



Project acronym: **EDSA**  
Project full name: **European Data Science Academy**  
Grant agreement no: **643937**

## **D1.5 Demand and supply analysis report**

Deliverable Editor: **Alexander Mikroyannidis (OU)**  
Other contributors: **Inna Novalija (JSI), Erik Novak (JSI), Simon Scerri (Fraunhofer),  
Diego Collarana Vargas (Fraunhofer)**  
Deliverable Reviewers: **Joos Buijs (TUE), Huw Fryer (SOTON)**  
Deliverable due date: **31/01/2018**  
Submission date: **26/01/2018**  
Distribution level: **P**  
Version: **1.0**

This document is part of a research project funded  
by the Horizon 2020 Framework Programme of the European Union



## Change Log

<b>Version</b>	<b>Date</b>	<b>Amended by</b>	<b>Changes</b>
0.1	20/11/2017	Alexander Mikroyannidis	Outline and responsibilities of contributors.
0.2	11/12/2017	Erik Novak, Inna Novalija	Content input.
0.3	10/01/2018	Diego Collarana, Simon Scerri	Content input.
0.4	11/01/2018	Alexander Mikroyannidis	Version for internal review.
0.5	25/01/2018	Alexander Mikroyannidis	Revised version.
1.0	26/01/2018	Alexander Mikroyannidis	Final QA.

## Table of Contents

Change Log.....	2
Table of Contents.....	3
List of Tables.....	3
List of Figures.....	3
1. Executive Summary.....	5
2. Introduction.....	6
3. Dashboard user interface.....	6
4. Dashboard software architecture.....	10
4.1 Harvesting jobs.....	12
4.2 Harvesting skills.....	14
4.3 Harvesting courses.....	15
5. Dashboard evaluation.....	16
5.1 Evaluation Goals.....	16
5.2 Evaluation Setup.....	17
5.3 Evaluation Results and Discussion.....	18
6. Conclusion.....	21
7. References.....	21

## List of Tables

Table 1: Summary of demand sources.....	13
Table 2: Summary of supply sources. ....	15

## List of Figures

Figure 1: Job search in the EDSA dashboard. ....	7
Figure 2: Job search with the map view selected.....	8
Figure 3: Job search with the courses view selected. ....	8
Figure 4: The dashboard search filters. ....	9
Figure 5: Data science learning pathways in the EDSA courses portal. ....	9
Figure 6: Building a personalised learning pathway. ....	10
Figure 7: Example of a job posting returned by the query “machine learning”. Even though the job posting is written in Spanish the methodology finds it relevant. ....	11

Figure 8: On the left the ten most demanded skills histogram, and on the right the number of job positions timeline, for the query “machine learning”. Hovering over the histogram column shows the number of queried jobs demanding the skill.-----	11
Figure 9: The EDSA Data acquisition process.-----	12
Figure 10: Top fifteen countries with most found job postings. The greatest number of job postings were found for UK, followed by France and Germany. -----	14
Figure 11: Top fifteen most demanded skills. They are mostly comprised of high-level skills, such as “database” and “computer science”, and specific programming languages. -----	14
Figure 12: The distribution of online courses by course providers. The most courses were acquired from Coursera, followed by Hackr.io.-----	16
Figure 13: The NASA Task Load Index used during the Dashboard usability evaluation. -----	17
Figure 14: The Post-Study Usability Questionnaire used during the Dashboard usability evaluation.	18
Figure 15: Workload analytics from Task 1 (T1) - Task 5 (5). Mean (M) and Standard deviation (STD) are calculated. Green tones show positively lowest ratings, and red tones the opposite higher ratings. -----	19
Figure 16: Boxplot of self-perceived success in accomplishing each task, according to NASA LTX results.-----	19
Figure 17: Boxplot of how insecure, discouraged, irritated, stressed, and annoyed were the participants during the task execution, according to NASA LTX results.-----	20
Figure 18: Post-Study Usability Questionnaire (PSSUQ)-----	20



## 1. Executive Summary

WP1 primarily deals with the design and execution of a series of studies on the demand for data science training. These studies have produced an evidence base of demand for data science training as well as listing gaps in any existing training across Europe. The results of this study have been analysed in order to inform the iterations of the proposed curricula (WP2) and resultant training (WP3).

Following the outcomes of the M18 project review, WP1 has been amended to also include an analysis of the supply of learning resources, in order to match the identified demand. The results of this demand and supply analysis are reflected in the EDSA dashboard. Supply information is collected semi-automatically, with a focus on relevant institutions from Europe as well as online offerings. The information is analysed automatically to identify relevant skills, which are then mapped to job descriptions.

This deliverable reports the types of supply analyses conducted in relation to the demand analysis and how this work has been implemented in the EDSA dashboard. This deliverable also describes how the demand and supply data are presented in the dashboard and how users are able to interact with this data in order to explore the current demand and supply.

## 2. Introduction

In today's job market the required skills are constantly evolving. This can be seen in more technical fields such as web development and data science where new tools and libraries are developed and available to the public with an increasing rate. This is visible in both research and industry sectors where a job position might require a previously unseen skill and the applicant needs to learn it to be qualified. Finding the courses that would give the skill knowledge can be tedious and does not guarantee its sufficiency.

The EDSA dashboard<sup>1</sup> connects the job market skill demand with the courses that give the required skill knowledge. The EDSA dashboard enables users to search for their desired job position, find out what is the required skill set and which are the appropriate learning materials and courses to acquire the missing skills. We have focused on job positions that require data science skills and courses that are provided by acknowledged course providers. Additionally, the dashboard shows the most demanded skills and hiring location for the given results.

The main contributions of the EDSA dashboard are a) creating a sizable data set of data science related job postings containing the job postings title, description, locations and other information, and b) developing a dashboard which for a given query shows relevant job postings as well as courses and lectures which give the appropriate skills. The dashboard is daily updated with new job postings showing the most recent changes. Basic statistics such as the most popular job locations and skills are also shown.

The remainder of this deliverable is organised as follows. First, an overview of the user interface of the dashboard is provided, followed by an in-depth look into the back-end architecture. We then present the mechanisms behind harvesting various types of data for the dashboard, namely jobs, skills and courses. Finally, the user evaluation of the dashboard is presented and the deliverable is concluded.

## 3. Dashboard user interface

The EDSA dashboard allows users to explore both the current data science skills demand and supply. Users of this dashboard are able not only to explore the current demand in the data science market, but also find learning materials and training relevant to the skills they will need to secure a specific job position. Additionally, users are supported in building personalised learning pathways, consisting of courses and learning materials that will help them reach their learning goals.

In particular, the EDSA dashboard allows users to:

- View the current demand for data science jobs and skills across Europe.
- Filter demand by required skills and region.
- View trends and statistics regarding data science jobs and skills for a given timeframe.
- Explore the current supply of courses and learning materials that will help them acquire certain skills.
- Build personalised learning pathways towards acquiring certain skills.

The following figures show different views of the EDSA dashboard. Figure 1 shows a job search performed via the dashboard. This view is deliberately kept as simple as possible. In this view, queries typed into the search box at the top result in a simple list of related data science jobs. Selecting any job results in additional details of the post being displayed. The toolbar below the query entry box allows users to add or remove additional views.

---

<sup>1</sup> <http://edsa-project.eu/resources/dashboard/>



SEARCH...

SEARCH [Settings] [Next]

LIST SKILLS/TIMELINE MAP COURSES

**JOB LIST**

10770 JOBS FOUND OUT OF 4664880  
TIME INTERVAL: 12/11/2017 - TODAY

<p><b>GRADUATE TECHNICAL SALES REPRESENTATIVE</b></p> <p>Hp, Spain, Ireland PUBLISHED ON JANUARY 7, 2018</p> <p>DESCRIPTION ...Science or similar. Typically 1-2 years technical and/or solution experience in IT industry preferred Knowledge and Skills Required General knowledge across... Hace +30 días en Wizbii</p>	<p><b>BLUECAMP: IBM SPAIN TECHNOLOGY INTERNSHIP PROGRAM</b></p> <p>ibm, Madrid, Spain PUBLISHED ON JANUARY 7, 2018</p> <p>big data</p> <p>DESCRIPTION Would you like to put into practice your knowledge about technology in different areas? Are you interested in learning more about Big Data, Cloud... Hace +30 días en Wizbii</p>
<p><b>WAREHOUSE</b></p> <p>Johnson And Johnson, Madrid, Spain PUBLISHED ON JANUARY 7, 2018</p> <p>DESCRIPTION Caring for the world, one person at a time. Inspires and unites the people of Johnson &amp; Johnson. We embrace research and science. Bringing innovative ideas... Hace +30 días en Wizbii</p>	<p><b>INTERN-PS (I)</b></p> <p>Spain PUBLISHED ON JANUARY 7, 2018</p> <p>DESCRIPTION Intern-PS (I) **Description:**Big Data Junior Engineer (Recent Graduated) internship** Position Description**We arelooking for recent graduated in Bachelor... Hace 21 días en MegaJobs</p>
<p><b>RESEARCH AND DEVELOPMENT ENGINEER (H/F)</b></p> <p>Innoha, Barcelona, Spain PUBLISHED ON JANUARY 7, 2018</p> <p>big data</p> <p>DESCRIPTION Poste et missions: You will advise our clients to which technology stacks / cloud / big data solutions to use Automating the operations and production of... Hace +30 días en Wizbii</p>	<p><b>SOFTWARE ENGINEER (BRITE: BILL)</b></p> <p>Amdocs, Madrid, Spain PUBLISHED ON JANUARY 7, 2018</p> <p>big data</p> <p>DESCRIPTION ...ever more connected and digital world. At Amdocs, we are leading the digital revolution into the future. From virtualized telecommunications networks, Big Data... Hace +30 días en Jobs2Web</p>

**Figure 1: Job search in the EDSA dashboard.**

In Figure 2, a map view has been selected. Google maps are used for the map view incorporating zoom facilities. Selecting any anchor point in the map brings up details about the job. In Figure 3, the courses view has been selected. The courses view shows recommended courses related to the query, which are offered by the EDSA project consortium and external organisations. This view also displays recommended learning pathways based on the performed search.

**JOB LIST**

10770 JOBS FOUND OUT OF 4664880  
TIME INTERVAL: 12/11/2017 - TODAY

**GRADUATE TECHNICAL SALES REPRESENTATIVE**

Hp, Spain, Ireland  
PUBLISHED ON JANUARY 7, 2018

**DESCRIPTION**

...Science or similar. Typically 1-2 years technical and/or solution experience in IT industry preferred  
Knowledge and Skills Required General knowledge across... Hace +30 dias en Wizbii

**BLUECAMP: IBM SPAIN TECHNOLOGY INTERNSHIP PROGRAM**

ibm, Madrid, Spain  
PUBLISHED ON JANUARY 7, 2018

**big data**

**DESCRIPTION**

Would you like to put into practice your knowledge about technology in different areas? Are you interested in learning more about Big Data, Cloud... Hace +30 dias en Wizbii

**MAP**

Shows the number of locations that offer job positions for the given query. Clicking on the pin gives the location name and number of jobs found for that location.

Figure 2: Job search with the map view selected.

**JOB LIST**

10770 JOBS FOUND OUT OF 4664880  
TIME INTERVAL: 12/11/2017 - TODAY

**GRADUATE TECHNICAL SALES REPRESENTATIVE**

Hp, Spain, Ireland  
PUBLISHED ON JANUARY 7, 2018

**DESCRIPTION**

...Science or similar. Typically 1-2 years technical and/or solution experience in IT industry preferred  
Knowledge and Skills Required General knowledge across... Hace +30 dias en Wizbii

**BLUECAMP: IBM SPAIN TECHNOLOGY INTERNSHIP PROGRAM**

ibm, Madrid, Spain  
PUBLISHED ON JANUARY 7, 2018

**big data**

**DESCRIPTION**

Would you like to put into practice your knowledge about technology in different areas? Are you interested in learning more about Big Data, Cloud... Hace +30 dias en Wizbii

**WAREHOUSE**

Johnson And Johnson, Madrid, Spain  
PUBLISHED ON JANUARY 7, 2018

**DESCRIPTION**

Caring for the world, one person at a time. Inspires and unites the people of Johnson & Johnson. We embrace research and science. Bringing innovative ideas... Hace +30 dias en Wizbii

**LEARNING PATHWAYS**

The learning pathway in showing the courses that will give you the skills related to the query.

**PERSONALIZE YOUR PATHWAY**

- Mathematics of computing
- Computing methodologies
- Big Data (Data Science) applications design
- Business Process Management

**COURSES**

The recommended courses provided by EDSA and other course providers.

- DATA SCIENCE MATH SKILLS
- OPEN DATA SCIENCE
- DATA SCIENTIST BASIC
- PROCESS MINING: DATA SCIENCE IN ACTION

Figure 3: Job search with the courses view selected.



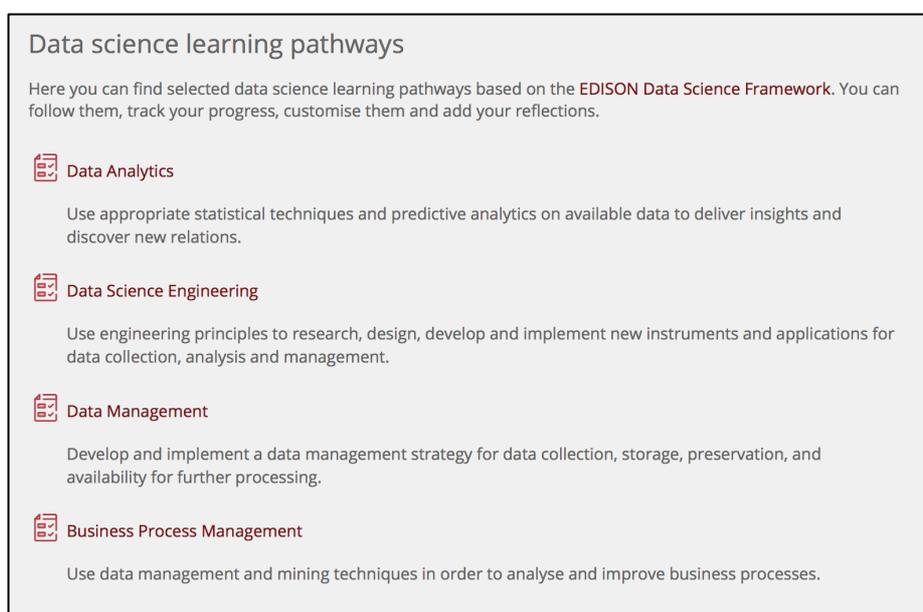
Figure 4 shows additional filters that users can enable when searching. These filters allow users to specify the required skills, the location of jobs, as well as the start and end time of job adverts.

The image shows a search interface with a dark grey background. At the top, there is a search bar with the placeholder text 'SEARCH...' and a 'Search' button. Below the search bar, there are three filter sections: 'Required Skills:' with a text input field containing 'Type a skill'; 'Search Locations:' with a text input field containing 'City or country names'; and 'Select Time:' with two date input fields labeled 'Start date' and 'End date', and a 'Clear' button followed by 'Last week' and 'Last month' options.

**Figure 4: The dashboard search filters.**

The EDSA dashboard not only allows users to explore the current listings of jobs and associated learning resources, but also structures the recommended learning resources into learning pathways, which can be further customised and personalised by users. In order to build their personalised pathways, users of the dashboard start by searching for certain job positions. Based on their searches, the dashboard recommends courses and learning pathways for gaining the required skills. Users may follow these pathways or further personalise them by visiting the EDSA courses portal.<sup>2</sup>

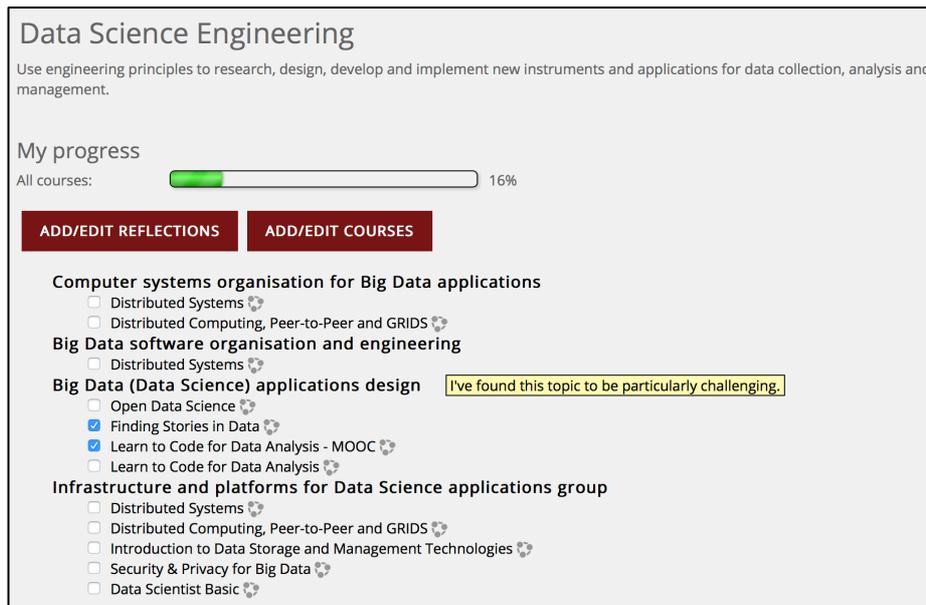
Figure 5 shows the list of data science learning pathways currently offered by the EDSA courses portal. These pathways have been adapted from the EDISON Data Science Framework<sup>3</sup> and consist of recommended data science topics, as well as courses for acquiring certain sets of skills related to these topics. Users can use these pathways as templates in order to build their own pathways by adding courses, monitoring their progress towards completing their pathways, as well as reflecting on the contents of the pathways and on what they have learned, as shown in Figure 6.

The image shows a section titled 'Data science learning pathways'. Below the title, there is a paragraph: 'Here you can find selected data science learning pathways based on the EDISON Data Science Framework. You can follow them, track your progress, customise them and add your reflections.' Below this paragraph, there are four pathway entries, each with a red icon of a document with a checkmark and a title: 'Data Analytics' (Use appropriate statistical techniques and predictive analytics on available data to deliver insights and discover new relations.), 'Data Science Engineering' (Use engineering principles to research, design, develop and implement new instruments and applications for data collection, analysis and management.), 'Data Management' (Develop and implement a data management strategy for data collection, storage, preservation, and availability for further processing.), and 'Business Process Management' (Use data management and mining techniques in order to analyse and improve business processes.).

**Figure 5: Data science learning pathways in the EDSA courses portal.**

<sup>2</sup> <http://courses.edsa-project.eu>

<sup>3</sup> <http://edison-project.eu/edison/edison-data-science-framework-edsf>



**Data Science Engineering**  
Use engineering principles to research, design, develop and implement new instruments and applications for data collection, analysis and management.

My progress  
All courses:  16%

[ADD/EDIT REFLECTIONS](#) [ADD/EDIT COURSES](#)

- Computer systems organisation for Big Data applications**
  - Distributed Systems
  - Distributed Computing, Peer-to-Peer and GRIDS
- Big Data software organisation and engineering**
  - Distributed Systems
- Big Data (Data Science) applications design** I've found this topic to be particularly challenging.
  - Open Data Science
  - Finding Stories in Data
  - Learn to Code for Data Analysis - MOOC
  - Learn to Code for Data Analysis
- Infrastructure and platforms for Data Science applications group**
  - Distributed Systems
  - Distributed Computing, Peer-to-Peer and GRIDS
  - Introduction to Data Storage and Management Technologies
  - Security & Privacy for Big Data
  - Data Scientist Basic

**Figure 6: Building a personalised learning pathway.**

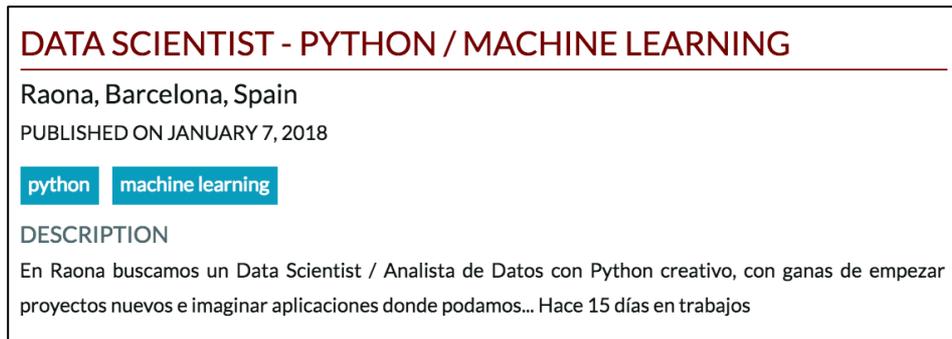
#### 4. Dashboard software architecture

In this section, we present the content retrieval methodology and describe the different components of the dashboard. The content is retrieved by inserting a query text in the search input. The user may add additional query conditions by selecting the Data Science skills, locations, countries and a time interval in which the job postings were published. Upon submitting, the query is used to fetch the content that matches the conditions. While all query values are used for retrieving job postings, only the input text and skills are used for retrieving the courses and video lectures content. Since courses and video lectures are available online the location and time interval are irrelevant for retrieving the supply content.

To retrieve the content, we first need to set an appropriate index. The job posting data set is indexed by Wikipedia concepts, Data Science skills, locations, countries and published date while the course and lecture data sets are indexed only by Wikipedia concepts. The query text is sent through wikification to acquire Wikipedia concepts which are used for retrieving the relevant content. Next, additional query conditions are used to filter out the content. The remaining content is used to calculate the most demanded skills and hiring locations. Finally, the query results are returned and used to update the dashboard components. This process is developed using QMiner [1], a data analytics platform for processing large-scale real-time streams containing structured and unstructured data.

The dashboard is composed of different components. The largest component is a list of job postings. Each job posting is presented by its extracted information, including the Data Science skills extracted from the title and description. Figure 7 shows an example of a job posting in the list. Since Wikifier supports cross and multi-linguality the list consists of job postings written in different languages.

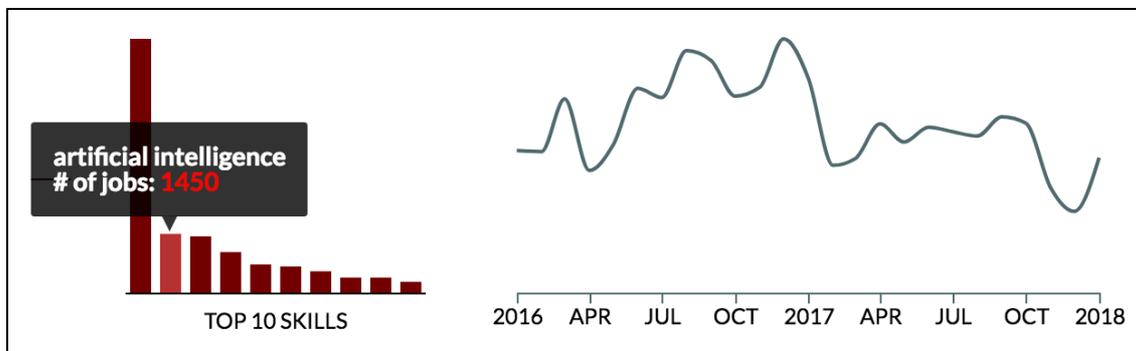




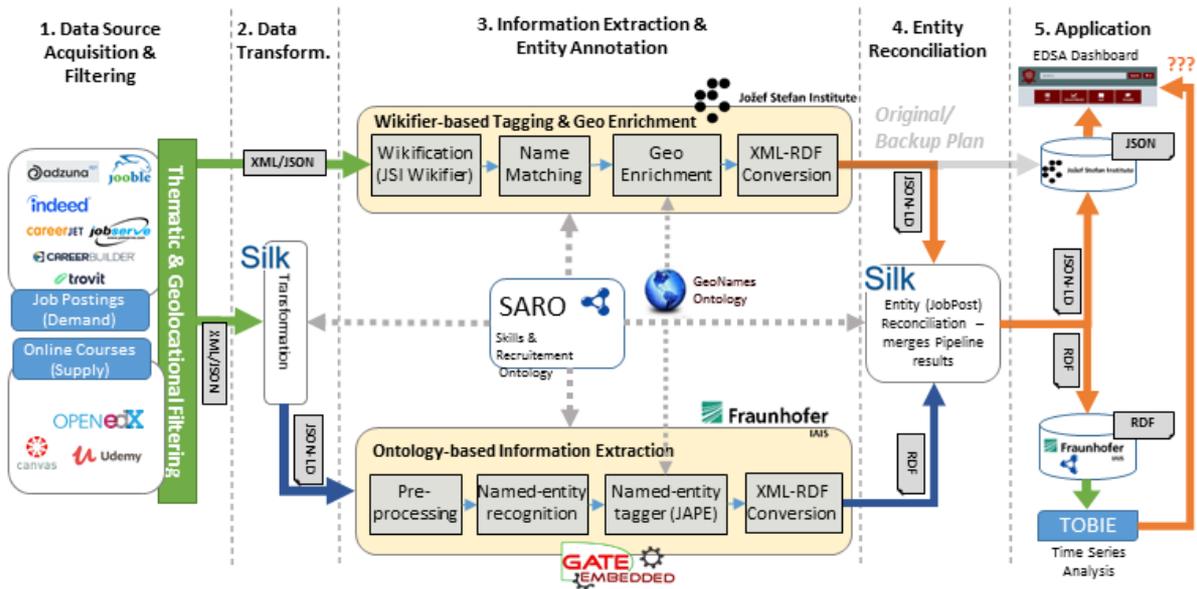
**Figure 7: Example of a job posting returned by the query “machine learning”. Even though the job posting is written in Spanish the methodology finds it relevant.**

If the user does not have the required skill set it can be acquired by enrolling into courses shown in the course list. The list shows courses offered by different online course providers that are relevant to the users input query. Figure 3 (section 3) shows the component containing the course list. Left and right arrows are used to navigate through the list where each course is presented by its name and a course provider. Additionally, the user can watch lectures to get a deeper understanding of a problem. Similar to courses the video lectures list shows relevant content found on VideoLectures.NET. Clicking the lecture redirects the user to the video lecture homepage.

The dashboard also shows them most demanded skills and job posting timeline. The timeline shows how the ratio between queried and all job postings changed since the start of the year 2016. Additionally, this shows a trend of the skill demand in the queried job posting subset. Figure 8 shows the visualizations used to show the skill demand and timeline.



**Figure 8: On the left the ten most demanded skills histogram, and on the right the number of job positions timeline, for the query “machine learning”. Hovering over the histogram column shows the number of queried jobs demanding the skill.**



**Figure 9: The EDSA Data acquisition process.**

The EDSA data acquisition process, shown in Figure 9, consists of two parallel pipelines that operate on the extracted (and transformed) demand and supply data. The Wiki-based Tagging and Geo-enrichment (WTG) pipeline shown on the top (responsible: JSI) complements the Ontology-based Information Extraction (OBIE) pipeline shown below (responsible: Fraunhofer), by discovering new skills that are not known to the ontology (SARO<sup>4</sup>). At the moment, these need to be manually added to be recognised in the future. The OBIE pipeline relies on a customised GATE<sup>5</sup> pipeline. The OBIE pipeline is better in terms of precision, whereas the WTG pipeline compensates for incomplete recall.

The results from both pipelines are integrated using the Silk Framework,<sup>6</sup> so that all the knowledge discovered is attached to the same data. Following the entity reconciliation process, results are stored for direct access by the EDSA dashboard. Separately, a snapshot of the dataset is routinely stored in order to enable time series analysis (the results are not integrated in the EDSA dashboard).

#### 4.1 Harvesting jobs

Open job positions can be found using job search services. These services aggregate job postings by location, sector, applicant qualifications and skill set or type. One such service is Adzuna,<sup>7</sup> a search engine for job ads which mostly covers English speaking countries. Another service is Trovit,<sup>8</sup> a leading search engine for classified ads in Europe and Latin America. The service is available in 13 different languages and provides listings of jobs as well as cars, real estate and other products. When applying for a job position the applicant requires to have a certain skill set. If the requirements are not fulfilled, he can enrol in courses to get the missing skills. Additionally, watching certain lectures can

<sup>4</sup> <http://vocol.iais.fraunhofer.de/saro/>

<sup>5</sup> <https://gate.ac.uk/>

<sup>6</sup> <http://silkframework.org/>

<sup>7</sup> <https://www.adzuna.com/>

<sup>8</sup> <https://www.trovit.com/>



give a deeper understanding of a particular problem which can increase the probability of getting accepted for a job position.

Jooble<sup>9</sup>, Indeed<sup>10</sup> and XING<sup>11</sup> have also been integrated as additional sources for demand data (routine extraction of job postings). Table 1 summarises all demand sources.

**Table 1: Summary of demand sources.**

<i>Demand Data Source</i>	<i>Responsible for Integration in Dashboard</i>
Adzuna	JSI
Trovit	JSI
Jooble	Fraunhofer
Indeed	Fraunhofer
XING	Fraunhofer

Since we needed a continuous flow of data, we developed a pipeline for acquiring job postings, courses and lectures. This will allow us to provide the dashboard, presented above, with the most recent data. For job postings we targeted the portals like Adzuna with an emphasis on positions in Data Science.

For data acquisition and enrichment, we collected data either using dedicated APIs, including Adzuna API,<sup>12</sup> as well as custom web crawlers. The data was formatted to JSON to aid further processing and enrichment.

The next step of data preprocessing is wikification - identifying and linking textual components to the corresponding Wikipedia pages [2]. This is done using Wikifier,<sup>13</sup> which also supports cross and multi-linguality enabling extraction and annotation of relevant information from job postings, courses and video lectures in different languages. Wikification will allow us to search for job postings, courses and lectures in multiple languages.

Next, we use the Skill and Recruitment Ontology (SARO) [3] to extract Data Science skills from job postings. For each job posting we match the Wikipedia concepts with the skills found in SARO ontology and declare the matched concepts as Data Science skills. These skills are then added to the job posting profile. Finally, to allow searching by locations and countries the job postings were further enriched by using GeoNames<sup>14</sup> ontology to include the latitude and longitude and the corresponding GeoNames ID and the location name.

---

<sup>9</sup> <https://jooble.org>

<sup>10</sup> <https://indeed.com/>

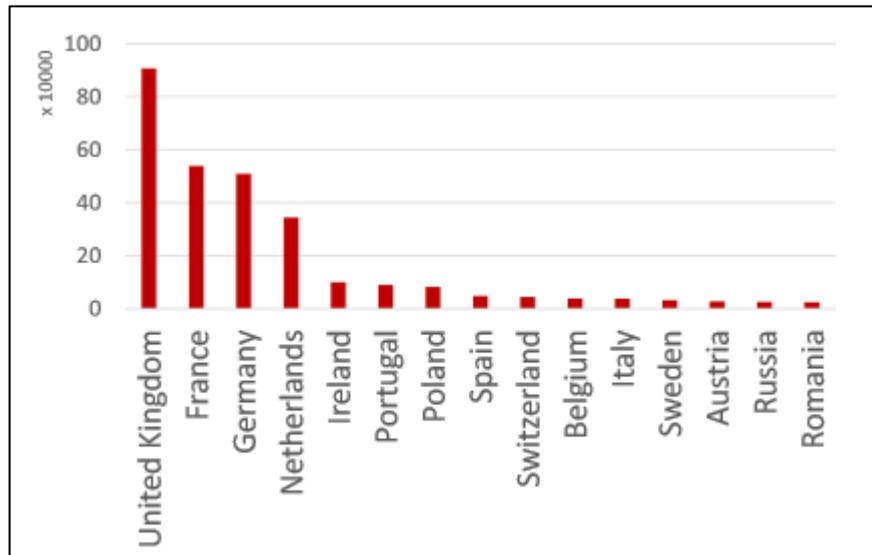
<sup>11</sup> <https://www.xing.com/jobs>

<sup>12</sup> <https://developer.adzuna.com/>

<sup>13</sup> <http://wikifier.org/>

<sup>14</sup> <http://www.geonames.org/>

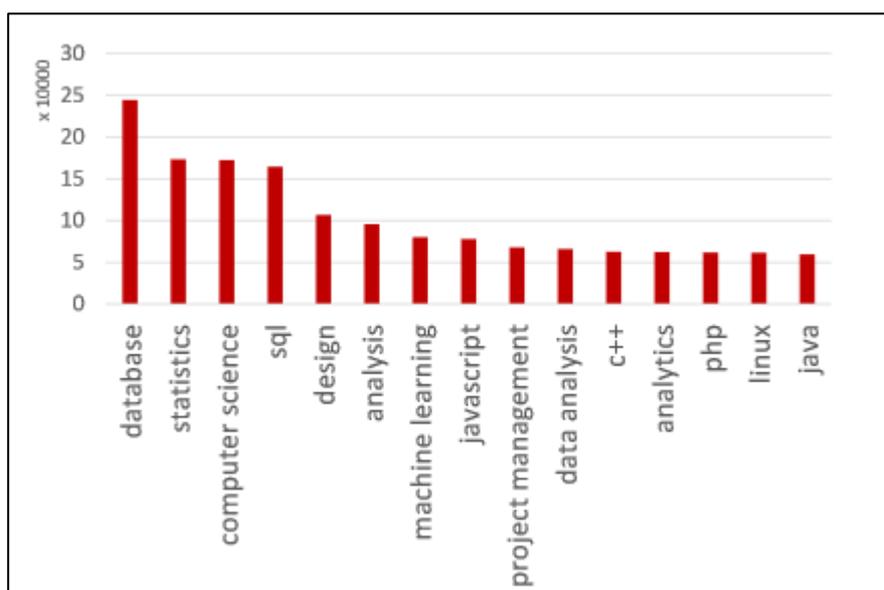
The job postings data set contains almost 3.3M job postings acquired in the period of 18 months. Job postings were located for 144 different countries, the majority of them from Europe. Figure 10 shows the top fifteen countries with most found job postings. The UK dominates other countries with 906k job postings, followed by France with almost 539k.



**Figure 10: Top fifteen countries with most found job postings. The greatest number of job postings were found for UK, followed by France and Germany.**

#### 4.2 Harvesting skills

There were 650 unique Data Science skills extracted from the data set. These include soft skills, such as leadership and management, knowledge of a particular domain, such as machine learning and artificial intelligence, and programming languages. Figure 11 shows the most demanded skills in the data set.



**Figure 11: Top fifteen most demanded skills. They are mostly comprised of high-level skills, such as “database” and “computer science”, and specific programming languages.**



### 4.3 Harvesting courses

For courses we targeted different course providers, including Coursera,<sup>15</sup> providing courses from top universities, and Hackr.io,<sup>16</sup> a service which finds the best online programming courses & tutorials. We also targeted VideoLectures.NET to acquire video lectures containing the Data Science tag. The tags are given manually by the VideoLectures team.

Canvas<sup>17</sup>, edX<sup>18</sup> and Udemy<sup>19</sup> have also been integrated as additional sources for supply data (routine extraction of relevant courses). Table 2 summarizes all supply sources.

**Table 2: Summary of supply sources.**

<i>Demand Data Source</i>	<i>Responsible for Integration in Dashboard</i>
Coursera	JSI
VideoLectures	JSI
Canvas	Fraunhofer
edX	Fraunhofer
Udemy	Fraunhofer

The course data set contains information for over 63k courses, including their title, description and course providers. The data set is comprised of over 8k courses available online and 55k offline courses. Figure 12 shows the distribution of online courses by course providers. The most courses were acquired from Coursera with above 4k, followed by Hackr.io at 2k.

---

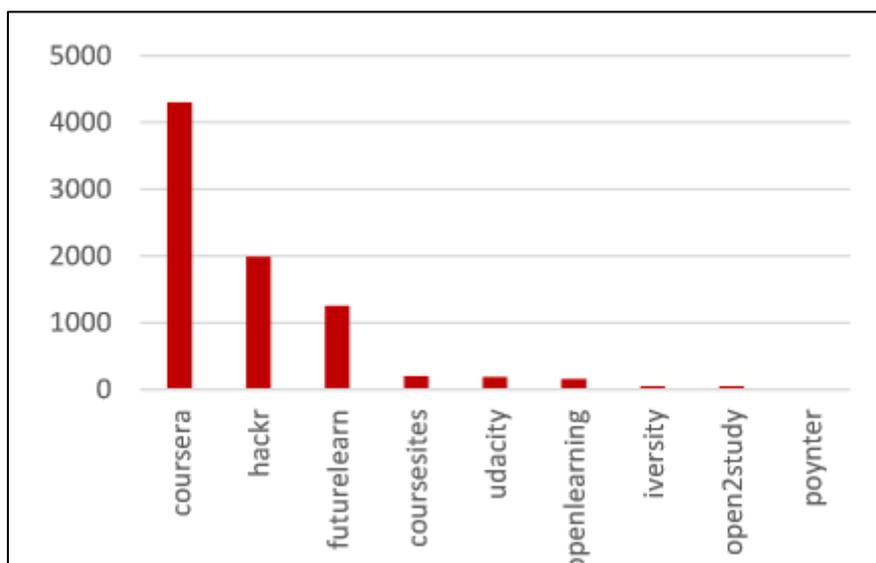
<sup>15</sup> <https://www.coursera.org/>

<sup>16</sup> <https://hackr.io/>

<sup>17</sup> <https://www.canvas.net/>

<sup>18</sup> <https://www.edx.org/>

<sup>19</sup> <https://www.udemy.com/>



**Figure 12: The distribution of online courses by course providers. The most courses were acquired from Coursera, followed by Hackr.io.**

VideoLectures.NET<sup>20</sup> is an award-winning free and open access educational video lectures repository. It contains videos of individual lectures as well as lectures given at renowned conferences.

We acquired a data set of over 20k lectures published on VideoLectures.NET. It contains information about the lectures available on the video repository including title and description and link to the lecture.

## 5. Dashboard evaluation

### 5.1 Evaluation Goals

In order to validate the usability and usefulness of the EDSA Dashboard, we conducted a formative usability evaluation. The evaluation was performed to identify to what the foreseen users can:

1. Easily perform basic tasks such as a search for a job or understand the learning paths.
2. Find the dashboard easy and enjoyable to use.

The following tasks were identified beforehand, and they defined the evaluation exercise. The evaluators were asked to:

#### 1. Search/Identification

- T1. Search for a relevant data science job in Germany
- T2. Search for a job based on specific skills (e.g. Python and Java)
- T3. Identify what steps are missing in one's learning path to become a data scientist

#### 2. Analysis/Exploration

- T4. Analyse which country has more suitable job offers
- T5. Identify which are the top 3 relevant skills for a data scientist

<sup>20</sup> <http://videlectures.net/>



## 5.2 Evaluation Setup

### Number of participants

6 Student Assistants (Data Science area) from Fraunhofer IAIS participated in the evaluation.

### Moderator

1 moderator carried out the evaluation. The moderator was open to discussion with the participants, explaining the objective of the tasks without giving any additional instructions on how to complete a task.

### Dashboard version

The version used during the was the last deployment on 22.11.2017

<http://edsa-project.eu/resources/dashboard/>

### Metrics

We defined the following evaluation metrics:

1. **NASA Task Load Index (TLX)**<sup>21</sup>: In all cases we used the standard NASA Task Load Index (TLX) to measure workload in loosely time-constrained tasks.
2. **Thinking Aloud protocol** [4] to uncover usability issues.
3. To measure the usability, we will use a simplified version of **Post-Study Usability Questionnaire (PSSUQ)** [5].

#### NASA Task Load Index

*Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.*

Name	Task	Date
Mental Demand      How mentally demanding was the task?		
Physical Demand      How physically demanding was the task?		
Temporal Demand      How hurried or rushed was the pace of the task?		
Performance      How successful were you in accomplishing what you were asked to do?		
Effort      How hard did you have to work to accomplish your level of performance?		
Frustration      How insecure, discouraged, irritated, stressed, and annoyed were you?		

**Figure 13: The NASA Task Load Index used during the Dashboard usability evaluation.**

<sup>21</sup> <https://humansystems.arc.nasa.gov/groups/tlx/>

The Post-Study Usability Questionnaire Version 3		Strongly agree							Strongly disagree	NA
		1	2	3	4	5	6	7		
1	Overall, I am satisfied with how easy it is to use this system.	<input type="radio"/>								
2	It was simple to use this system.	<input type="radio"/>								
3	I was able to complete the tasks and scenarios quickly using this system.	<input type="radio"/>								
4	I felt comfortable using this system.	<input type="radio"/>								
5	It was easy to learn to use this system.	<input type="radio"/>								
6	I believe I could become productive quickly using this system.	<input type="radio"/>								
7	The system gave error messages that clearly told me how to fix problems.	<input type="radio"/>								
8	Whenever I made a mistake using the system, I could recover easily and quickly.	<input type="radio"/>								
9	The information (such as online help, on-screen messages and other documentation) provided with this system was clear.	<input type="radio"/>								
10	It was easy to find the information I needed.	<input type="radio"/>								
11	The information was effective in helping me complete the tasks and scenarios.	<input type="radio"/>								
12	The organization of information on the system screens was clear.	<input type="radio"/>								
13	The interface* of this system was pleasant.	<input type="radio"/>								
14	I liked using the interface of this system.	<input type="radio"/>								
15	This system has all the functions and capabilities I expect it to have.	<input type="radio"/>								
16	Overall, I am satisfied with this system.	<input type="radio"/>								

*\*The "interface" includes those items that you use to interact with the system. For example, some components of the interface are the keyboard, the mouse, the microphone, and the screens (including their graphics and language).*

**Figure 14: The Post-Study Usability Questionnaire used during the Dashboard usability evaluation.**

**Procedure**

The moderator carried out the experiment at each participant’s desks, in order to provide a comfortable environment for them. The evaluation introduction also gave some background of the EDSA project, before explaining the tasks requested. Participants were told to speak freely and loud what they are thinking during their attempts to carry out the tasks in the EDSA Dashboard; all the while, the moderator took note of all their remarks. After each task, the NASA TLX questionnaire was given to each participant. At the end of all tasks, the PSSUQ questionnaire was also filled out by the participants.

No time limit was applied but the participants were asked to stop attempting to complete the tasks when they gave up. The time required was recorded, but no timer was shown to the participants so as not to introduce time pressure for the participants while performing the task.

**5.3 Evaluation Results and Discussion**

In general, all the evaluation showed positive scores, indicating that the participants were able to complete the tasks and the workload demand using the Dashboard was low. Additionally, the usability scores showed an overall user satisfaction with the Dashboard.

The most prominent difficulty in the Dashboard’s use was the intuitiveness during first use, but this showed signs of improvement when performing the next tasks. A How-to Video for new users is therefore highly recommended.

Figure 15 shows the results of the NASA TLX. Overall, all the NASA indexes indicate a positive result, Task 3 and 5 were the ones with the lowest level of required effort. In contrast, Task 1 showed the highest level of frustration, although this may be related to the lack of guidance at first use (see above).

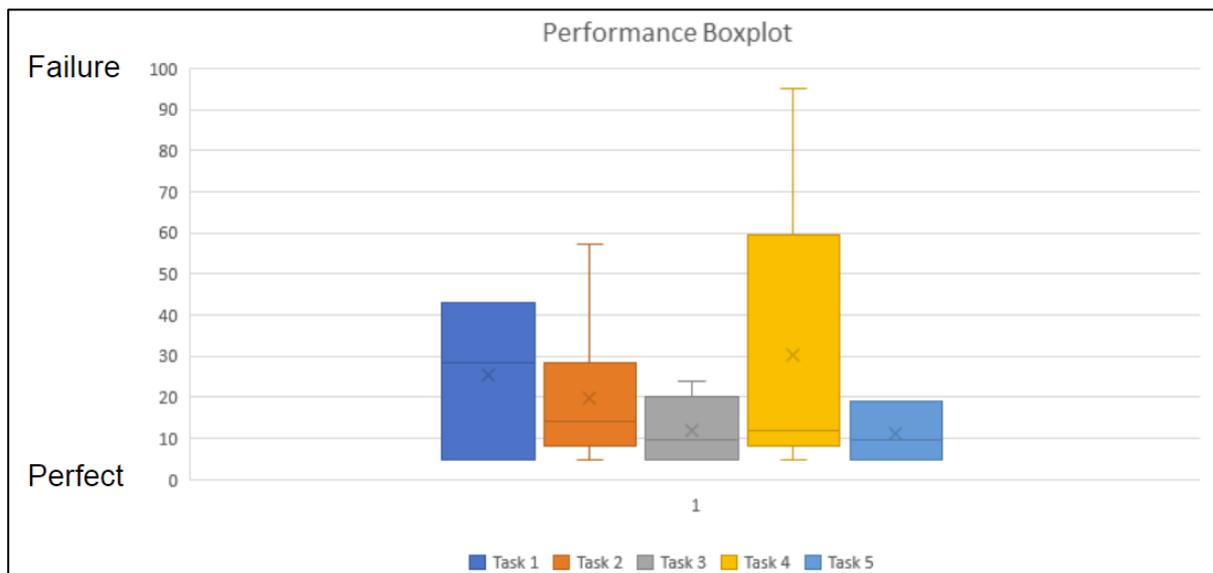


Task 4 was the task that required most effort to accomplish and with the lowest performance of all tasks. In general, the map view proved to be somewhat confusing, as well as the interpretation of its values.

	T1		T2		T3		T4		T5	
	M	STD	M	STD	M	STD	M	STD	M	STD
Mental Demand	15.08	13.30	19.05	13.75	15.08	13.59	20.63	23.27	12.70	5.26
Physical Demand	18.25	17.48	14.29	13.47	14.29	13.47	13.49	13.59	10.32	6.96
Temporal Demand	26.98	20.33	34.13	15.89	20.63	12.80	24.60	16.36	20.63	8.98
Performance	25.40	16.65	19.84	17.26	11.90	7.14	30.16	32.37	11.11	5.94
Effort	27.78	24.19	29.37	19.52	13.49	5.08	30.16	25.74	11.11	6.54
Frustration	29.37	22.24	23.81	17.39	13.49	10.07	19.84	24.03	8.73	3.27

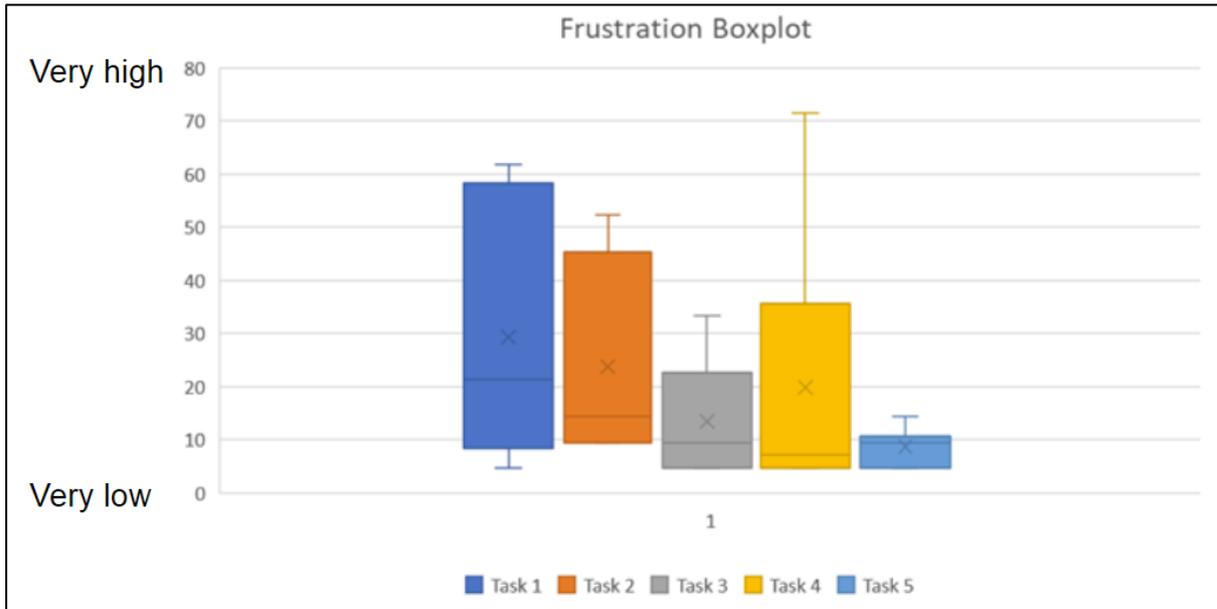
**Figure 15: Workload analytics from Task 1 (T1) - Task 5 (5). Mean (M) and Standard deviation (STD) are calculated. Green tones show positively lowest ratings, and red tones the opposite higher ratings.**

Figure 16 shows the boxplot of the NASA TLX score in terms of performance. The users were asked to rate how successful they felt accomplishing a task, from Perfect to Failure. Overall all the tasks were completed. Task 4 was the only one (1 user) that failed to be completed.



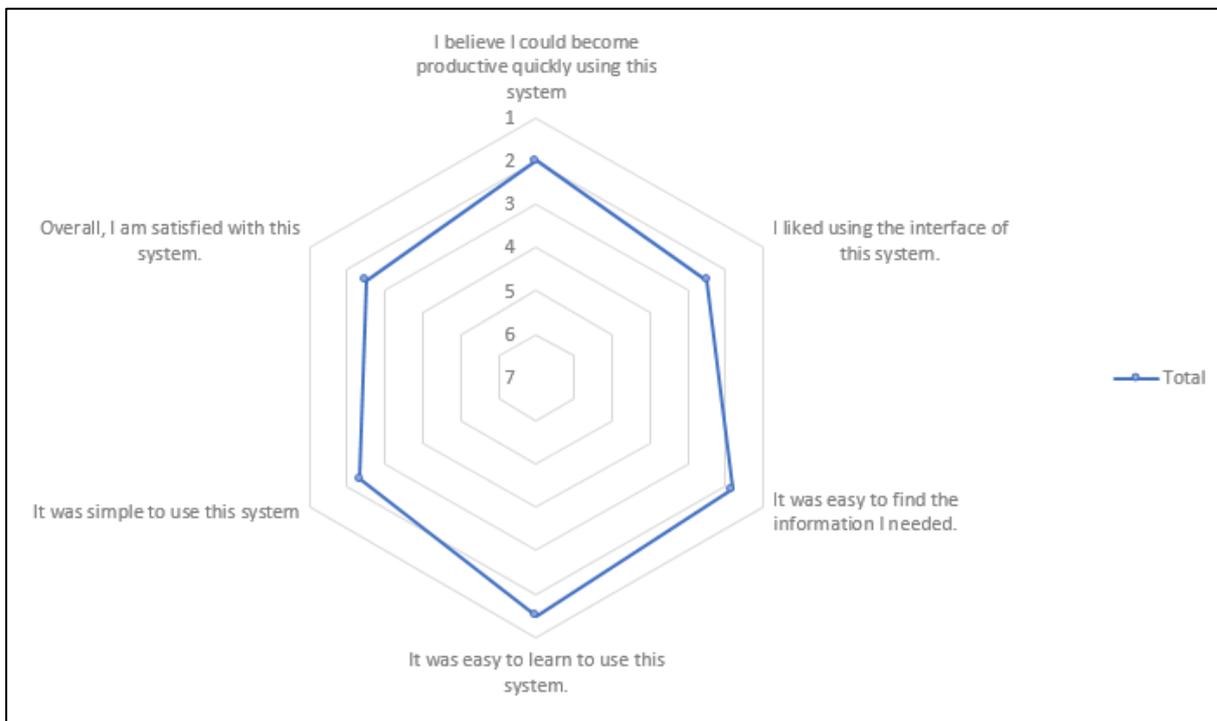
**Figure 16: Boxplot of self-perceived success in accomplishing each task, according to NASA LTX results.**

Figure 17 shows the boxplot of the NASA TLX score in terms of frustration. The users were asked to rate how frustrated they felt while performing the task. Overall the mean of users felt low frustration while using the Dashboard. Task 1, 2 show some exceptions which may be related to the effect of first-use. Task 4 shows a higher degree of frustration related to the user who could not complete the task (as shown in Figure 16).



**Figure 17: Boxplot of how insecure, discouraged, irritated, stressed, and annoyed were the participants during the task execution, according to NASA LTX results.**

Figure 18 shows the results of the PSSUQ questionnaire about the general usability of the Dashboard. All the shown dimensions achieved good scores, showing that the Dashboard is usable and participants found it easy to learn, and were satisfied with it.



**Figure 18: Post-Study Usability Questionnaire (PSSUQ)**

The most relevant Think Aloud remarks gathered during the evaluation process were:



1. Most users had problems locating the search query fields.
2. Most users had positive reaction to Learning Pathways once they located it.
3. Most users were confused with the purpose of the buttons on top of the page.

## 6. Conclusion

Linking the demand for data science skills with the supply of learning resources that offer these skills is crucial for bridging the data science skills gap. Towards this goal, EDSA has developed an interactive dashboard that enables its users to explore both the current data science skills demand and supply.

This deliverable has presented the types of supply analyses conducted in relation to the demand analysis and how this work has been implemented in the EDSA dashboard. In particular, this deliverable has described how the demand and supply data are presented in the EDSA dashboard and how users are able to interact with this data in order to explore the current demand and supply.

The main contributions of this work are: a) creating a sizable data set of data science related job postings containing the job postings title, description, locations and other information, and b) developing a dashboard that offers relevant job postings for a given query, as well as courses and learning pathways for gaining the appropriate skills. The dashboard is updated daily with new job postings and courses and will continue to be maintained beyond the lifetime of the project, thus offering a sustainable service to the data science community.

It should be noted that the dashboard will continue to be maintained and further developed after the end of the project. Further improvements will include the addition of a time series analysis based on the data being collected, which is to be implemented by Fraunhofer. More details about this planned work are provided in the Project Exploitation Report (D5.4). Additionally, usability issues that were identified in the evaluation of the dashboard, such as usability improvements of the map view and the search query fields as well as other layout issues, will be taken upon by the team that will continue to support the dashboard after the project's end.

## 7. References

- [1] B. Fortuna, J. Rupnik, J. Brank, C. Fortuna, V. Jovanoski, M. Karlovcec, B. Kazic, K. Kenda, G. Leban, A. Muhic, et al. qminer: Data analytics platform for processing streams of structured and unstructured data, software engineering for machine learning workshop. In Neural Information Processing Systems, 2014.
- [2] L. Ratinov, D. Roth, D. Downey, and M. Anderson. Local and global algorithms for disambiguation to wikipedia. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, pages 1375–1384. Association for Computational Linguistics, 2011.
- [3] E. Sibarani, S. Scerri, N. Mousavi, and S. Auer. Ontology-based skills demand and trend analysis, July 2016.
- [4] Joseph S. Dumas and Janice C. Redish. 1999. A Practical Guide to Usability Testing (1st ed.). Intellect Books, Exeter, UK.
- [5] J. R. Lewis. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use". In: International Journal of Human-Computer Interaction 7.1 (1995), pp. 57.