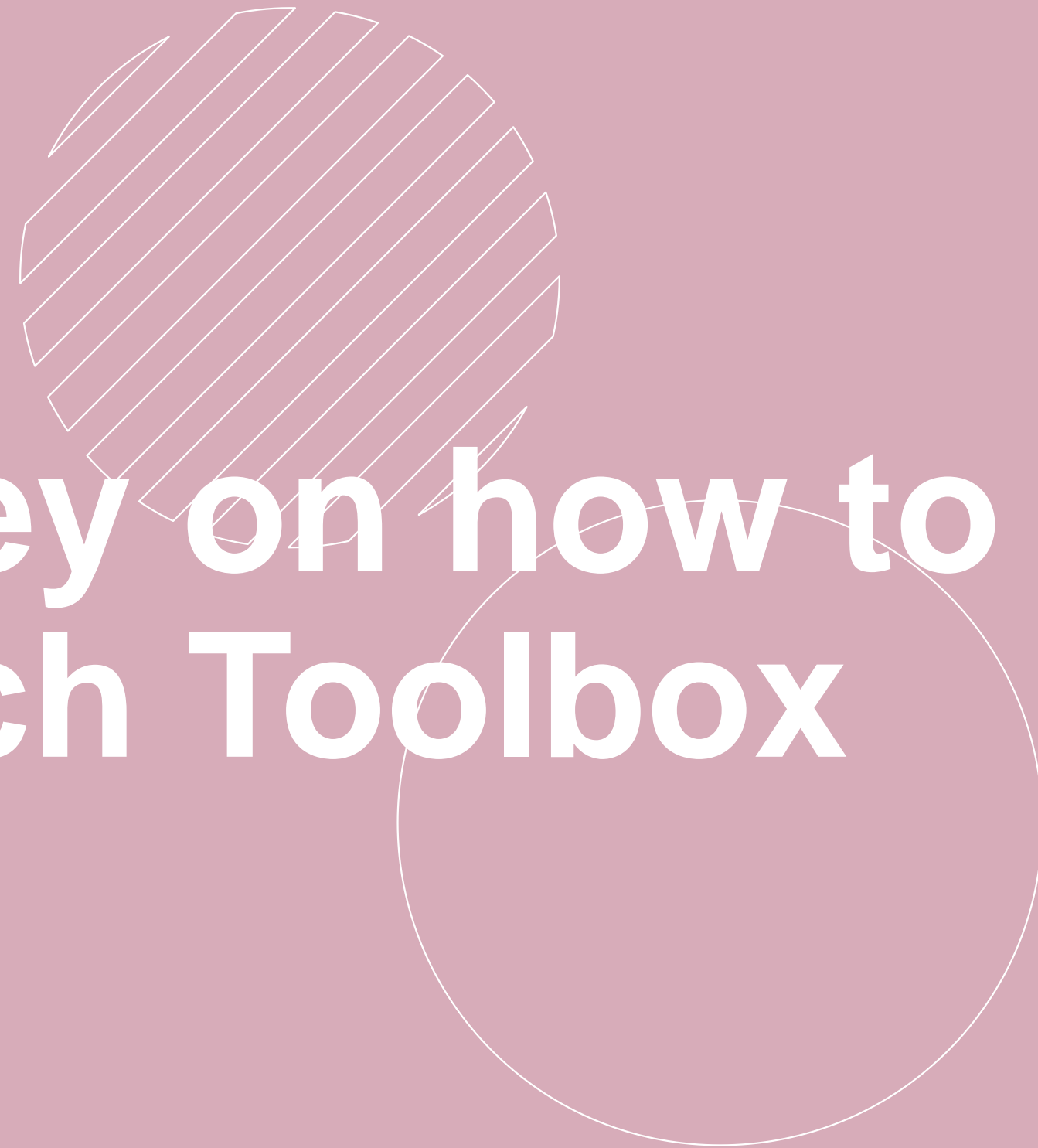


# Session 4.

## A practical journey on how to use the DataBench Toolbox



# Panelists



**Tomás Pariente**

Associate Head of AI, Data & Robotics Unit, Atos Research and Innovation



**Daniel Sáez Domingo**

Strategic Intelligence & Technology Transfer Director, ITI. Coordinator of the Data Cycle Hub (DCH) and EuHubs4Data



**Marko Grobelnik**

Artificial Intelligence Lab, Jozef Stefan Institute

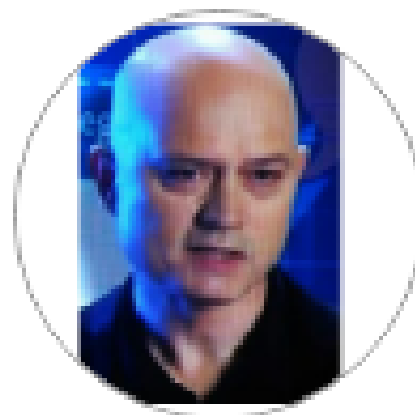


EUROPEAN BIG DATA VALUE FORUM 2020



DataBench

# Demo of the DataBench Toolbox



Tomás Pariente Lobo

Associate Head of AI, Data & Robotics Unit at Atos

Evaluation schemes for Big data and AI Performance of high Business impact (DataBench project sponsored session), Parallel session on Data Driven Innovation workshop

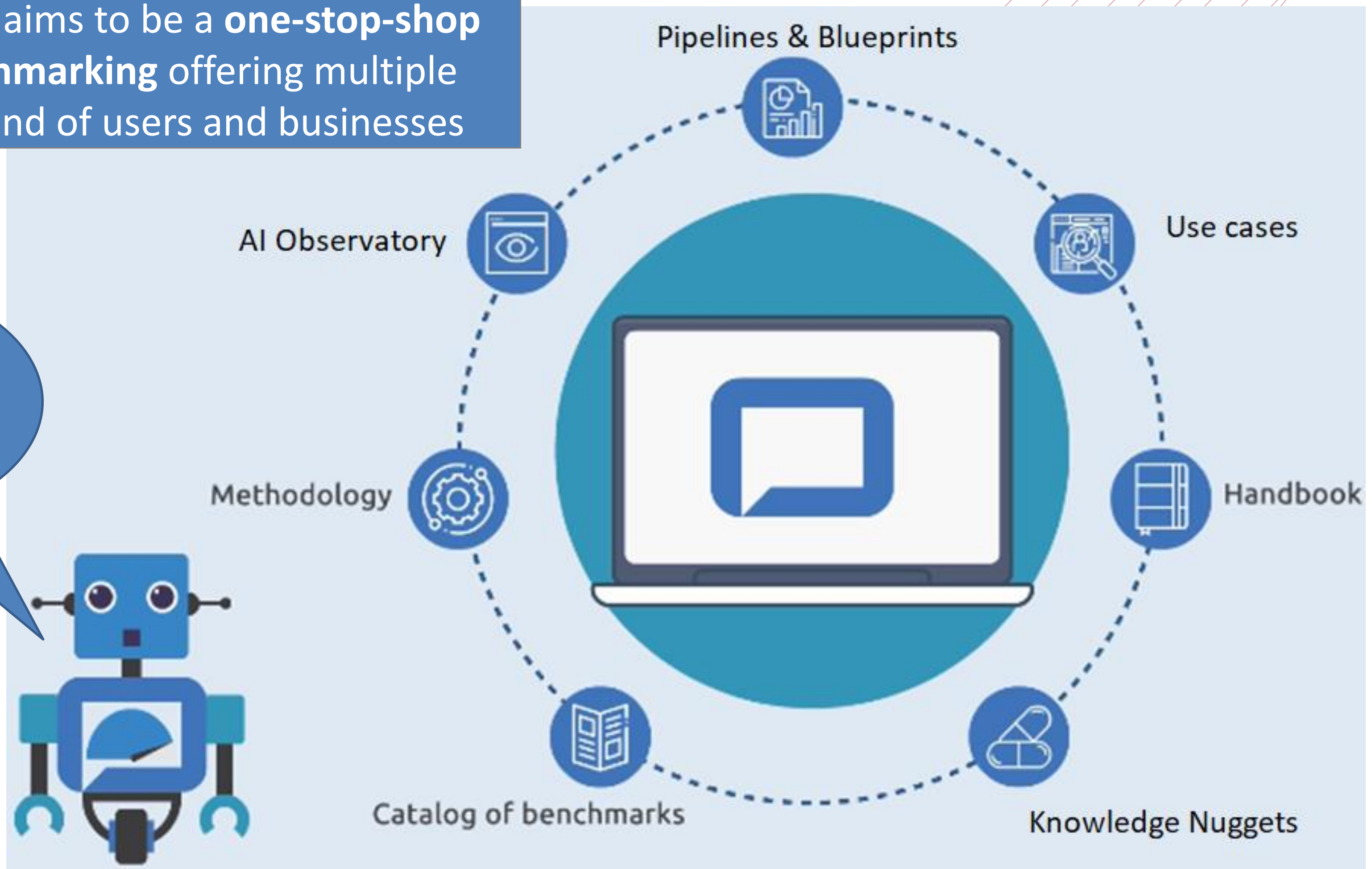
AI, Data & Robotics Unit at Atos Research & Innovation, Madrid Area, Spain. Tomás Pariente Lobo has more than 30 years of experience in IT. His technical expertise is mainly in Artificial Intelligence, Big Data, Linked Data and knowledge management. Since June 2006 Tomas works as a project manager and technical coordinator for EU-funded projects leading a group of researchers dealing with all aspects related to the data value chain, with special focus on data architectures, data analysis and technologies such as Natural Language Processing and semantics..



# DataBench Toolbox

The DataBench Toolbox aims to be a **one-stop-shop for Big Data / AI Benchmarking** offering multiple benefits for different kind of users and businesses

Connecting the  
Toolbox with all  
DataBench  
results



# Toolbox

## Showcasing knowledge about benchmarking

Who are our stakeholders?

Benchmark providers / communities

BDV PPP Projects / Researchers

Technical users

Business Users / CTOs...

What can we do for them?

Visibility, integration

Benchmarking  
Alignment with BDV RM / Blueprints

Catalogues / Search  
Deployment / Execution  
Pipelines / Blueprints

Knowledge,  
Use cases / Industries,  
Correlation Tech / KPIs

How?

Operationalize work  
done in DataBench

Catalogue of benchmarks vs  
BDV Reference model and  
data pipelines

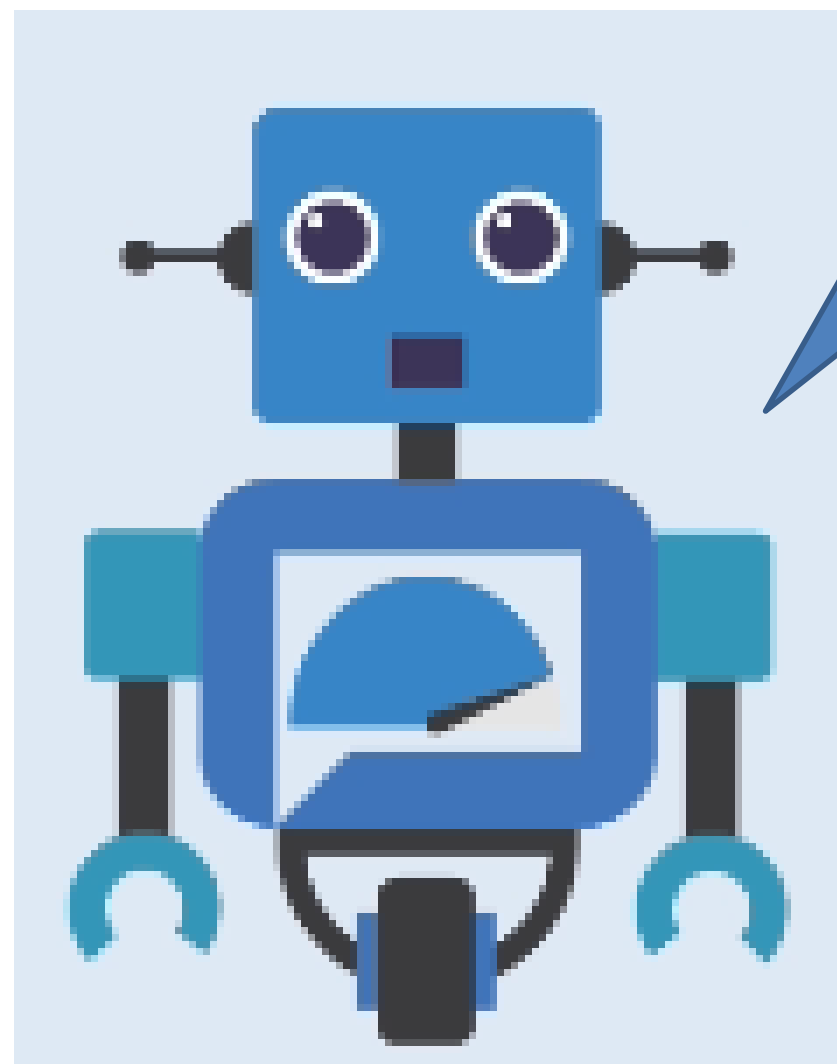
Reports, methods,  
indicators, self-assessment,  
surveys

Relation to use cases &  
industries, examples,  
Handbook

AI Observatory, blueprints,  
admin...



DEMO



DataBench



# AlBench Observatory



Marko Grobelnik

senior researcher at Jozef Stefan Institute

Evaluation schemes for Big data and AI Performance of high Business impact  
(DataBench project sponsored session)

---

Marko Grobelnik is a researcher in the field of Artificial Intelligence (AI). Focused areas of expertise are Machine Learning, Data/Text/Web Mining, Network Analysis, Semantic Technologies, Deep Text Understanding, and Data Visualization. Marko co-leads Artificial Intelligence Lab at Jozef Stefan Institute, cofounded UNESCO International Research Center on AI (IRCAI), and is the CEO of [Quintelligence.com](https://www.quintelligence.com) specialized in solving complex AI tasks for the commercial world. He collaborates with major European academic institutions and major industries such as Bloomberg, British Telecom, European Commission, Microsoft Research, New York Times. Marko is co-author of several books, co-founder of several start-ups and is/was involved into over 50 EU funded research projects in various fields of Artificial Intelligence. Marko represents Slovenia in OECD AI Committee (ONE AI), in Council of Europe Committee on AI (CAHAI), and Global Partnership on AI (GPAI). In 2016 Marko became Digital Champion of Slovenia at European Commission.

---



EUROPEAN BIG DATA VALUE FORUM 2020

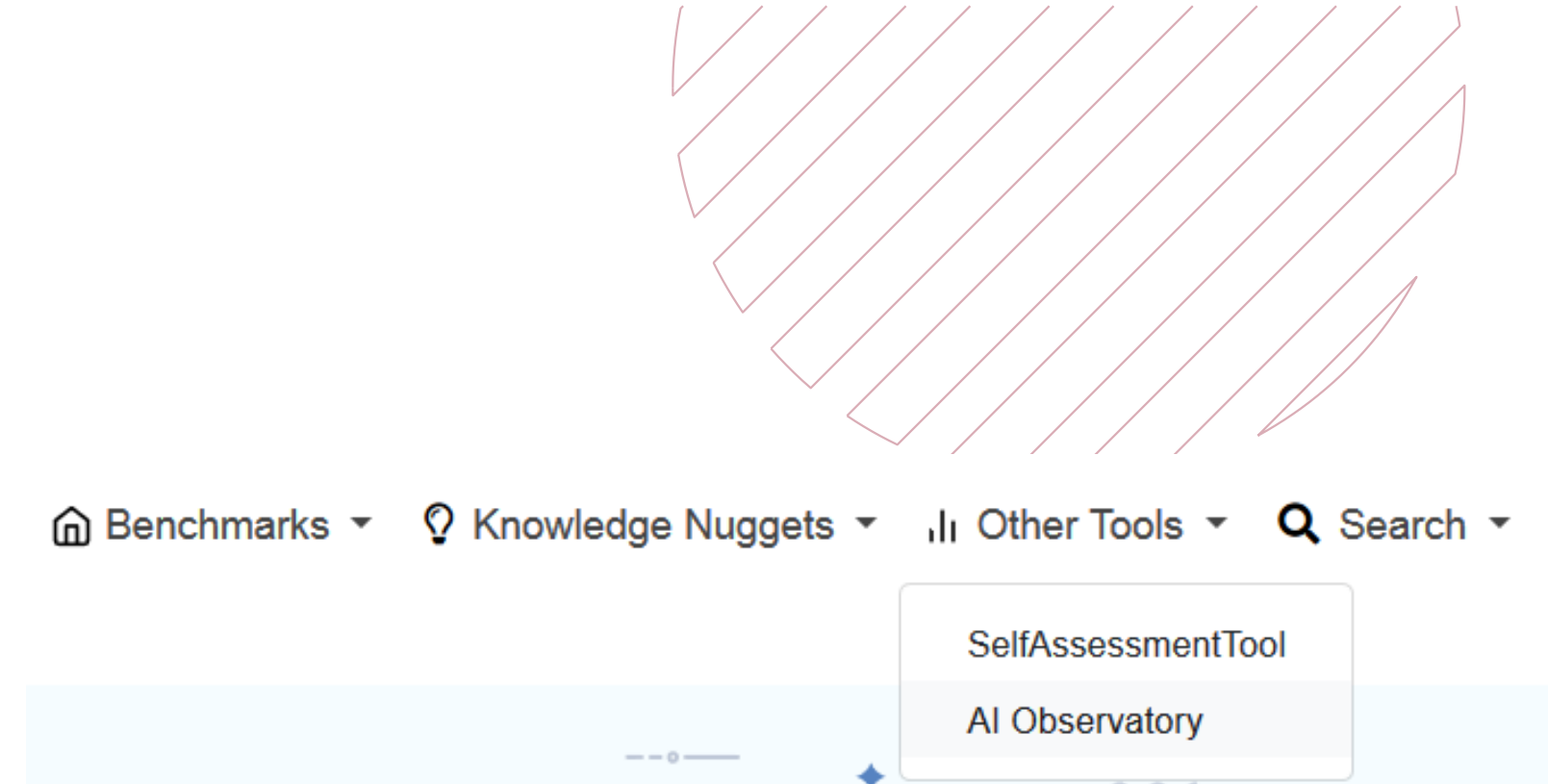
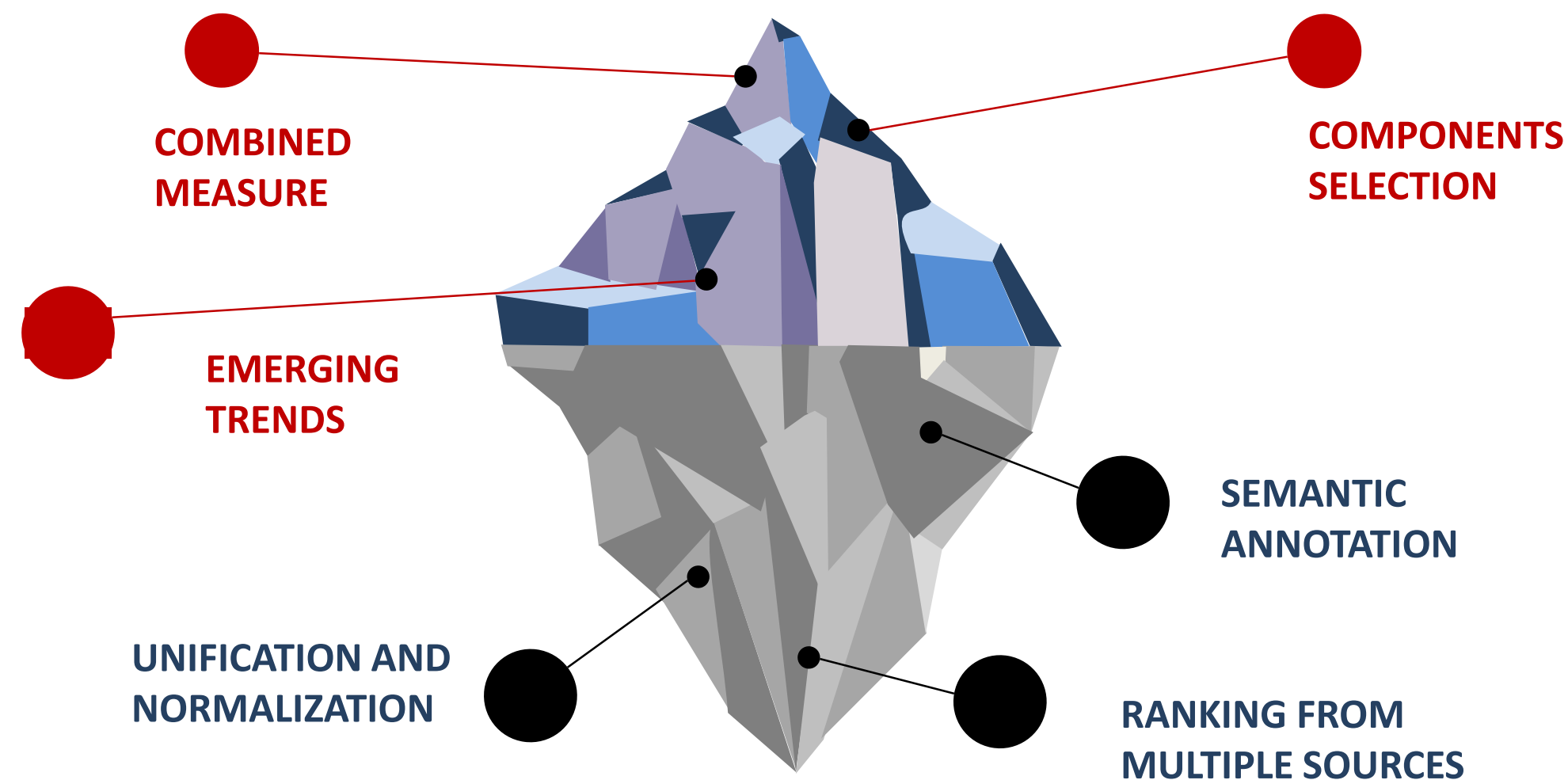


DataBench

# AlBench Observatory

**AlBench Observatory** is a tool to **observe the popularity, importance and visibility of topic terms** related to the Artificial Intelligence and Big Data along several dimensions **on a monthly basis**.

**DataBench Popularity Index** is a measure to **rank** topic terms in the area of **Artificial Intelligence and Big Data** (with specific accent on tools and technologies related to **Benchmarking**).




[aibench.ijs.si](http://aibench.ijs.si)




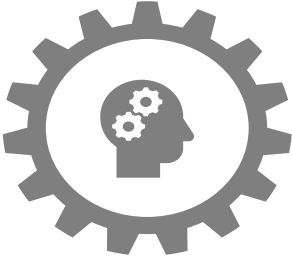
**DataBench**




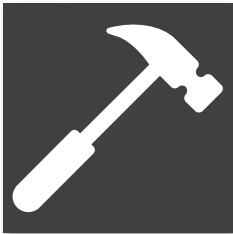
# Data Sources to assess Popularity Index


- 


Microsoft Academic Graph  
[Academia & Patents]
- 

Media News from Event Registry  
[News and Social media]
- 

GitHub projects  
[Development]
- 

Horizon2020 projects  
[Research projects]
- 

Job ads from Adzuna  
[Job market & skills]
- 

Search volume from Google Trends  
[Global popularity]
- 

Investments from Prequin  
[Investment trends]

month: All						
Search:						
Topic	Papers	Projects	News	Github	Jobs	Total
infrastructure	4.45	2.47	10	2.67	10	5.92
customer	6.18	10	1.71	1.06	1	3.99
stat, tool	3.96	2.03	1.13	9.69	1.45	3.65
Artificial Intelligence	3.7	1.06	1.6	10	1.12	3.5
label	10	1	3.84	1.59	1.01	3.49
misc	5.42	2.22	3.52	2.3	1.14	2.92
relational dbms	1.76	1.07	1.21	1.29	8.25	2.72
Machine Learning	1.49	1.06	1.24	8.56	1.12	2.69

month: All categories: computer_vision						
Search:						
Topic	Papers	Projects	News	Github	Jobs	Total
Amazon Rekognition	2.42	10	4.43	7.26	5.5	5.92
Scape	10	1	5.71	1.78	5.5	4.8
Affectiva	1.82	10	2.66	1	5.5	4.2
SenseTime	1.09	1	10	2.96	5.5	4.11
Google Cloud Vision	1	1	1.82	10	5.5	3.86
Kairos	5	1	4.71	1.78	5.5	3.6
Ever AI	4.1	1	2.61	3.74	5.5	3.39

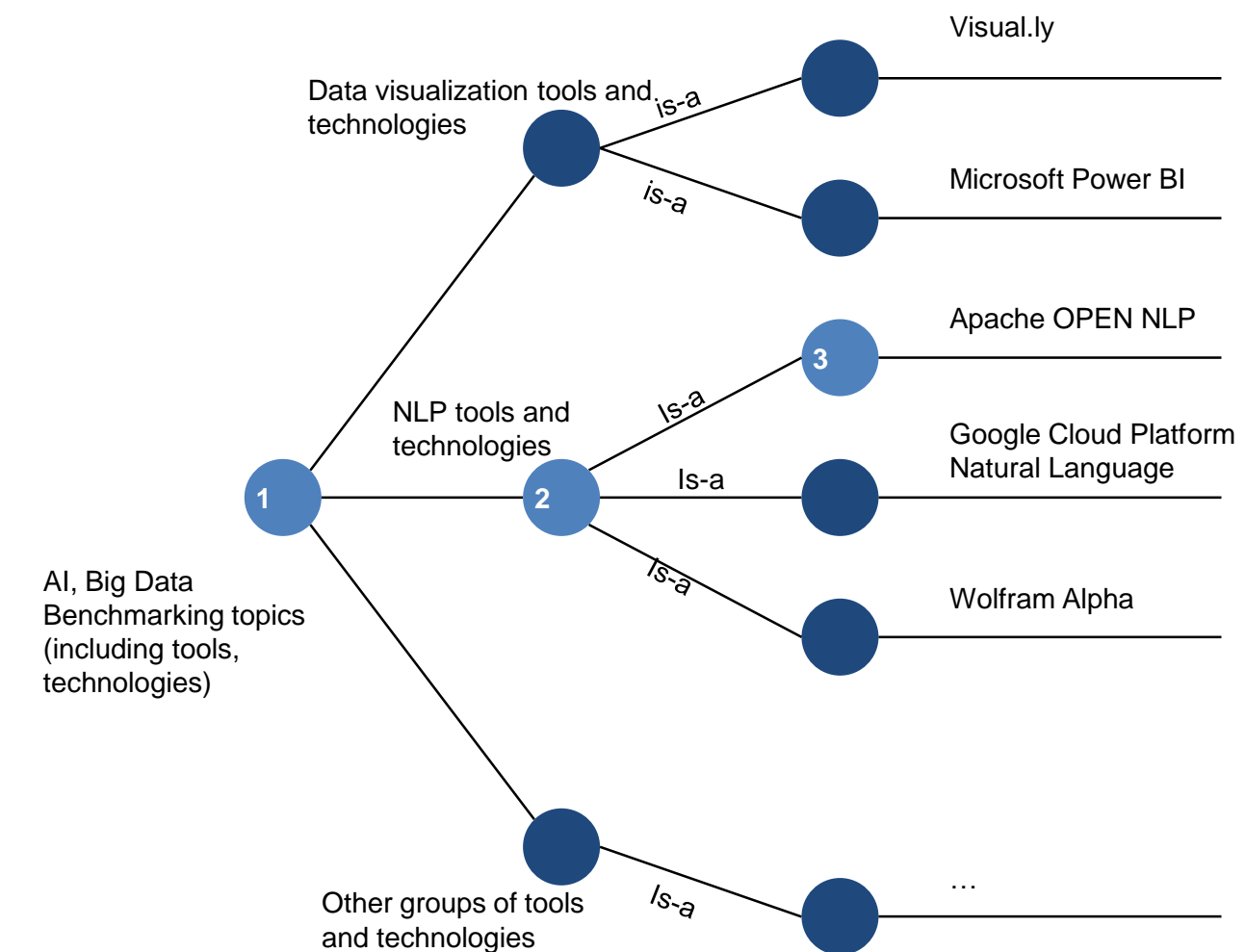
Global and & Dynamic data sources



# Ontology of AI & Big Data

**Microsoft Academic Graph (MAG) taxonomy** has been expanded with **DataBench terms** – over 1,700 tools and technologies related to Benchmarking, Big Data, Artificial Intelligence.

New concepts have been **aligned with MAG topic, MAG keyword, Wikipedia** (for analysis in wikification) and **Event Registry concepts**.



## Example

**Term:** Apache OPEN NLP

**Parent:** [databench.ijs.si/nlp]

**URI:** databench.ijs.si/apache\_open\_nlp

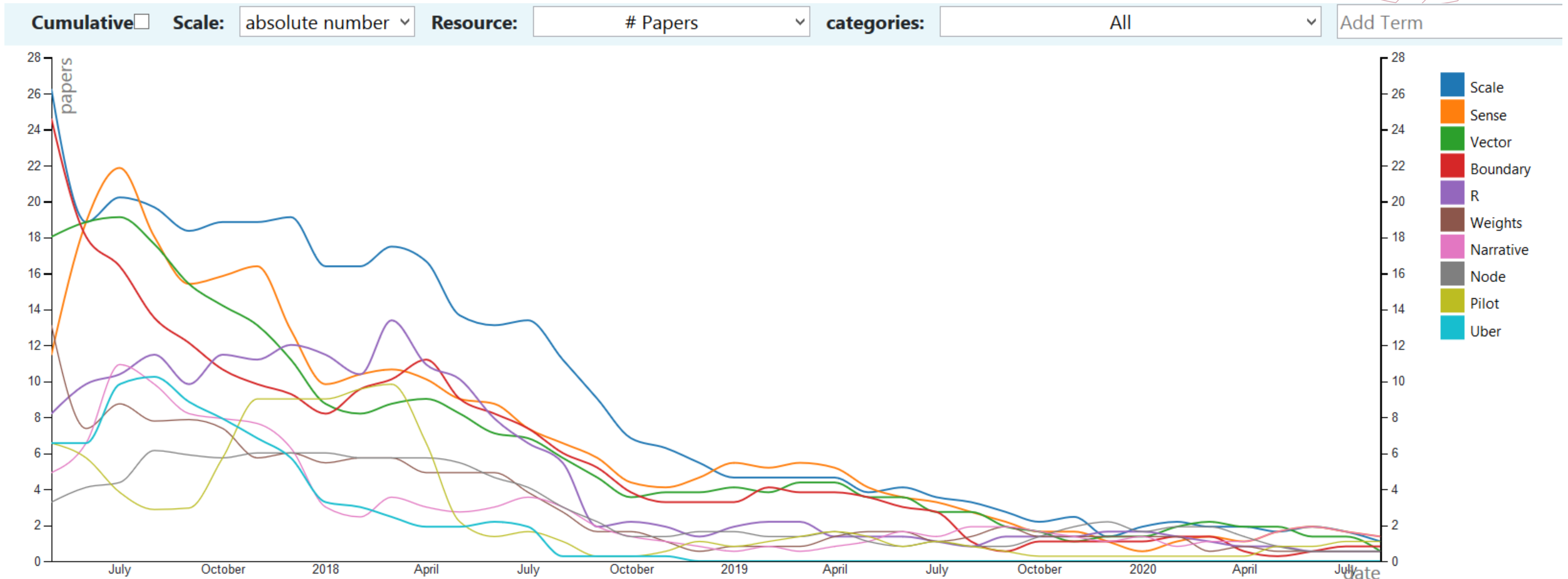
**Wiki concept:** [https://en.wikipedia.org/wiki/Apache\\_OpenNLP](https://en.wikipedia.org/wiki/Apache_OpenNLP)

**ER concept:** "Apache OPEN NLP"



**DataBench**

# Time Series – popularity over time

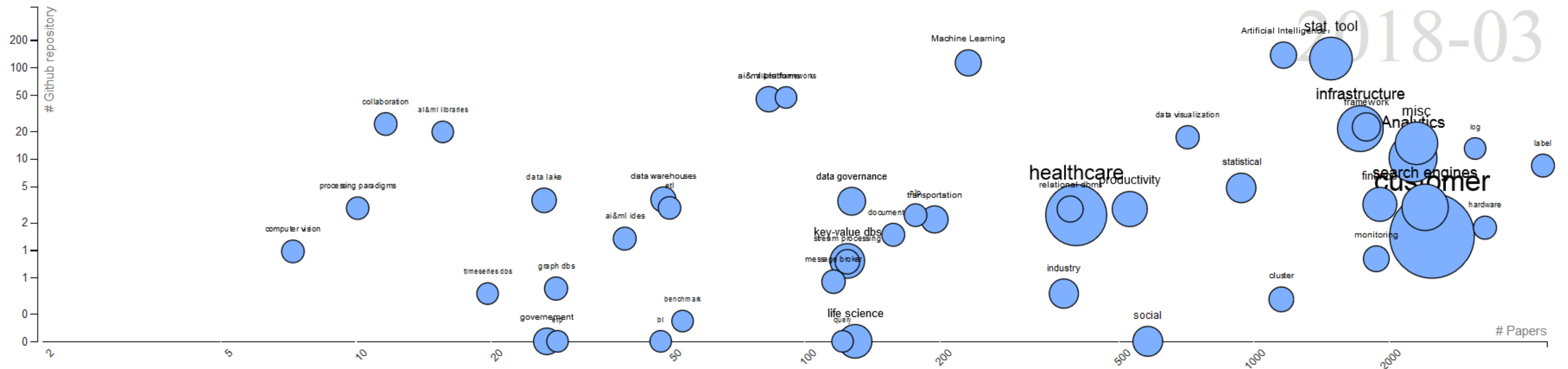
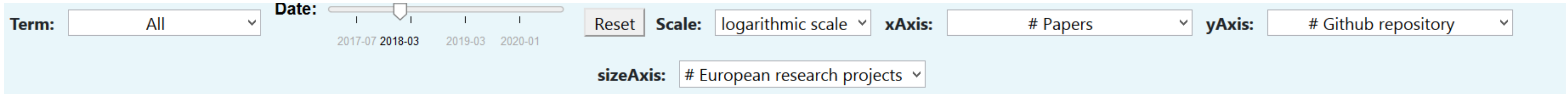
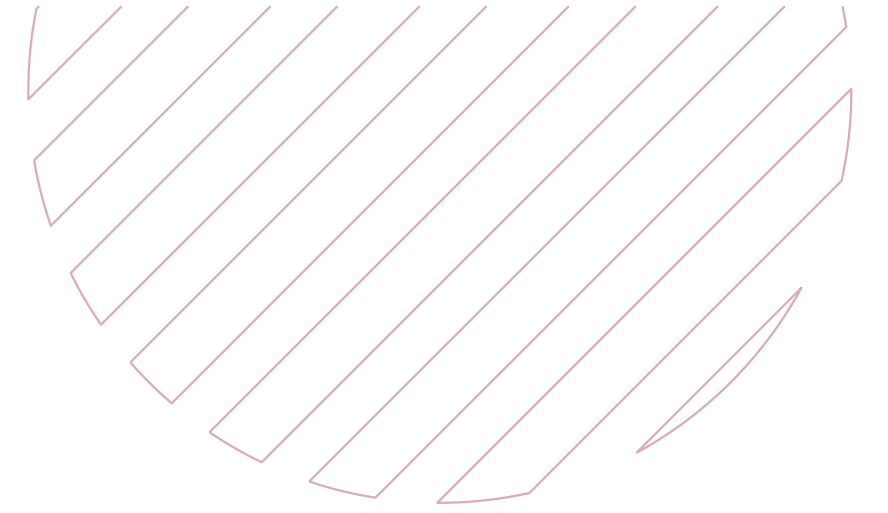


Time series for Tools and Technologies



DataBench

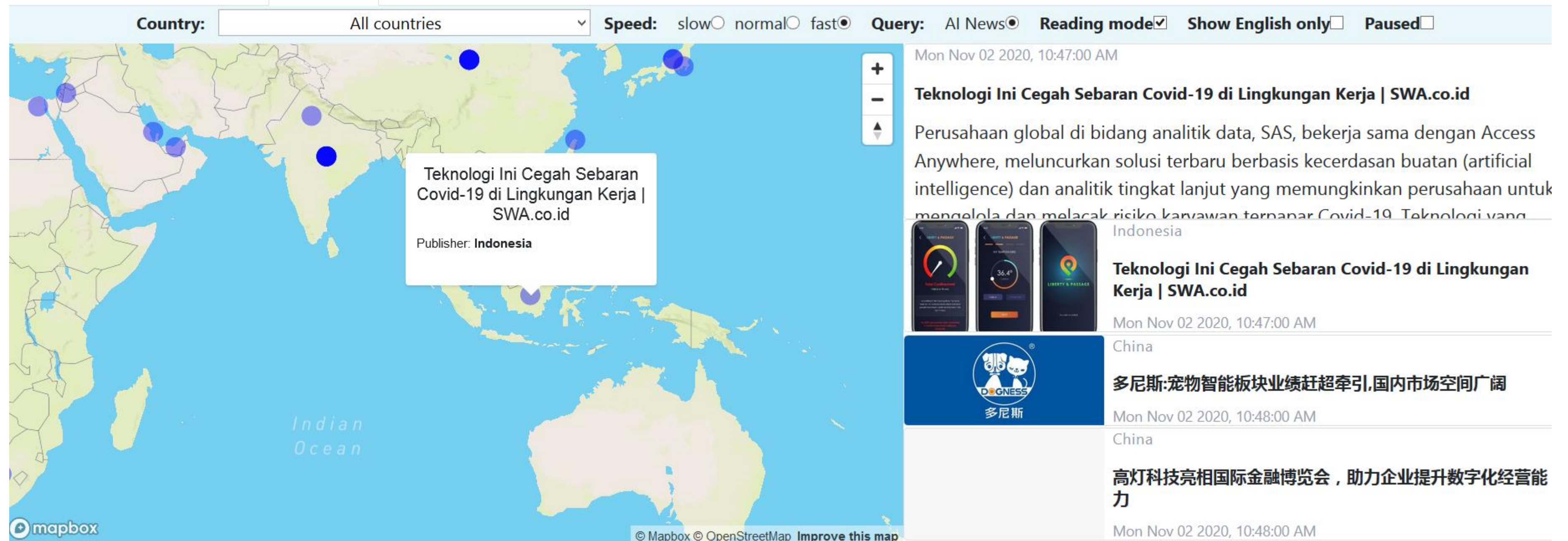
# Moving Indicators



DataBench



# Live News – media coverage



# Fostering adoption of DataBench results. Needs from the point of view of Digital Innovation Hubs



**Daniel Sáez Domingo**

Strategic Intelligence &  
Technology Transfer Director, ITI.  
Coordinator of the Data Cycle  
Hub (DCH) and EuHubs4Data



EUROPEAN BIG DATA VALUE FORUM 2020



DataBench



# Data Driven Innovation Hubs are key instruments to foster DDI



BDVA SRIA -  
[http://bdva.eu/sites/default/files/BDVA\\_SRIA\\_v4\\_Ed1.1.pdf](http://bdva.eu/sites/default/files/BDVA_SRIA_v4_Ed1.1.pdf)

*Environments for data driven innovation and experimentation, where research on BDV technologies and novel BDV applications can be quickly tested, piloted and exploited*

*Technical and non-technical activities, bringing data, technology, application developments together, and fostering skills, competences and best practices*

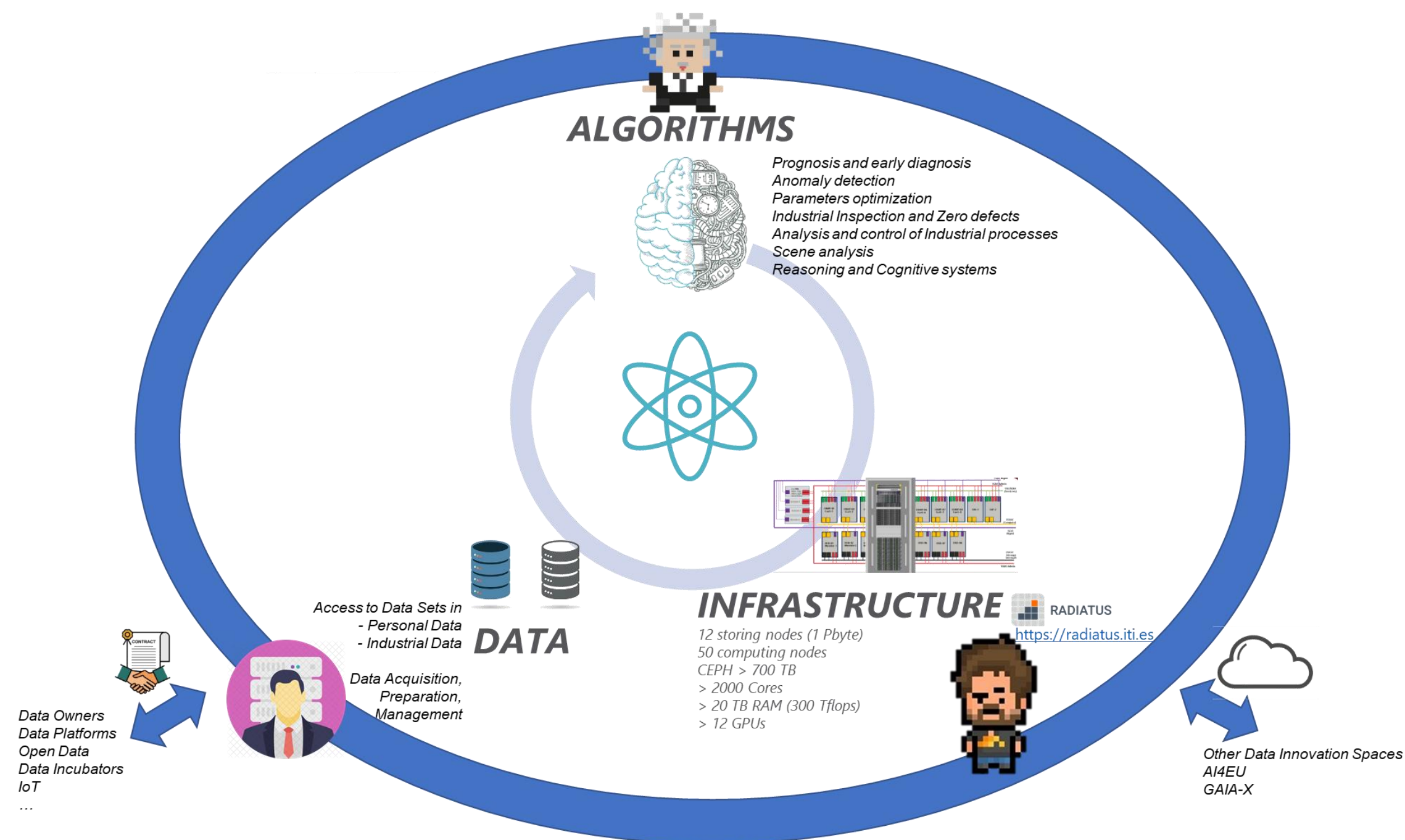
*Rely on existing national and European initiatives, federate, complement and leverage activities of similar national incubators / environments and existing PPPs*

*Needs of large and small companies, to ensure access to economic opportunities offered by BDV, testing the viability of actual business deployments*

***Services to accelerate data driven innovation***

# Local Offer ...

- **One-Stop-Shop in the region**
- **Support Digitalization of SMEs and Public Admin.**
- **Neutral infrastructures and information about Data driven technologies and Data Sharing options**
- **Personalized assessment to companies in order to take decisions about technology investments or business models in the digital economy.**
- **Deploy pilot projects and technology tests in lab environments as a previous phase to the investment in a technology or the launch of a product to market.**












# ... but Connected to Europe

- **The Needs**

- Use Cases
- Metrics
- Projects
- Networks
- Maturity models
- Demonstrators
- Orchestrators
- Data Sharing Schemes
- Safe-mode tests

- **The Services**

- Technology Selection 
- Data maturity & Digital assessment 
- Access to Data & infrastructures
- Strategic approach around Data 
- Support to experiments
- R&D and PoC
- Prototyping and concept validation
- Ecosystem Building 
- New Business Models 
- Funding
- Data & AI Technology Watch 
- Ethics and legal on Data
- Skills/Training 
- Brokering/Consortium Building/Innovation Scouting
- Data governance services



# Contacts



[www.databench.eu](http://www.databench.eu)



[info@databench.eu](mailto:info@databench.eu)



[@DataBench\\_eu](https://twitter.com/DataBench_eu)



[DataBench](https://www.facebook.com/DataBench)



[DataBench Project](#)



[DataBench](#)



[DataBench Project](#)



# DataBench

EUROPEAN  
**BIGDATA  
VALUE** FORUM

**BERLIN + VIRTUAL  
3-5 NOVEMBER 2020**



This project has received funding from the European Horizon 2020 Programme for research, technological development and demonstration under grant agreement n° 780966