



EUROPEAN COMMISSION
EUROSTAT

First hand demo of winning project
1st EU Big Data Hackathon

EU Jobs and Skills Explorer

Belgrade, Serbia, 12th July 2017.



Agenda

1. Team
2. About IN2data (what we do)
3. EU Big Data Hackathon Executive Summary
4. EU Big Data Hackathon Timeline
 - Team Selection
 - Pre-Hackathon Activities
 - Competition
 - Policy Question
 - Platform Selection
 - Solution/Output
 - Panel Members
5. Prototype Live Demo
6. Prototype Credentials
7. Q&A

TEAM

About Us



Igor Kaluđer, PhDc
igor.kaludjer@in2data.eu



Leo Mršić, PhD
leo.mrsic@in2data.eu



Robert Kopal, PhD
robert.kopal@in2data.eu



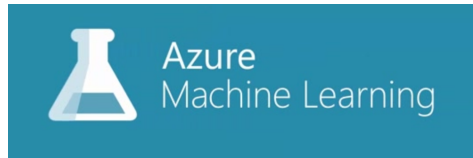
Milica Mičić, PhD
milica.micic@in2.rs

Average 18+ years of work experience per expert, active in business/EDU/GOV community, PhD/PhDc level in advanced analytics, Professors/lecturers, 50+ scientific papers, 15+ books, 100+ conferences/lectures, data economy professionals, data monetization experts...

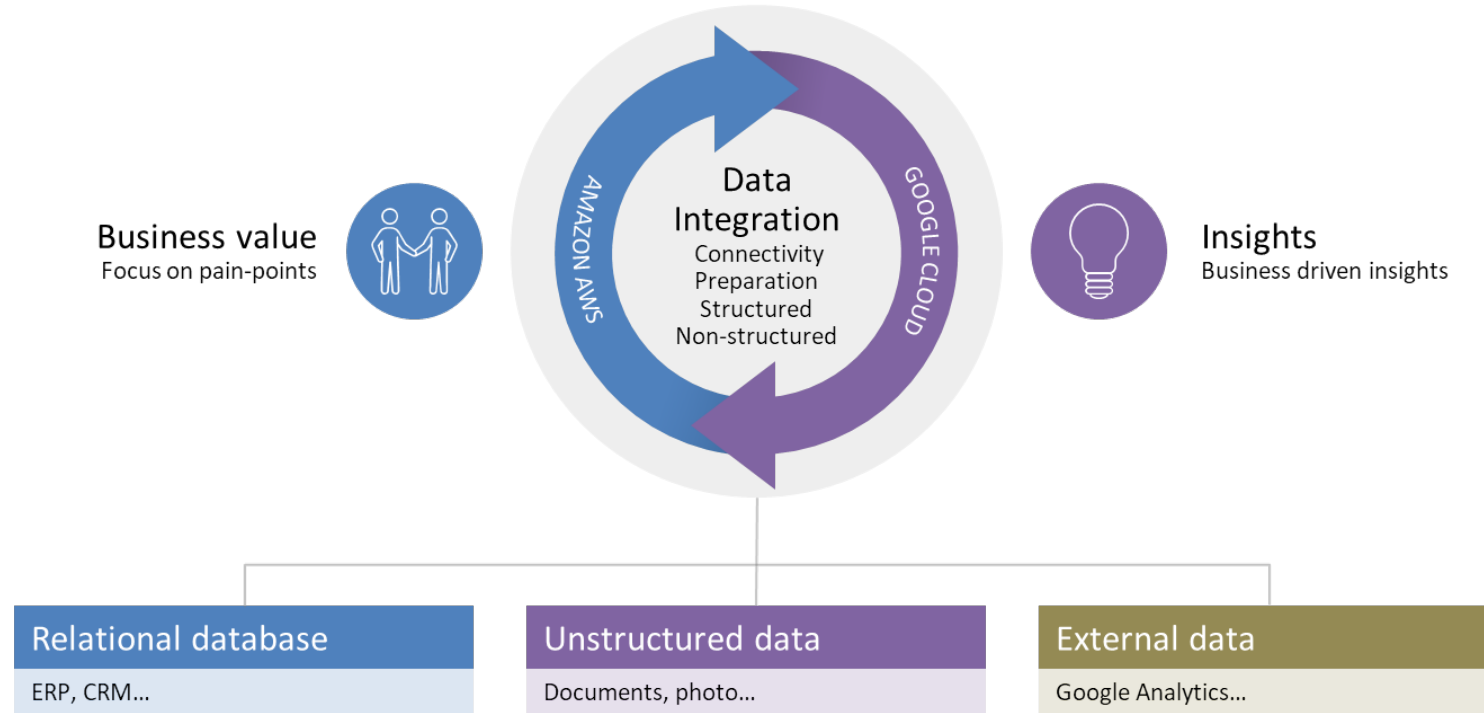
IN2DATA – WHAT WE DO

About IN2data Company

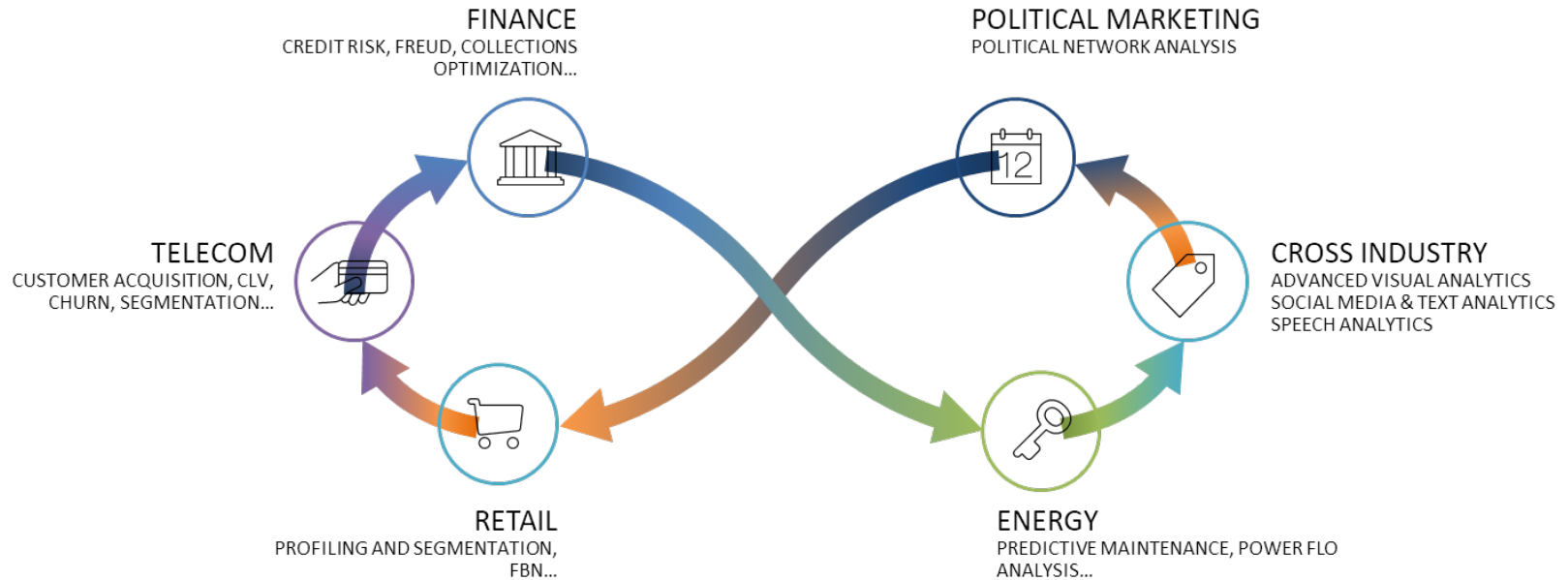
Unique team, 8 team members (data scientists),
Data economy oriented, Vendor agnostic...



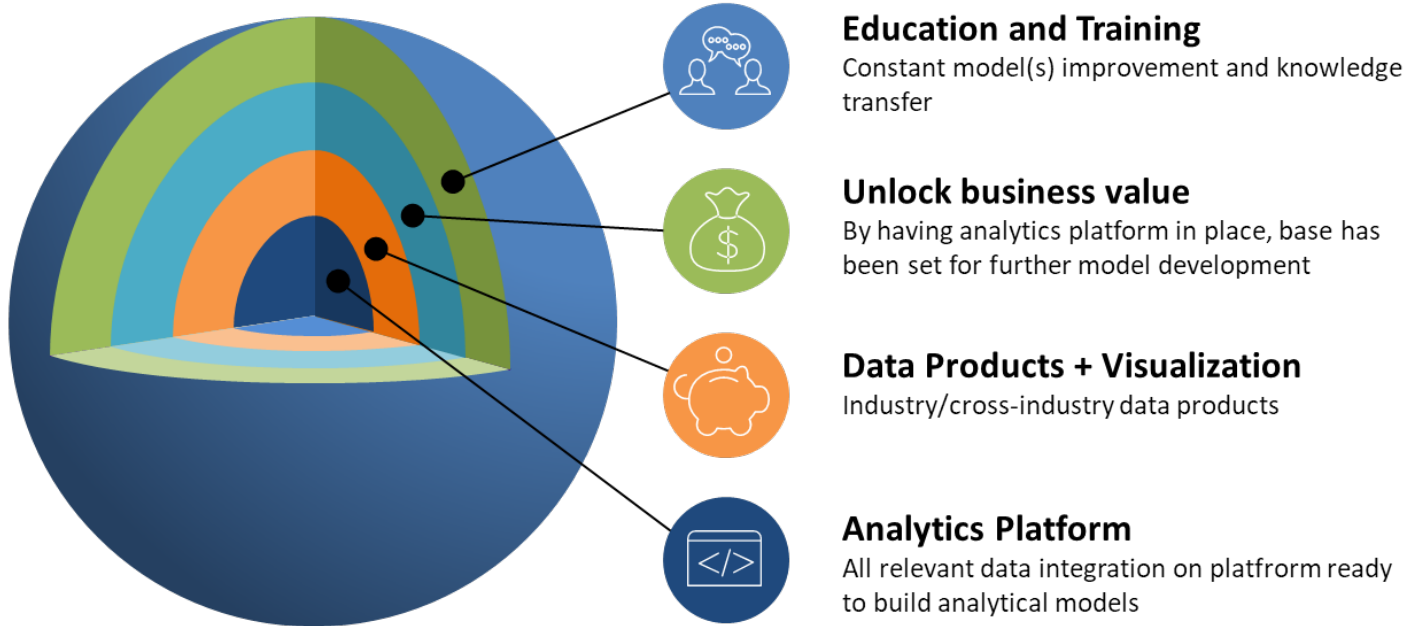
Analytics Platform



Data Products / 60+

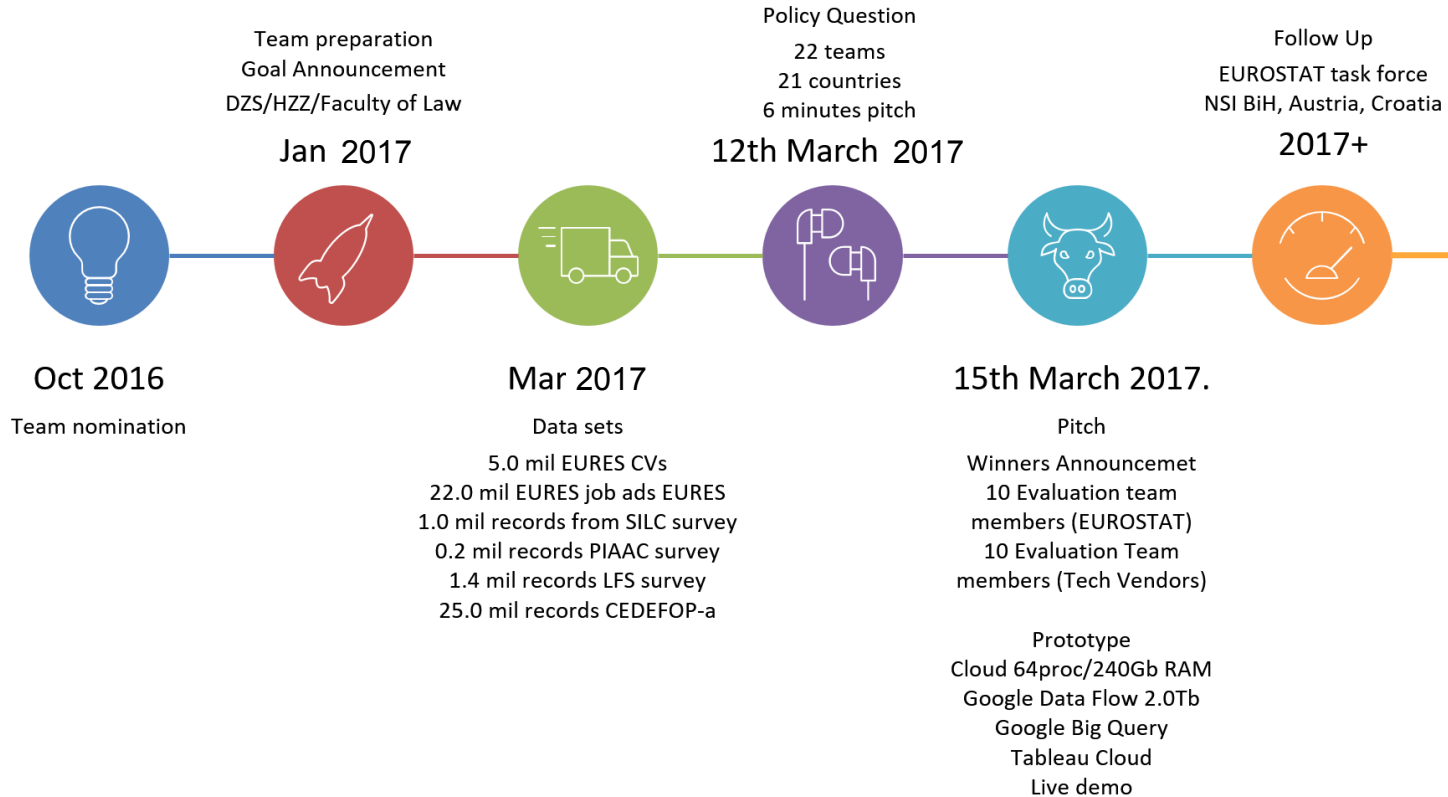


Business Model



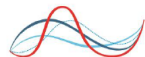
EU BIG DATA HACKATHON EXEC SUMMARY

EU Big Data Hackathon



FIRST EU BIG DATA HACKATHON IN DETAIL

1st EU Big Data Hackathon



DRŽAVNI ZAVOD ZA STATISTIKU
REPUBLIKE HRVATSKE
CROATIAN BUREAU OF STATISTICS

OBJAVA REZULTATA JAVNOG POZIVA ZA SUDJELOVANJE NA EUROPSKOM BIG DATA HACKATHONU

Izuzetno nam je zadovoljstvo obznaniti da je javni poziv Državnog zavoda za statistiku za prijavu natjecateljskih timova za sudjelovanje na Europskom Big Data Hackathonu izazvao veliki interes stručne javnosti te smo zaprimili mnoštvo nadasve kvalitetnih pojedinačnih i timskih prijava.

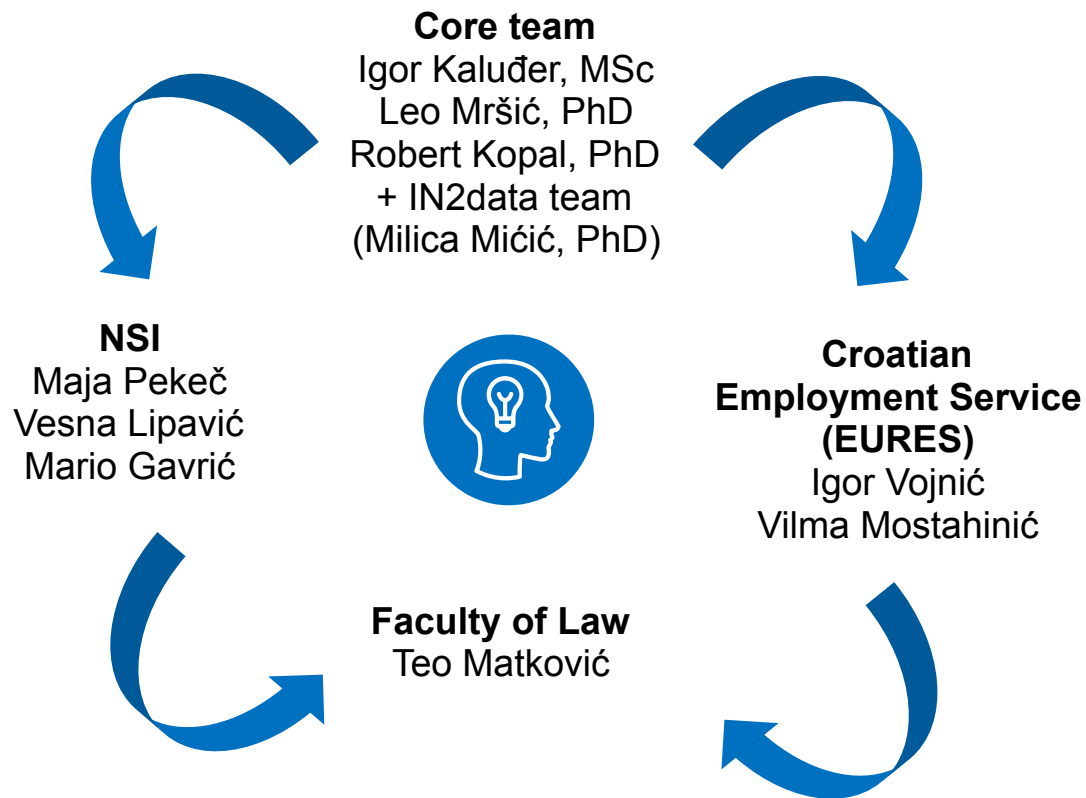
Nakon pomnog pregleda svih pristiglih prijava, stručni ocjenjivački sud kojeg su činili Saša Cvetojević, direktor Insaco d.o.o., Luka Ljubičić, pomoćnik ravnatelja za informatiku Hrvatskog zavoda za mirovinsko osiguranje te Marko Krištof, ravnatelj Državnog zavoda za statistiku, donio je odluku da će na Europskom Big Data Hackathonu u ožujku 2017., Hrvatsku predstavljati tim kojeg čine doc. dr. sc. Robert Kopal, dr. sc. Leo Mršić i mr. sc. Igor Kaluder.

Zahvaljujemo svim pojedincima koji su iskazali svoj interes i motiviranost za sudjelovanje na Europskom Big Data Hackathonu te moramo naglasiti da izbor nije bio nimalo lak. Pred odabranim timom nalazi se veliki izazov te im želimo puno sreće!

Zagreb, 11. studenoga 2016.



Team Croatia



Competition

First Big Data hackathon organized by European Commission and EUROSTAT

- ✓ **22** teams (65 experts)
- ✓ **21** countries
- ✓ **2 days & 3 nights**
- ✓ 6 minutes pitch per country



Data Sets

4.7 mil EURES CVs

35.0 mil EURES job ads

7.4 mil web scraped job ads

1.0 mil records from EU Statistics on Income and Living Conditions Survey (SILC)

0.2 mil records from OECD Programme for the International Assessment of Adult Competences Survey (PIAAC)

1.4 mil records from EU Labour Force Survey (LFS)

...other surveys

Policy Question

Policy question

How would you support the design of policies for reducing mismatch between jobs and skills at regional level in the EU through the use of data?

Demonstrate your work through a tool that performs data analytics on the provided data and potentially additional data sources.

Your proposal should take into account various approaches in matching "demand for" and "supply of" skills either by providing relevant training, outsourcing or hiring employees. The proposal should be supported by relevant data, statistical analysis and visualisation. You are also invited to establish links between various data sets and to identify and use other data sources beyond those currently at hand. The tool should integrate or at least consider the aspect of data quality for the analysis and visualisation.

The tool could interactively provide answers to specific questions as shown in the following **examples**:

- How can mismatches between skills demand and skills supply in the labour market be measured? Which regions have reported the biggest mismatches?
- How can mismatches between skills demand and skills provided by education system be measured?
- Which regions provide higher training opportunities (formal-informal education, on the job training)? How can this be measured?
- Can you identify regions with high/low labour mobility (senders/receivers of skills)?
- Which regions have the highest potential to reduce skills mismatch through intra-EU labour mobility?
- How business environment (legal, financial, economic, labour, entrepreneurship culture, etc.) is linked to skills demand/supply/(mis)match in the EU regions?
- What are the most (and least wanted) jobs at regional level in the EU? Can the top (bottom) 10 jobs be identified at regional level in the EU?
- What are the most (and least wanted) skills at regional level in the EU? Can the top (bottom) 10 skills be identified at regional level in the EU?

Policy Question

How would you **support** the **design of policies** for **reducing mismatch** between **jobs and skills** at **regional level** in the EU through the use of data?

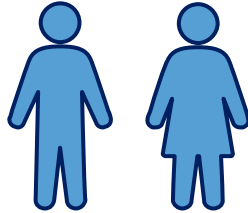
Demonstrate your work **through a tool that performs data analytics** on the provided data and potentially additional data sources.

Your proposal should take into account **various approaches** in matching "demand for" and "supply of" skills either by providing relevant training, outsourcing or hiring employees. The proposal should be **supported by relevant data, statistical analysis and visualisation**. You are also invited to establish **links between various data sets** and to identify and use other data sources beyond those currently at hand. The tool should integrate or at least consider the aspect of data quality for the analysis and visualisation. The tool could interactively provide answers to specific questions as shown in the following **examples**:

- How can **mismatches** between **skills demand and skills supply** in the labour market be **measured**? Which **regions** have reported the **biggest mismatches**?
- How can **mismatches** between **skills demand and skills** provided by **education system** be **measured**?
- Which **regions** provide **higher training opportunities** (formal-informal education, on the job training)? How can this be **measured**?
- Can you **identify regions** with **high/low labour mobility** (senders/receivers of skills)?
- Which **regions** have the **highest potential to reduce skills mismatch** through intra-EU labour mobility?
- How **business environment** (legal, financial, economic, labour, entrepreneurship culture, etc.) is **linked to skills** demand/supply/ (mis)match in the EU regions?
- What are the **most** (and **least wanted**) **jobs** at regional level in the EU? Can the top (bottom) 10 jobs be identified at regional level in the EU?
- What are the **most** (and **least wanted**) **skills** at regional level in the EU? Can the top (bottom) 10 skills be identified at regional level in the EU?

Skill Decomposition & Matching

Qualification



Occupation
(ESCO ≈5.000)



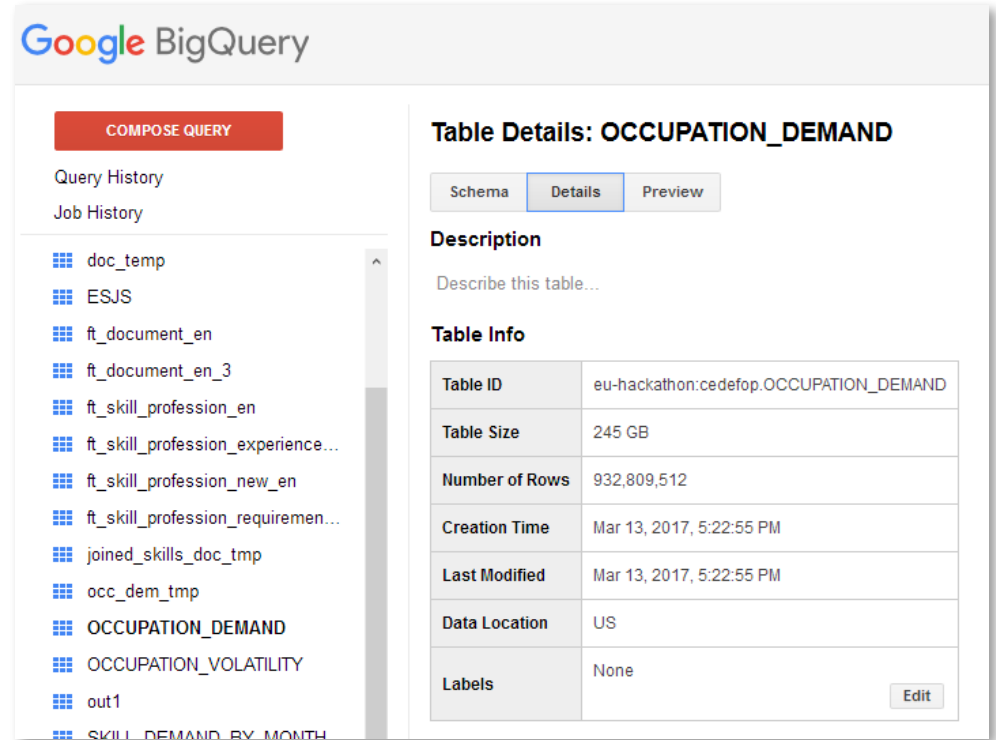
Skills/Competences
(ESCO ≈6.000)



Three overlapping resumes for John Bravo, PMP Certified Project and Organizational Manager. The resumes are white with black text and feature a small portrait photo of John Bravo in the top right corner. Each resume includes sections for Key Skills Assessment, Professional Experience, and Certifications & Memberships. The resumes are stacked, with the top one being the most prominent. The resumes are for the same individual, John Bravo, and show consistent information across all three, such as his title, contact information, and professional experience. The resumes are presented in a way that shows they are multiple copies of the same document, slightly offset to show depth.

Output

- ✓ Platform selection
- ✓ Data preparation
- ✓ Data Storage
- ✓ Data Optimization
- ✓ Project Scope Definition
- ✓ Data Processing / Analytics
- ✓ Visualization
- ✓ Presentation Bundle

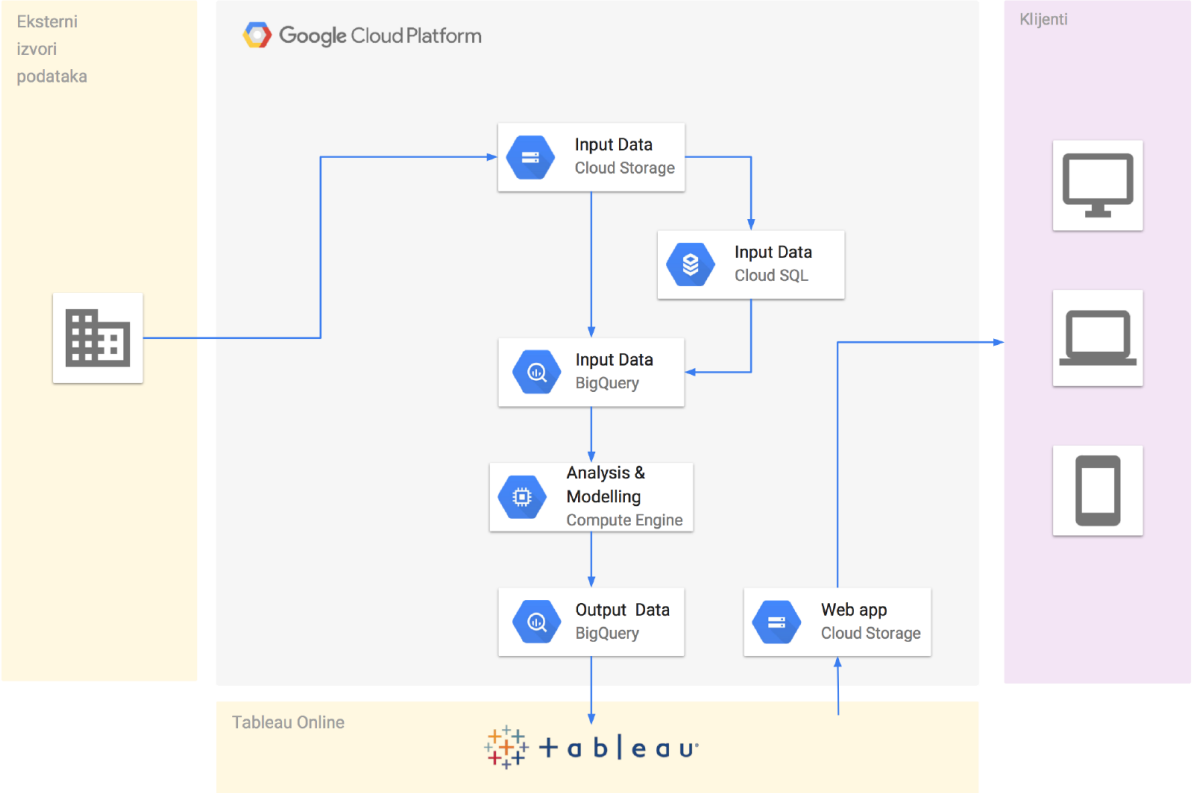


The screenshot displays the Google BigQuery interface. On the left, there is a sidebar with a red 'COMPOSE QUERY' button and a list of tables including 'doc_temp', 'ESJS', 'ft_document_en', 'ft_document_en_3', 'ft_skill_profession_en', 'ft_skill_profession_experience...', 'ft_skill_profession_new_en', 'ft_skill_profession_requiremen...', 'joined_skills_doc_tmp', 'occ_dem_tmp', 'OCCUPATION_DEMAND', 'OCCUPATION_VOLATILITY', 'out1', and 'SKILL_DEMAND_BY_MONTH'. The main area on the right is titled 'Table Details: OCCUPATION_DEMAND' and features three tabs: 'Schema', 'Details' (which is selected), and 'Preview'. Below the tabs, there is a 'Description' section with the text 'Describe this table...'. The 'Table Info' section contains a table with the following data:

Table ID	eu-hackathon:cedefop.OCCUPATION_DEMAND
Table Size	245 GB
Number of Rows	932,809,512
Creation Time	Mar 13, 2017, 5:22:55 PM
Last Modified	Mar 13, 2017, 5:22:55 PM
Data Location	US
Labels	None

An 'Edit' button is located at the bottom right of the 'Table Info' section.

Platform



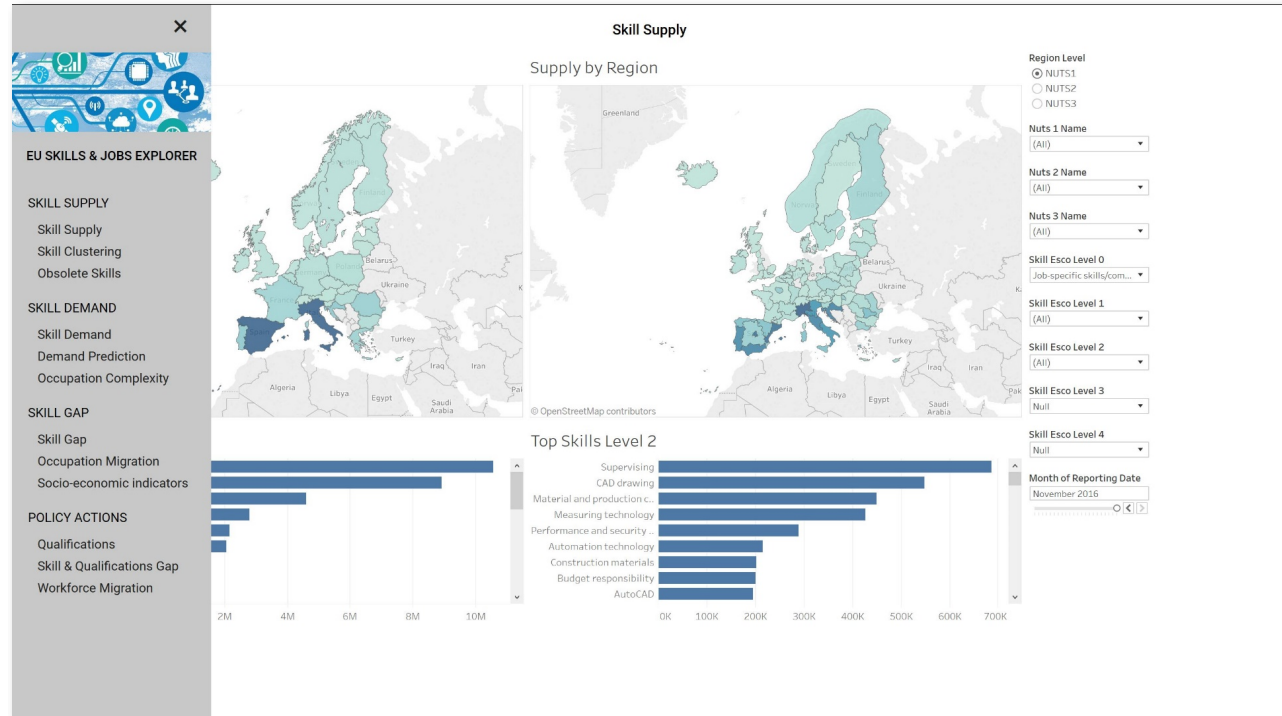
PROTOTYPE

Solution

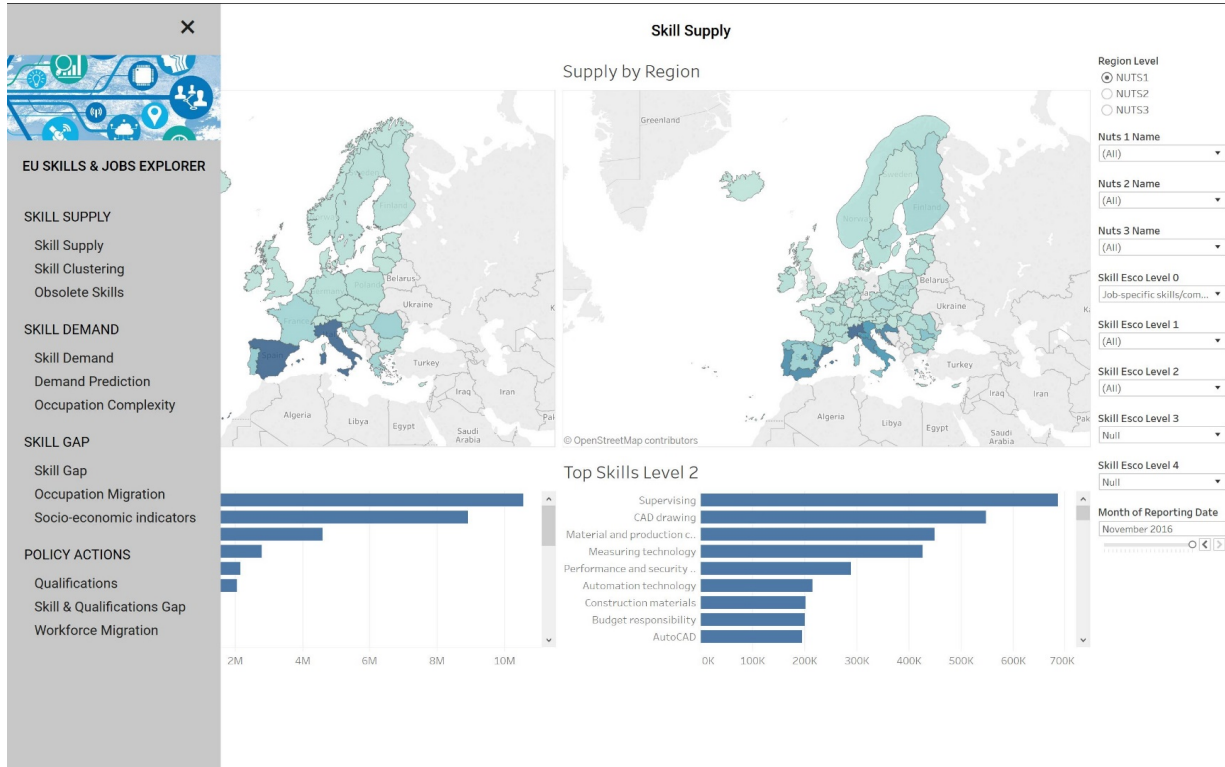
Live cloud/web based application, visual, interactive dashboards (region maps included)

Key areas (4):

- 1) Skill supply
 - Skill Supply
 - Skill Clustering
 - Obsolete Skills
- 2) Skill Demand
 - Skill Demand
 - Skill Clustering
 - Demand Prediction
 - Occupation Complexity
- 3) Skill Gap
 - Skill Gap
 - Occupation Migration
 - Socio-economic Indicators
- 4) Policy Actions
 - Qualifications
 - Skill & Qualifications Gap
 - Workforce Migration



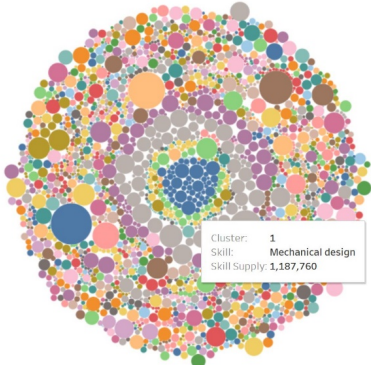
Skill supply



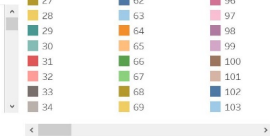
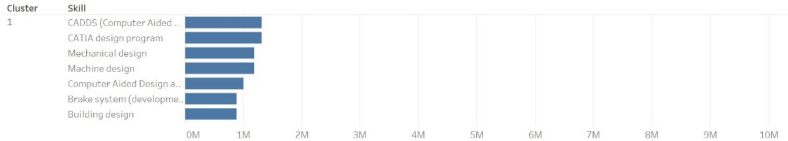
Skill Clustering



Skill Clustering



Skill Supply



Occupation Migration



Occupation Migration

Current Occupation	Desired Occupation									
	Armed forces occupations	Clerical support workers	Craft and related trades	Elementary occupations	Managers	Plant and machine operators	Professionals	Service and sales workers	Skilled agricultural, forestry and fishery workers	Technicians and associate professional occupations
Armed forces occupations	150	362	363	198	858	420	889	974	110	982
Clerical support workers	34	66,309	4,916	11,034	12,105	5,785	53,706	25,241	2,048	35,745
Craft and related trades	7	4,864	53,607	8,576	1,726	10,680	11,248	7,648	2,232	13,250
Elementary occupations	3	8,290	6,433	18,990	1,763	4,654	10,666	12,859	2,431	7,484
Managers	66	12,186	3,590	2,723	72,003	2,819	69,472	11,232	1,354	35,418
Plant and machine operators		4,506	7,278	6,028	1,234	25,942	5,524	5,417	1,288	6,189
Professionals	84	41,741	12,856	10,170	69,488	6,696	566,011	33,845	4,815	140,465
Service and sales workers	99	36,321	10,407	19,638	11,940	9,106	55,914	74,195	2,864	38,501
Skilled agricultural, forestry and fishery workers		730	1,014	1,779	679	1,658	2,378	984	3,918	1,415
Technicians and associate professional occupations	155	41,735	15,748	12,143	36,680	9,289	149,531	32,248	2,754	161,189

Occupation Level

- 0
- 1
- 2
- 3

Reporting Month

November 2016

Nuts0 Name

- (All)
- BELGIQUE-BELGIË
- ČESKÁ REPUBLIKA
- DANMARK
- DEUTSCHLAND

Nuts1 Name

- (All)
- ÅLAND
- ALFÖLD ÉS ÉSZAK
- ATTIKH (ATTIKI)
- BADEN-WÜRTTEMBERG

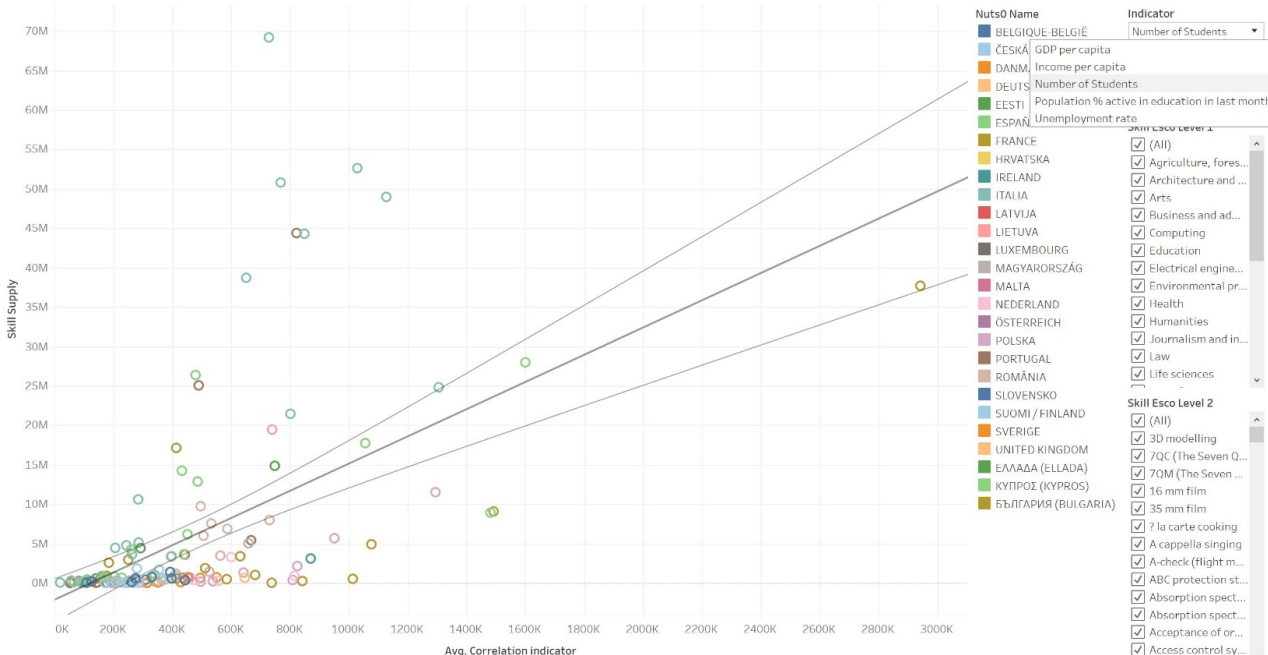
Nuts2 Name

- (All)
- Abruzzo
- Åland
- Alentejo

Socio-economic Indicators



Socio-economic indicators



Occupation Complexity



Occupations

Physical and engineering science tec..	384
Engineering professionals (excluding ..	345
Sales and purchasing agents and bro..	322
Client information workers	321
Administrative and specialised secre..	318
Shop salespersons	317
Material-recording and transport cle..	316
Sales, marketing and public relations ..	311
Manufacturing, mining, construction,...	306
Administration professionals	305
Software and applications developer..	304
General office clerks	294
Information and communications tec..	294
Electrical equipment installers and re..	290
Business services and administration..	289
Machinery mechanics and repairers	285
Sheet and structural metal workers. ...	284
Manufacturing labourers	283
Electrotechnology engineers	279
Mining, manufacturing and construct..	274
Mobile plant operators	274
Architects, planners, surveyors and d..	272
Business services agents	271
Transport and storage labourers	271
Other clerical support workers	269
Building finishers and related trades ..	262
Blacksmiths, toolmakers and related ..	257
Financial and mathematical associat..	255
Sales, marketing and development m..	255
Building frame and related trades wo..	246
Database and network professionals	246
Other sales workers	245
Secretaries (general)	243
Mining and construction labourers	242

Occupation Complexity

Skills

3D modelling
Accounting (adult educati..
Accounting / auditing
Accounting responsibility
Active Server Pages (ASP)
Actuarial science
Administration of wages
Adobe Photoshop
Advanced Business Applic..
Advertising services
Agility
Agricultural science
Agriculture
Air decontamination
Aircraft electronics
Aircraft technology
Airport management
Ambulance care (nursing)
Apache web server
APL (programming langua..
Applications programming
Applied mathematics (tea..
Applied statistics
Architectural models
Archiving
Arts and crafts/hobby ma..
Asbestos decontamination
Assembly (wood)
Assembly of parts
Asylum and immigration l..
Attitudes
Audiovisual engineering
Auditing

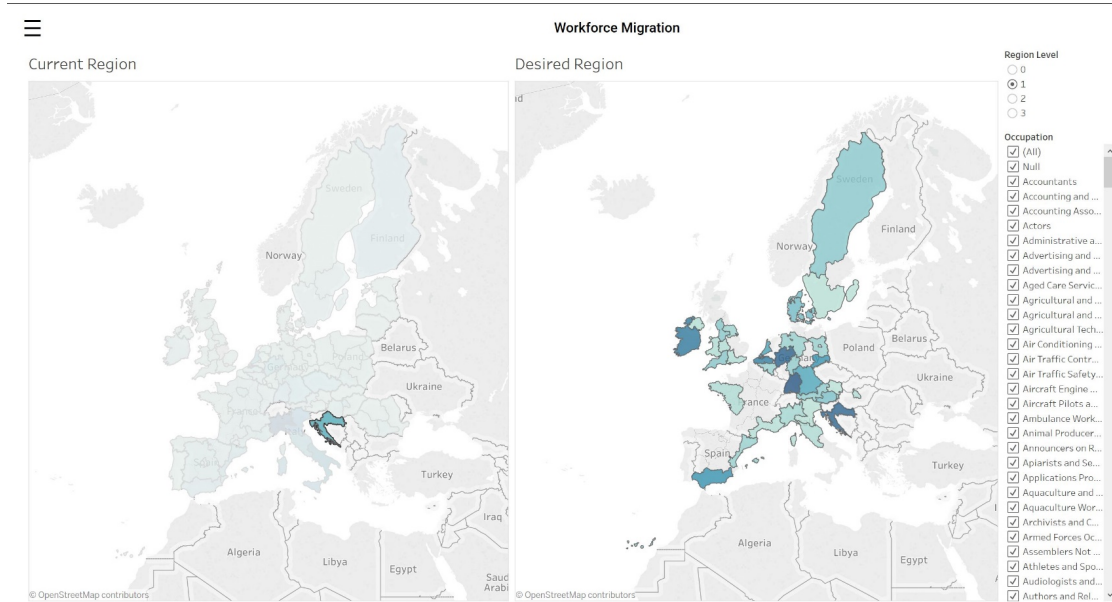
Occupation Level

- 0
- 1
- 2
- 3
- 4

Skill Level

- 0
- 1
- 2
- 3

Workforce Migration



Statistical panel members

Members of the statistical panel of evaluators of the European Big Data Hackathon

Eurostat	Dario Buono	Dario.Buono@ec.europa.eu	Yes	Yes
Eurostat	Christine Kormann	Christine.Kormann@ec.europa.eu	Yes	Yes
Eurostat	Jean-Louis Mercy	Jean-Louis.Mercy@ec.europa.eu	Yes	Yes
DG-EMPL	Aldo Laudi	Aldo.Laudi@ec.europa.eu	Yes	Yes
DG-EAC	Geir Ottestad	Geir.Ottestad@ec.europa.eu	Yes	Yes
CEDEFOP	Pascaline Descy	Pascaline.Descy@cedefop.europa.eu	Yes	Yes
DIGIT	Tanya Chetcuti	Tanya.Chetcuti@ec.europa.eu	Yes	Yes
DG-CNECT	Stefano Abruzzini	Stefano.Abruzzini@ec.europa.eu	Yes	Yes
Academia	Silvia Biffignandi	silvia.biffignandi@unibg.it	Yes	Yes
Academia	Tobias Blanke	tobias.blanke@kcl.ac.uk	Yes	Yes

Industry panel members

Members of the industry panel of evaluators of the European Big Data Hackathon

Oracle	Mats Stellwall	mats.stellwall@oracle.com	Yes	Yes
SAS	Michel Philippens	michel.philippens@sas.com	Yes	
Amazon	Yotam Yarde	yotamy@amazon.com	Yes	Yes
Microsoft	Frank Callewaert	frankcal@microsoft.com	Yes	Yes
SAP	Gregory Neuveglise	gregory.neuveglise@sap.com gneuveglise@yahoo.fr	Yes	Yes
Everis	Sonia Taborcia Hernández	Sonia.Taborcia.Hernandez@everis.com	Yes	Yes
IBM	Herman Nielens	herman.nielens@be.ibm.com	Yes	Yes
QLIK	Jurgen Hamelrijckx	Jurgen.Hamelrijckx@qlik.com	Yes	Yes
AGILOS	Edwin Willems	edwin.willems@agilos.com	Yes	Yes
Accenture	Thibaut Roisin	thibaut.roisin@accenture.com	Yes	Yes

EU BD HACK / PROTOTYPE CREDENTIALS

Live Prototype

url: **eubdhack.in2data.eu**

username: **eubdhack@in2data.eu**

password: **EUBDHACK2017**

Q&A