Co-funded by the Erasmus+ Programme of the European Union

DA.RE.

Phase 1
New Big Data Initiatives

Intellectual Output 1

Towards a Data-driven Mind-Set

Dissemination Level	:	Public
Author(s)	:	C. Cristalli, M. Gatto, D. Isidori, R. Paci (LOC, AEA srl), E. Merelli, M. Piangerelli, L. Tesei, (UNICAM), J.H. Johnson (OP) J. Barbosa, P. Leitão, F. Piras (IPB) B. Kavšek (UP), C.J. Romero (VSL) M. Amador (MAISIS) J. Borlinić, B. Horvat (ABELIUM) N. Stojanovic (NISSATECH)
File Name	:	Da.Re. Intellectual Output 1

Project Acronym	:	Da.Re.
Project Title	:	Data Science pathways to re-imagine education
Project Reference	:	2016-1-IT02-KA203-024645 - KA2 HEI
Project Duration	:	36 months
Website	:	dare-project.eu

Co-funded by the Erasmus+ Programme of the European Union

This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.
Contract no. 2016-1-IT02-KA203-024645 - KA2 HEI

DA.RE.
DATA SCIENCE PATHWAYS
TO RE-IMAGINE EDUCATION

Revision History

REVISION	DATE	INVOLVED PARTNERS	DESCRIPTION
0.1	15/05/2017	Loccioni (AEA srl)	Index draft
0.2	24/07/2017	UNICAM	Completed part regarding Training Paths in Italy
0.3	28/08/2017	UNICAM	First complete draft
0.4	29/08/2017	UNICAM	Complete document

Co-funded by the Erasmus+ Programme of the European Union

Contract no. 2016-1-IT02-KA203-024645 - KA2 HEI

Executive Summary

The intellectual output O1 is the direct result of Phase 1 activities. It comprises: a report with the detailed market needs analysis in the data science field, methodology guidelines for structuring Data Science modules and learning resources to address those needs and a report containing an inventory of business and training opportunities. These documents are also published on the project website (www.dare-project.eu) and are available to both partner and non-partner HEIs, adult training providers, SMEs, public authorities and other stakeholders at national and EU level to foster the spreading of best practices in Data Science education and plan further cross- sectoral Data-Science based initiatives. In addition, the datasets to be used in Phase 2 are provided as separate files.

This output is delivered at the end of Phase 1 (M1-M12) in order to pave the way for the development of new Data Science modules and resources. The O1 as a whole aims at identifying the state of the art of both education and its relationship with the business in the field of Data Science.

More specifically, O1 allows Da.Re. Consortium to update the state of art in the following fields:

- Data Science competences, including innovative competency models, best practices for HEIs/SMEs/PA cooperation, key areas and innovation trends. This work results in an inventory of several data analytics skills. This can include quantitative skills such as machine learning, statistical modelling and predictive analytics; computer science and IT skills like proficiency in programming languages and SQL/No-SQL database management; as well as qualitative skills (e.g. teamwork and presentation skills);

- Data Science education, including a “knowledge map” and the identification of existing training opportunities, educational resources and teaching methodologies as well as the detection and analysis of the project target groups’ training needs;

- Desk research on Data Analytics models, also exploiting the catalogue of models developed within the 7FP projects or LLP in Europe.

The document also describes the different conclusions made comparing the research carried among the countries involved in the project. Finally, an introduction towards the creation of the Da.Re. Pilot Course is also given, providing the expected learning outcomes and profiles of the students to be engaged.

Partner P2 (UNICAM) coordinated the tasks leading to l.O1 by providing common guidelines, tools and templates for the collection of information.

DA.RE.
DATA SCIENCE PATHWAYS
TO RE-IMAGINE EDUCATION

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

The Intellectual Output O1 is the result of the activities carried on during Phase 1 of the Da.Re. project. It presents the analysis of the state of the art about market needs, training paths and training needs linked to Data Science. The analysis was conducted by the Consortium partners in their respective Countries. This deliverable establishes the background for designing and structuring the learning modules of the pilot course to be realized during Phase 2.

More specifically, the work done for the intellectual output O1 allowed the Da.Re. Consortium to acquire knowledge about the state of art in the following fields:

- Data Science market needs: the analysis of the current market needs in terms of competences and experience required for Data Scientist open job positions.
- Data Science education: the identification of existing training opportunities, educational resources and teaching methodologies in order to detect gaps with respect to the training needs.
- Data Science training needs: the recognition of new competences, best practices for HEIs/SMEs/PA cooperation, key areas for innovation trends;

The results of the related activities were combined to perform a comparative analysis among Italy, UK, Portugal, Slovenia and Serbia in the European context.

The activities of Phase 1 of the project were structured as follows:

Activity 1.1: "Mapping of knowledge and initiatives of Data Science and Big Data Analytics within the selected countries, with a focus on identifying key skills for Data Scientists. This work resulted in an inventory of both business and academic opportunities". The results of this activity are reported in Chapter 2 except the inventory of academic opportunities that is detailed in Chapter 3.

Activity 1.2: "Surveying of existing training paths, educational resources and most applied teaching methodologies. An inventory on training opportunities will provide an overview of what is available in the Consortium countries and at EU level. In this activity, the datasets needed for the teaching and learning activities were identified and created". The results of this activity are reported in Chapter 3.

Activity 1.3: "Training-needs analysis: in this activity, partners submitted questionnaires and held interviews/consultations with target groups during participatory events, to draft a list of training needs to meet the Data Science market requirements". The results of this activity are reported in Chapter 4.

Activity 1.4: “A comparative analysis and matching of the results obtained in the previous phases will be carried out. This will result in the intellectual output O1”. The results of this activity are reported in Chapter 5.

In order to understand better the content of the present document, the reader should consider the following issue. The problem in giving a precise figure to the number of data scientists in the UK depends on many things. First, data scientist is not very well defined and is often used interchangeably with other term such as ‘data analyst’ and even sometimes ‘engineer’. Secondly many people use data science, even though they would not consider themselves to be data scientists. Thirdly, data science includes many specialisms, from the engineers who design and set up distributed multiprocessor systems to the statisticians, mathematicians, physicists and others who devise new scientific techniques to abstract value from heterogeneous data. This is highly relevant to the DA.RE. project.

Chapter 2. Mapping of Knowledge and Initiatives of Data Science and Big Data Analytics

The conducted research during Activity 1.1 took into account the perspective of the job market in terms of skills and experience required for Data Scientists. The objective was to draw a comprehensive and clear picture of what companies are looking for regarding the figure of Data Scientist.

Along with the implementation of the project activities, the Da.Re. Consortium targeted a double advantage from this research:

- knowing the job market and the skills required by a sample of companies in the project Countries;
- identifying the training needs to be filled by the project.

Moreover, the market needs research gave to the Da.Re. Consortium the opportunity of better defining the next trajectories for properly developing the training modules and resources for the pilot course to be realized in Phase 2.

Partners started by identifying the most popular job-seeking websites in their respective Country in order to search for current Data Scientist open-positions, thus understanding what exactly the market required. The effective use of data requires new skills and demands for new professions, usually referred as the Data Scientist. Several 7FP projects or LLP projects like [1] [2] [3] put as a fundamental aspect the analysis of the required sector specific skillsets for data scientists and the development of an adaptable data science curricula to meet industry expectations.

Other sources besides websites were used to reach this objective such as existing national surveys and direct interviews. This approach worked well for Italy, UK and Portugal. However, in the case of Slovenia and Serbia, the proposed approach did not give enough results to perform a significant analysis. Thus, for these two Countries, ad hoc approaches were devised and are reported in the relative sections below. It is worth mentioning that for UK, where the data scientist figure is already widely present, a further analysis of the existing jobs was performed.

More precisely, the research for Italy, UK and Portugal was implemented through the following steps:

- Creation of a suggested template for collecting the data. *The Coordinator - Loccioni - and the Phase Leader - UNICAM - created a suggested template that could be used by partners for collecting data. The aim was to cluster the most relevant information as homogeneously as possible thus facilitating the final comparative analysis and the way of disseminating the outcome*.
- Identification of the sources of information. *The partners analysed which online or offline resources were more appealing to the task objective. Concerning the websites, they*

identified those mostly used by both companies, for job offerings, and by people, surfing for job seeking.

- Collection of the job positions available. *By using common keywords (e.g. Data Scientist, Big Data, Data Analytics and so on) partners found available positions within their Country for Data Scientists, thus understanding the expectations and the demand of the market.*

- Analysis of the job positions selected. *This activity paved the way to a qualitative analysis. Data Scientists must possess competencies and skills in data mining and analytics, information visualisation and communication, as well as in statistics, engineering and computer science, and acquire experiences in the specific domain of their future work and specialisation* [4]*.Therefore, partners decided to organise the data by clustering them into three hubs: soft skills, hard skills and geographical distribution. The first one gathers the transversal skills required such as communicative and relational attitude, problem solving, team working and others. The second one looks for the technical skills needed for the specific job such as SQL, Python, Spark and so on. The last cluster aims at showing how in each Country the demand for data scientist is distributed. Some partners added other aspects based on specificities of their Country.*

The obtained results gave the chance to the Da.Re. Consortium partners to update their knowledge and networks related to the Data Science world. Particularly, it has been crucial for the involved companies for being aware of what kind of organisations are currently looking for Data Scientists within their Country and also within the same business.

The Higher Education Institutions have benefited as well. In fact, they got a clearer idea of the training needs currently most required by industries, thus affecting at decision-making level the skills framework to be provided to future students. This was a missing point in the higher education field and an element of great relevance for the project.

According to these premises the chapter is organised in order to give relevance to the results collected from each partner by dedicating a specific section to each one. To guarantee a homogeneous display of the data, graphs have been considered the main means for showing the outcomes in order to clearly communicate the results.

2.1 Market Needs Analysis in Italy

The analysis wants to provide a glance on the training needs most required by industry in Italy. The methodology used has taken into account the job positions required by companies on the data science and big data fields.

The research started with the selection of the 4 most popular job-seeking websites in Italy, which were identified as the followings:

- Indeed.com [5]

- Infojobs.it [6]

- Monster.it [7]
- Glassdoor.it [8]

Then we looked for the data scientist positions. From the results, the required hard and soft skills were extrapolated as well as the geographical distribution of the demanding companies. The data were organized in graphs in order to better show and communicate the outcomes.

We report that the data sample is about 40 positions. This number suggests that a wider analysis could be carried out in future initiatives.

2.1.1 Soft Skills Analysis

The results, presented in Figure 1, show that communicational and relational skills are the hottest skills on table for companies. The most required skill for Data Scientists is the ability to persuasively communicate technical results to non-experts as well as to be able to detect and solve problems creating new business opportunities. Moreover, attitude to team working is an important soft skill for a data scientist.

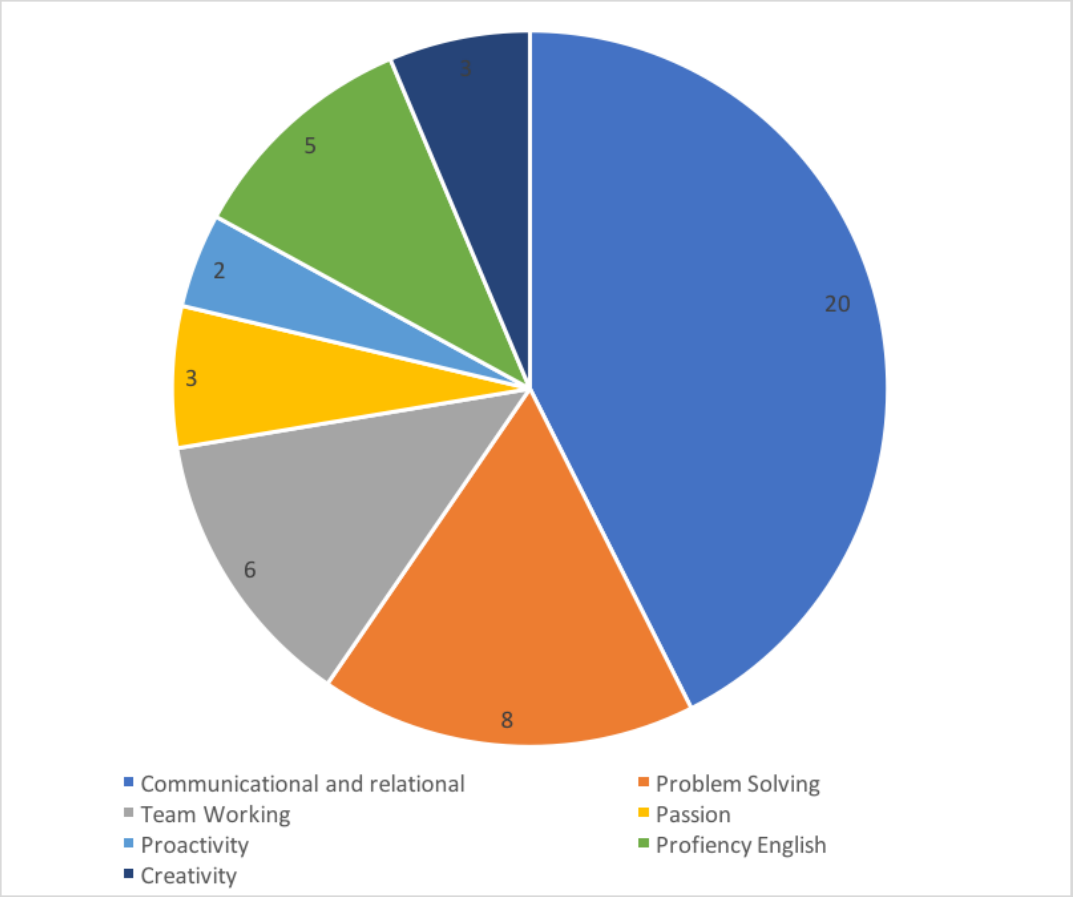


Figure 1 – Soft skill diagram

2.1.2 Hard Skills Analysis

The results, depicted in Figure 2, show that the Italian companies are mostly attracted from people experts in SQL, PYTHON and R tools. This gives an important suggestion on what could be taken as mainstream skills to be developed into training courses and activities.

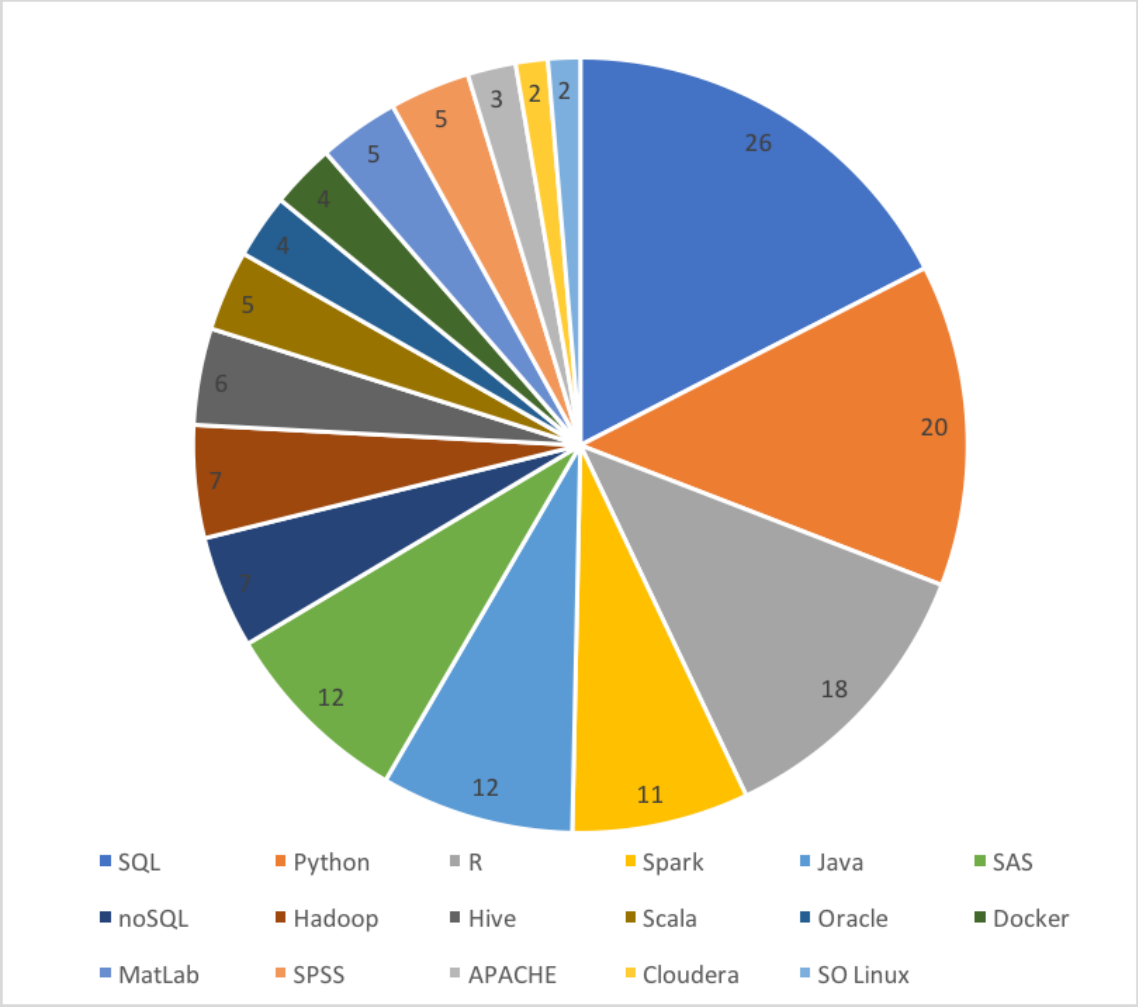


Figure 2 – Hard skills diagram

Given high number of collected hard skills, we identified the categories in Table 1 to provide an easier way of reading the obtained results.

Table 1 - Hard skills hubs

HARD SKILLS	CATEGORY
SQL; ORACLE; noSQL	Database Management System (DBMS)
HADOOP; APACHE; SPARK CLOUDERA; HIVE	Big Data Engine-Data Warehouse-Olap
JAVA; PYTHON; SCALA	Programming Languages
R; MATLAB	Integrated Programming and Analysis Tools
SPSS; SAS	Statistical Analysis Software
DOCKER; LINUX	DevOps - Development and Operations

Hard skills grouped by hubs are shown in Figure 3.

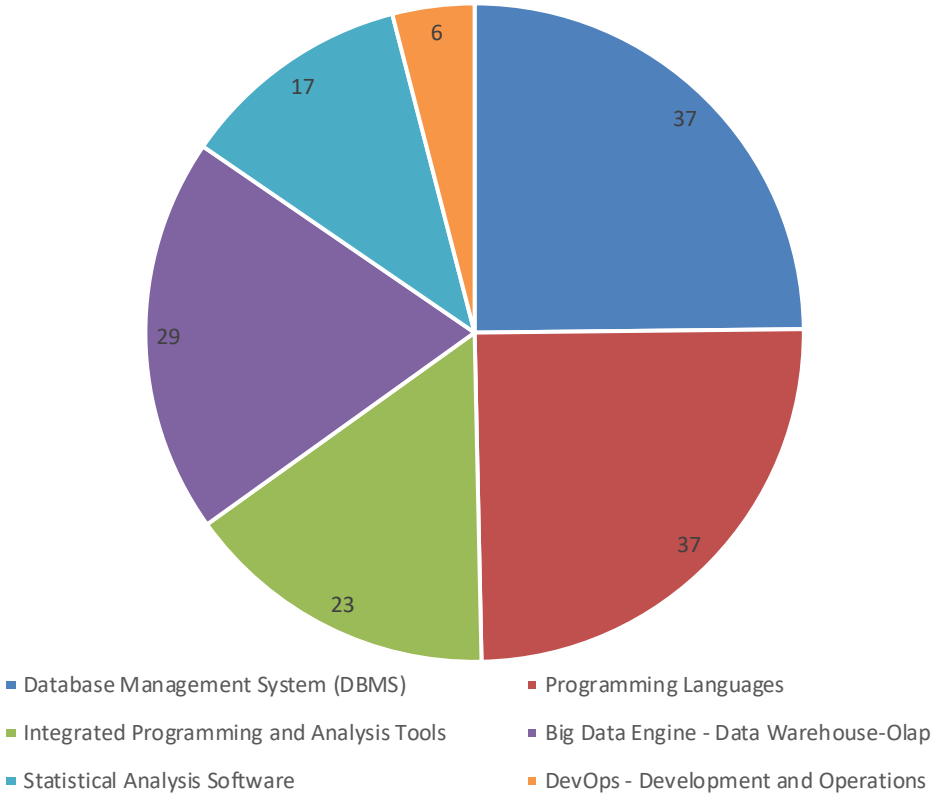


Figure 3 – Hard skills grouped by hubs

2.1.3 Geographical Distribution

Finally, we obtained data about the geographical distribution of the positions. The following graph shows a clear distinction between the northern/central and the southern parts of the Country. In fact, all the opened positions are placed in regions from the northern and central part of Italy. This is a significant datum to be questioned and useful for policy suggestion at national level. Moreover, it can be noticed that the positions are mostly placed in big cities, as shown in Figure 4.

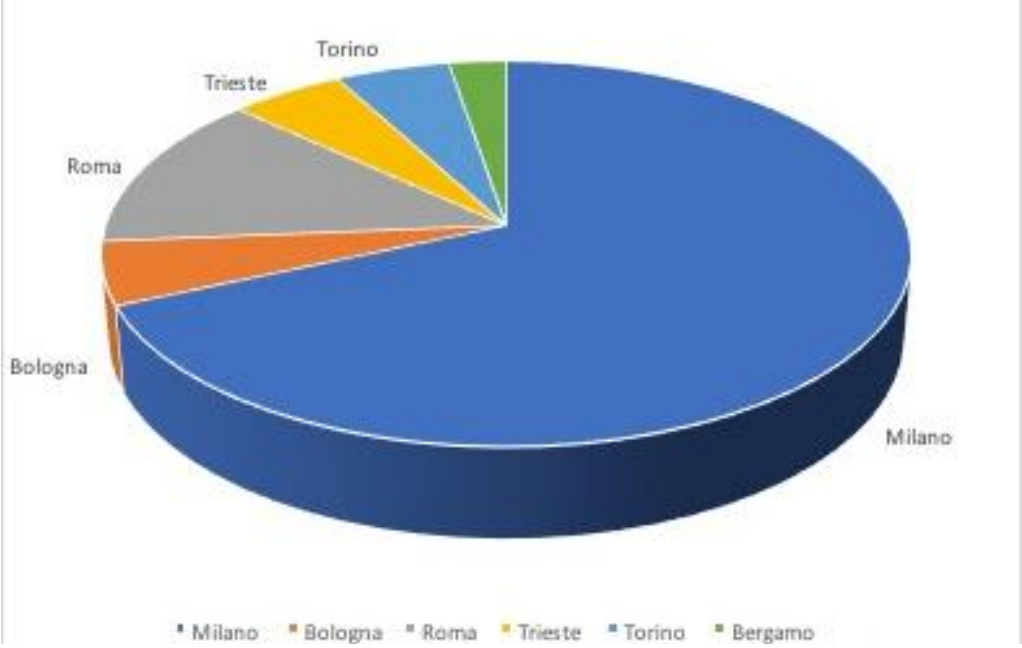


Figure 4 – Geographical distribution

2.1.4 Further Analyses

Type of companies

A further aspect investigated in Italy was the type of company offering job positions for Data Scientists.

The objective of the research was to find a *ratio* behind the peculiarities of those companies, thus identifying both which type of them is more interested in finding a Data Scientist and onto which market sector those profiles are more claimed.

The outcomes, depicted in Figure 5, confirm what could be expected, i.e., the fact that the ICT consulting companies are the most relevant market sector. As a matter of fact, almost half of the companies involved in the research belong to the ICT field. The rest of the companies run at the same percentage, with some prominence of those ones gatherable under the Head Hunter and Assurance fields.

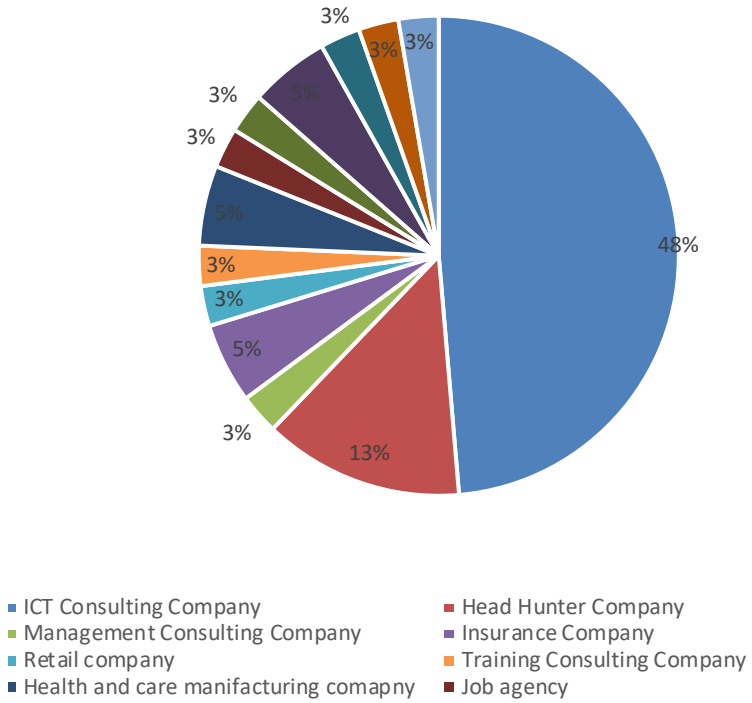


Figure 5 – Type of companies offering jobs in data science

2.2 Market Needs Analysis in UK

This analysis of market needs for data science in the UK uses mostly internet sources. It also uses an interview with the director of a science department and a business manager. The research shows that the UK is the largest employer of data scientists in Europe, and that across Europe the number of data scientists is increasing. There are thousands of data scientist jobs advertised in the UK and this provided the data for a study of the 'soft' and 'hard' skills required in the UK market.

The data suggest that most data science jobs are concentrated in London, but there are significant numbers of data science jobs in the regions. Data scientists are relatively well paid in the UK, earning about twice the national average salary. In Section 3.3 it will be seen that over a hundred UK universities provide Masters degrees in data science suggesting that a significant portion of the market needs for education and training are already met in the UK. For this reason, we have looked beyond the market needs for academic education and considered an industrial role for which there is no formal training available from universities or other sectors. In Section 4.2 we will define a 'data science bridge person' as someone from the business side of an organisation trained to know enough about data science to be able to hold productive conversations with specialists, either within or without their organisations.

2.2.1 How many data scientists are there in the UK

Figure 6 suggests that the UK had about 2100 data scientists in August 2016 and that this will rise to 2900 data scientists in August 2017, ahead of other European countries. This is probably a considerable underestimate, as discussed in Section 2.2 but the increase shown reflects an upward trend.

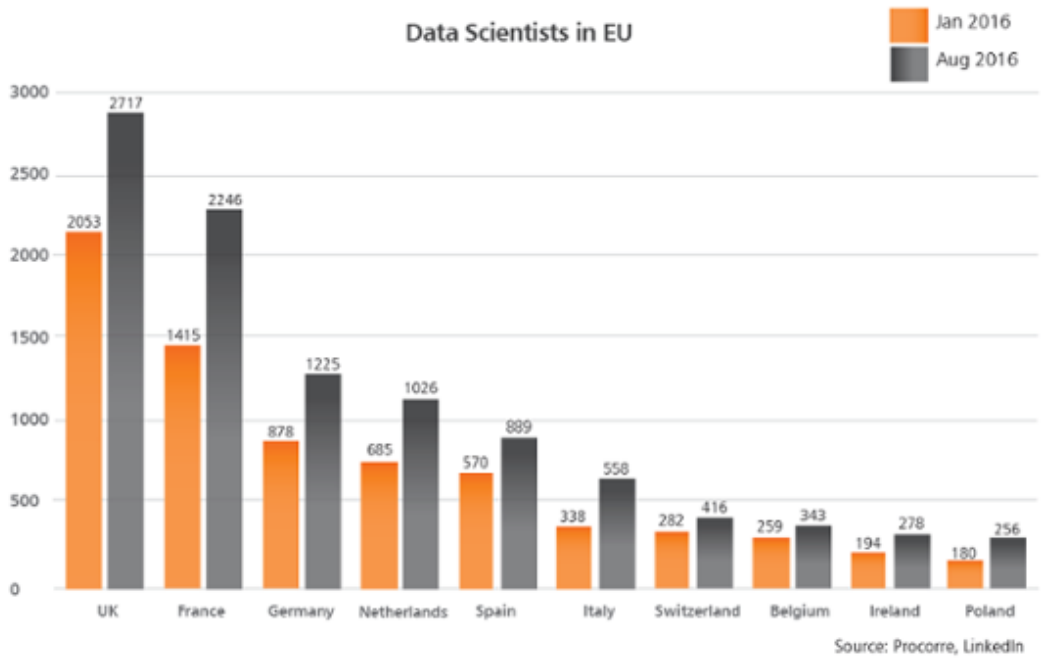


Figure 6 – Numbers of data scientists in European Countries [9]

In the UK, at least 100 universities provide masters in data science. Assuming these courses each attract 10 to 20 students (usually the minimum to make a masters course viable) the universities could be producing 1,000 – 2,000 new data scientists each year. This means that in the UK there could be 6,000 – 10,000 data scientists by 2020.

This upward trend can also be seen in Figure 7. which shows the number of searches on 'data scientist' in the UK over a decade given by Google Trends, where this search term is likely to be used by those seeking jobs in this area.

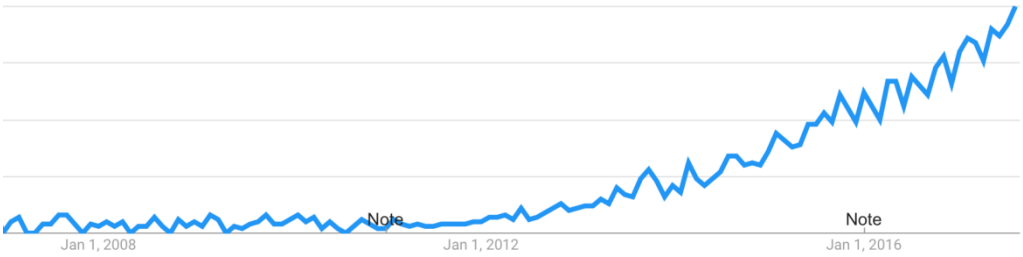


Figure 7 – Increase in the number of UK searches on 'data scientist' over a decade

Later in this report we will introduce the notion of 'bridge person' – someone who is business-based but sufficiently knowledgeable about data science to interface their business to specialist data scientists. By definition, there is the potential for there to be more 'bridge persons' than there are data scientists. Although in the UK there is an ample supply of education for data scientists, there is little or no systematic education aimed at training 'bridge persons'

As Figure 8 shows, a Google search on 'data scientist jobs' yields a large number of hits, including links to sites with thousands of data science jobs.

2.2.2 Soft skills analysis

In their analysis of data science jobs in the UK, Johnson *et al.* [10] made a web search on the term 'data science job' to obtain the results shown in Figure 8. Of the many results a number came from job agencies advertising thousands of jobs, including the glassdoor.com website. Another interesting result was from LinkedIn.com, a social media platform that supports professional networking.

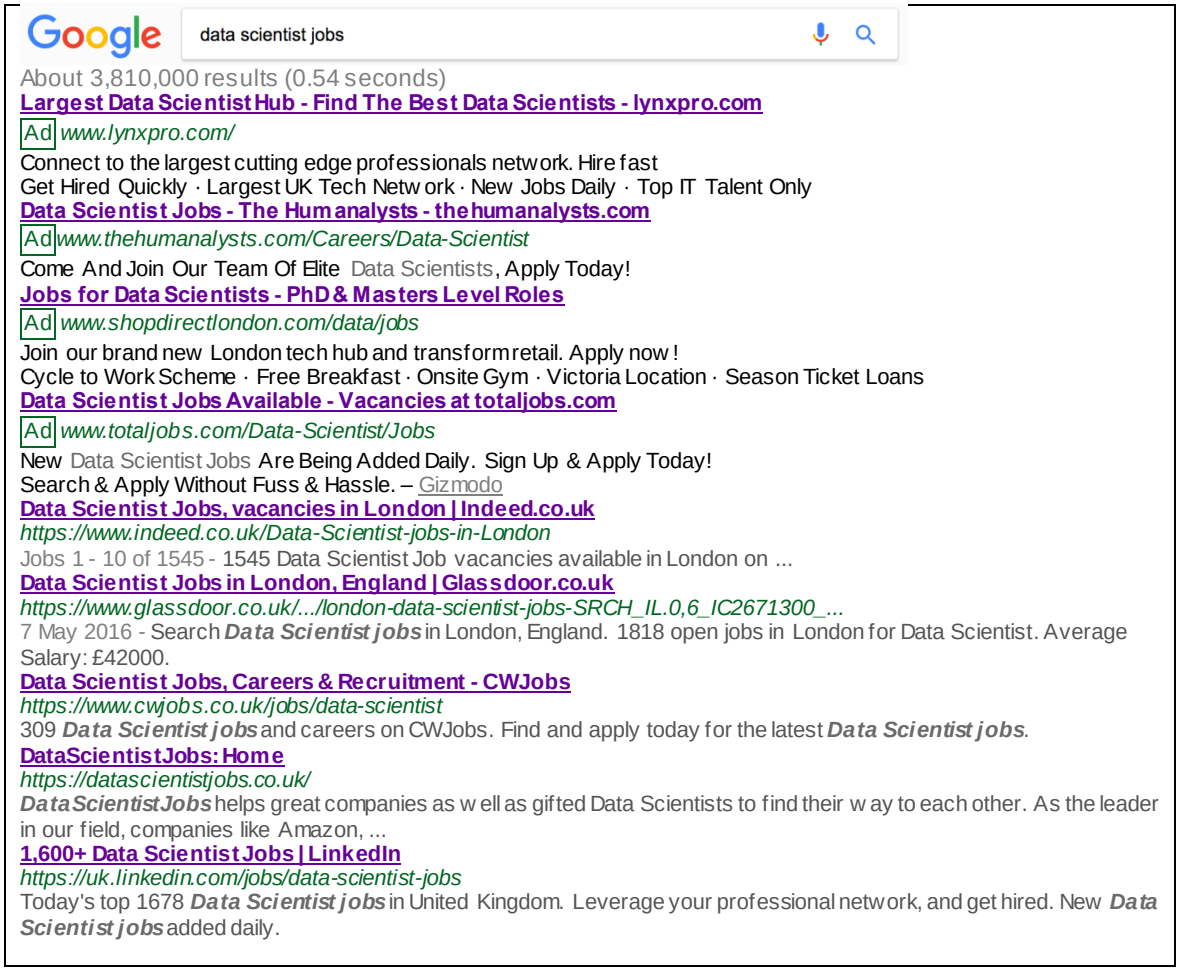


Figure 8 – In the UK thousands of data science jobs are advertised on Google

For the purpose of our analysis, the first ten jobs were taken from the Glassdoor and LinkedIn websites and the details of the jobs were abstracted and recorded. In Figure 9, we reported the first job as an example:

Oakmont Consulting, Greater London

Posted Date Posted 4 days ago Number of views 673 views
Seniority Level - Mid-Senior level
Industry - Computer Software, Information Technology and Services, Information Services
Employment Type - Full-time
Job Functions - Information Technology, Engineering, Finance

Job description: Data Scientist at **Global Investment Management Company**
Location: London
Salary: £Neg+ attractive bonus

Our Client

Our client, a premier global investment fund manager, has experienced unprecedented growth in the last 5 years and now has over $40bn under management. With offices in London, USA, Canada, Hong Kong and Australia, they now employ more than 500 people globally

About the Data Science Team

As part of its strategy for growth, the company is now looking to increase its use of alternative data sources to support the investment analysis processes. They are therefore looking to establish and build an exciting world class Data Science team, to scale and commercialize this use. This technology and data driven team will be responsible for data capture, storage, processing and presentation. The team will work closely with the business unit which supports the investment analysts.

About the role

Our client is looking to recruit an outstanding technical data-scientist to join this new team. Whilst the formation of the team has begun, the technical and design direction is being set experimentally and your role will be to apply your experience and expertise to help steer and influence that direction as data volumes grow larger and processing techniques more sophisticated.

Responsibilities include but are not limited to:

- Contribute their experience, knowledge and expertise to the Data Science Team
- Hands-on delivery with both tactical and strategic development work
- Selecting and testing analytic and visualisation tools
- Integrating with third party data sources
- Processing, cleansing, and verifying the quality and integrity of data for use in analysis
- Selecting features, integrating with existing (or building & optimizing new) classifiers using techniques
- Building supportable solutions for ongoing data collection and dashboard generation
- Work with the business to test complex investment hypotheses
- Working with firm-wide technical leadership to ensure alignment with firm technical direction and other teams.
- Solid C#, Python or other appropriate programming language
- Demonstrable analytic and problem solving skills
- Ability to coach other team members
- Energetic and driven

Desirable Skills

- Designing and implementing solutions using Cloud services from AWS or Azure; for example, -RedShift, Data Lake Analytics, Lambda, u-SQL
- Experience with common data science toolkits, such as R
- Good understanding of machine learning techniques and algorithms, such as Decision Trees or Neural Networks
- Experience with a Hadoop ecosystem
- Previous financial experience is not a mandatory requirement for this role

What they can do for you

- Exposure to a variety of cutting edge technologies
- Culture which champions, recognises & rewards creativity & contribution at all levels
- Amazing corporate offices

- Regular access to external training courses for personal & career development
- Free breakfasts -On-site gym facility -Various social clubs & events

Apply immediately for this outstanding opportunity

Seniority Level Mid-Senior level
Industry Computer Software, Information Technology and Services, Information Services
Employment Type Full-time
Job Functions: Information Technology; Engineering; Finance

Figure 9 – An example of an online job advertisement

As can be seen there is great detail about this job enabling us to analyse it both in terms of soft and hard skills.

The soft skill classification used by the Italian and Slovenian Analysis teams were respectively:

Table 2 – The classification scheme for Italy

Communication and relational	Problem Solving
Team Working	Passion
Proactivity	Proficiency English
Creativity	

Table 3 – The classification scheme for Slovenia

Ask the correct questions, define the problem	Good communication skills
The ability to solve the problem	Strategic thinking and planning
Analytical skills	The ability to make decisions
Passion to learn new things	Follow rules and regulations
Curiosity for observe things (patterns, relations, relationships)	Be moral and ethical
Passion for data analysis	The ability to tell stories using data
Patience and persistence	Project management
Creative thinking	Leadership skills – motivate and inspire others
Courage and self-esteem to stand by your findings, decision	A feel for aesthetics and other visualization skills
Pay attention to quality	Business oriented
Good teamwork skills	

Bearing in mind these classifications we abstracted the following soft skill classification scheme for the UK jobs

Table 4 – The soft skills identified by Vision Scientific and the Open University in the UK

team working	work in fast-paced changing environment
passion/proactive/self-starter	attention to detail/quality work
analytic & problem solving	developing your skills
work with business/customers	leadership & mentoring
curious/hacker/open/ independent researcher	creative/entrepreneurial
communicate/write reports/graphics	

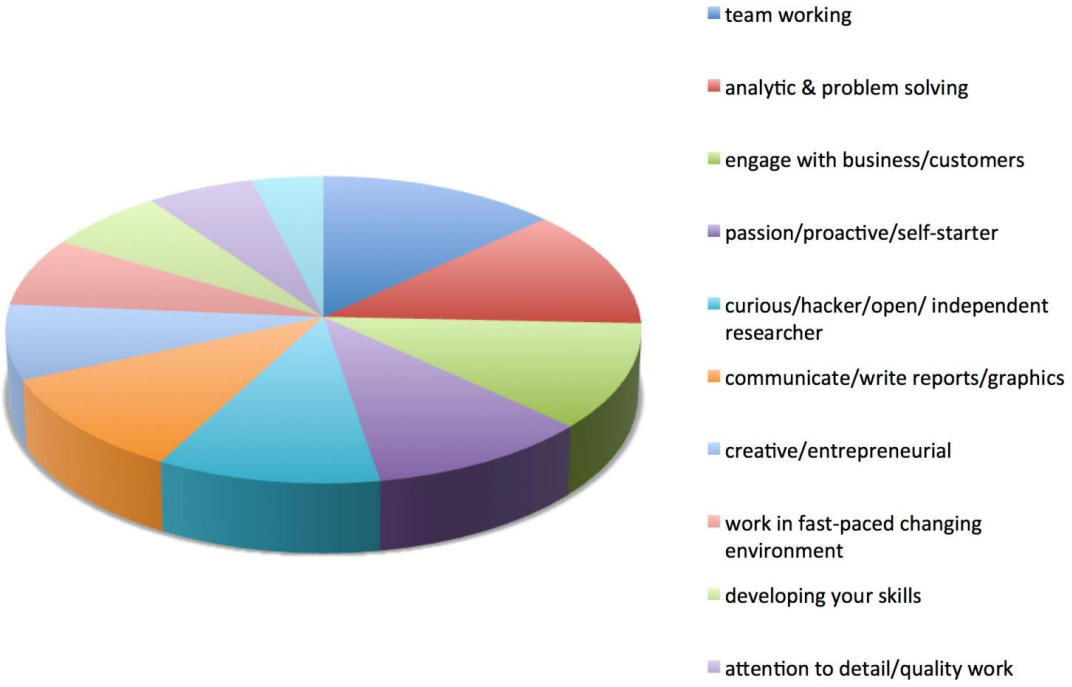


Figure 10 – Soft Skills Analysis for the UK

The soft skills required in the UK, shown in Figure 10, give an insight into the nature of the data science industry. This is a young industry with rapidly evolving technologies that require people to move fast and keep up with developments. This is an industry for intelligent people with analytic and problem solving skills, with passion for what they are doing and the ability to do it without being told. Team working is essential to all the jobs advertised. Interestingly the requirement to engage with the business and clients was required in 85% of the advertisements, related to the need to have excellent communications skills both verbal and written. Apart from being highly technical, data science is 'people' industry where communications within teams and with clients is very important.

2.2.3 Hard skills analysis

The hard skills analysis used the same advertisements as the soft skills analysis, abstracting the technical requirements. In this case there was no attempt at classification and 70 skills were abstracted with 40 occurring in two or more advertisements. We reported the hard skills in Table 5 and in Figure 11 and Figure 12.

Table 5 – Hard skills abstracted from the advertisements

Hard Skill	Number of adverts		Hard Skill	Number of adverts
Python	19		Amazon Redshift	1
Data Science Experience	15		Data Lake	1
R	15		Decision Trees	1
Machine Learning	13		Docker	1
Technical Degree	13		Elastic search	1

SQL	10		EMR	1
Spark	9		Flink	1
Scala	7		Full stack development	1
Scikit	6		Gateway	1
Java	5		Git	1
Predictive models	5		Google BigQuery	1
Visualisation	5		GraphX	1
AI	4		Hbase	1
Hadoop	4		Hive/HiveQL	1
Natural Language Processing	4		Julia	1
Pandas	4		Kera	1
Statistics	4		Lambda	1
Azure	3		Linux Bash	1
Javascript	3		Lucene	1
Pearl	3		ML/MLib	1
NumPy	3		NLTK	1
Visualisation	3		OpenShift	1
APIs	2		PHP	1
AWS	2		Pig	1
C/C++/C#	2		Qlik	1
Cassandra	2		RDF	1
Cloud	2		REST	1
D3	2		SDLC	1
Highcharts	2		Tableau	1
Jupyter	2		Zeppelin	1
Kafka	2			
Kinetica	2			
Mathlab	2			
Mathplotlib	2			
MongoDB	2			
Neural Networks	2			
Postgress	2			
Redshift	2			
Ruby	2			
Sqoop	2			
TensorFlow	2			

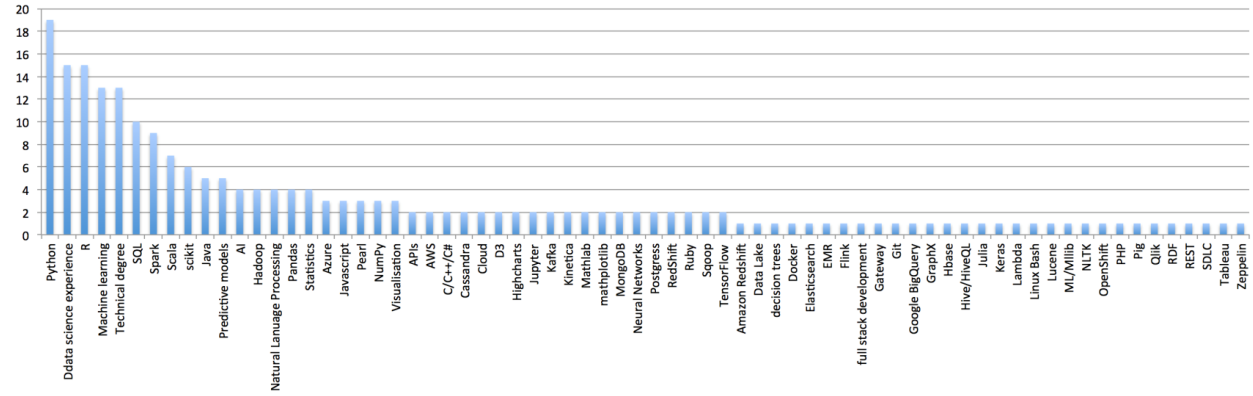


Figure 11 – Hard skills in the UK advertisements

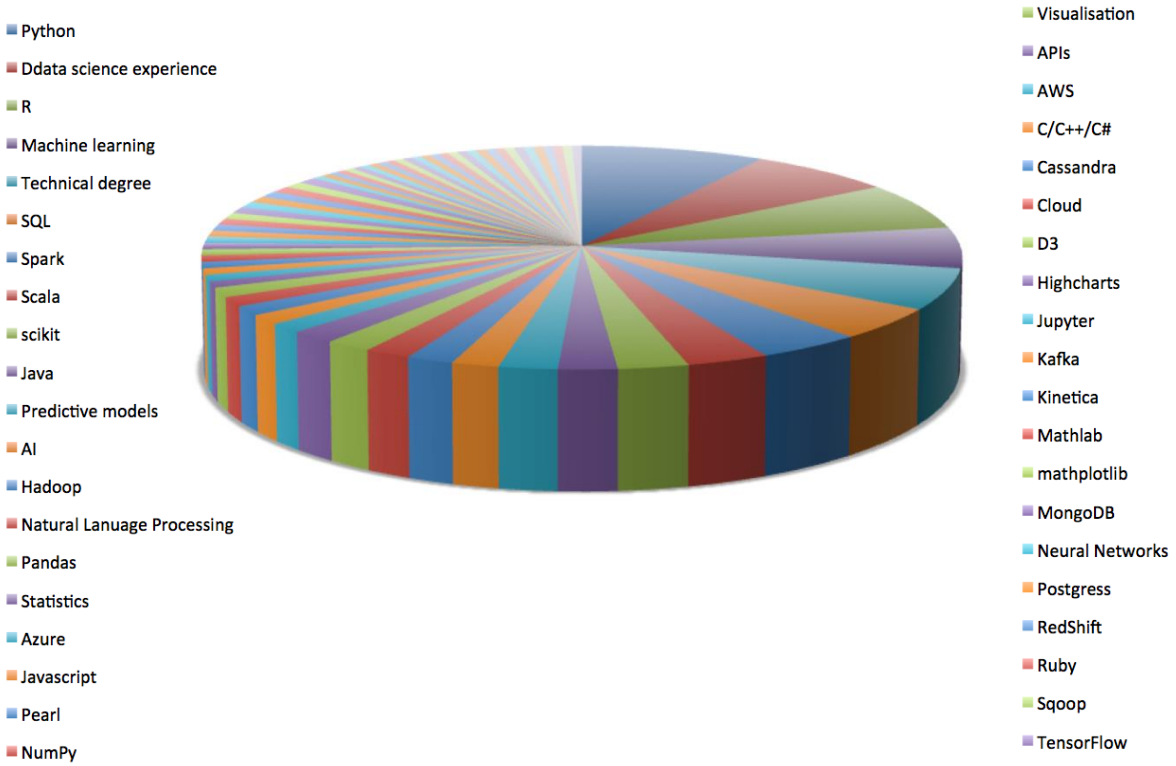


Figure 12 – UK Hard skills, key shows those appearing twice or more

Although our figures are based on a detailed analysis of a limited sample of jobs, the conclusions are similar to other studies. For example, the recruitment company adzuna.co.uk published a list of job skills, reported in Figure 13, based on the data science jobs that it advertises [11].

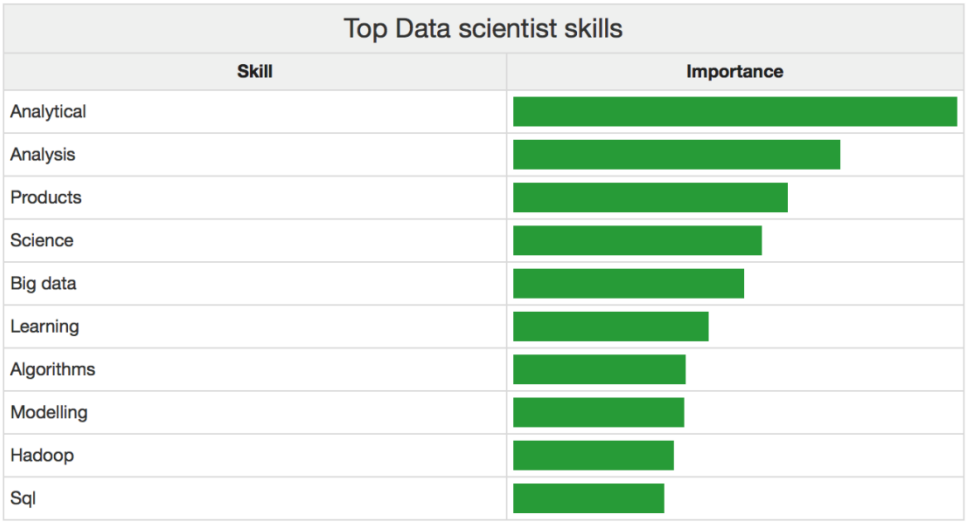


Figure 13 – Top data scientist job skills from [11]

2.2.4 Geographical distribution

As noted earlier, the number of data scientists in the UK in 2017 has been estimated to be about 3,000. The website https://www.adzuna.co.uk/jobs/salaries/data-scientist gives the map shown in Figure 14 showing the distribution of data science jobs across England and Northern Ireland. Those shown add up to 1442 jobs, which is probably an underestimate.

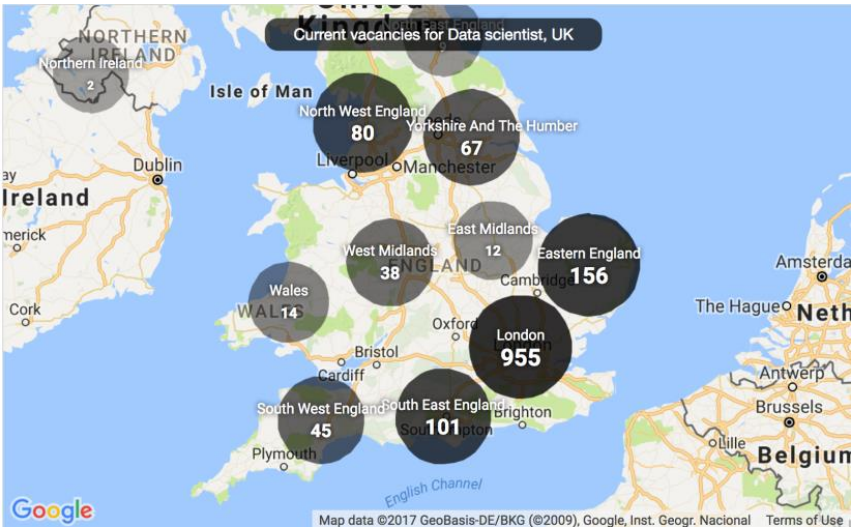


Figure 14 – The geographical distribution of jobs in the UK from [11]

For example, it is hardly credible that here are just two data science jobs in Northern Ireland – one would expect the devolved government to have more than this. Also this map does not include Scotland.

A simple web search on “Data Science jobs Edinburgh” shows there are many data science jobs in Scotland, as illustrated by Figure 15. Indeed a search on the indeed.co.uk website on “Data

Science jobs Edinburgh" gives 3 big data job adverts on the first page – more than are shown in Figure 15.

What then is the number of data scientists in the UK?

A report [12] produced by the Centre for Economics and Business Research (Cebr) states "we expect that the benefits of big data outlined above could lead to the creation of approximately 157,000 new jobs by 2017 and a further 41,000 new jobs by 2020." This two orders of magnitude greater than the 3,000 data science jobs figure we have found so far, but not all those jobs are as data scientists.

Data Scientist. RBS. Royal Bank of Scotland. Edinburgh

Leading the broader Data Scientist & Data Analyst community to identify and deliver opportunities to support the bank's strategic direction through better use…

Data Scientist. NERC. Natural Environmental Research Council. Edinburgh

£28,200 - £30,600 a year.

Utilise data (e.g. An MSc or PhD including Data Science or Data Analytics techniques. Can demonstrate a clear understanding of data privacy and ethical…

Data Scientist. British Geological Survey, Edinburgh

£28,200 - £30,600 a year

A vacancy has arisen for an enthusiastic Data Scientist in our Edinburgh office. You will play a leading role in developing and delivering insight from data…

Data Scientist, Eden Scott, Edinburgh

£30,000 - £50,000 a year

Eden Scott's client, a leading technology start-up based in the centre of Edinburgh are looking for a Data Scientist to join their growing team….

Data Scientist. AIR Worldwide. Edinburgh

Update and improve technical documents describing data processes for sourcing and updating data. The role would initially focus on extending and enhancing the…

Data Scientist. 7 Fifty Two. Edinburgh

7 Fifty Two are currently recruiting for a Data Scientist for our financial services client based in Edinburgh. Hadoop, Hive, Ambari, Pig, Big Data….

Data Scientists. Bright Purple Resourcing. Edinburgh

Our award winning client is looking to appoint several Data Scientists to join their team in the heart of Edinburgh….

Data Scientist. Churchill Frank. Edinburgh

£55,000 - £60,000 a year

A Data Scientist, you will join an established but growing team, high-visibility machine learning team that is developing and deploying solutions to some of the…

Data Scientist SoulTek. Edinburgh

SoulTek - Edinburgh

With the backing of giants like Lloyd's Bank our client is working on software that uses project data to support decision making….

Source: https://www.indeed.co.uk/Data-Scientist-jobs-in-Edinburgh

Figure 15 – 2 pages from indeed.co.uk show there are many data science jobs in Scotland

2.2.5 Further Analyses

Salaries for Data Scientists in the UK

The website https://www.adzuna.co.uk/jobs/salaries/data-scientist gives live statistics for data scientist vacancies from its database of over 1 million job ads. From this we abstracted the following facts (15-8-17):

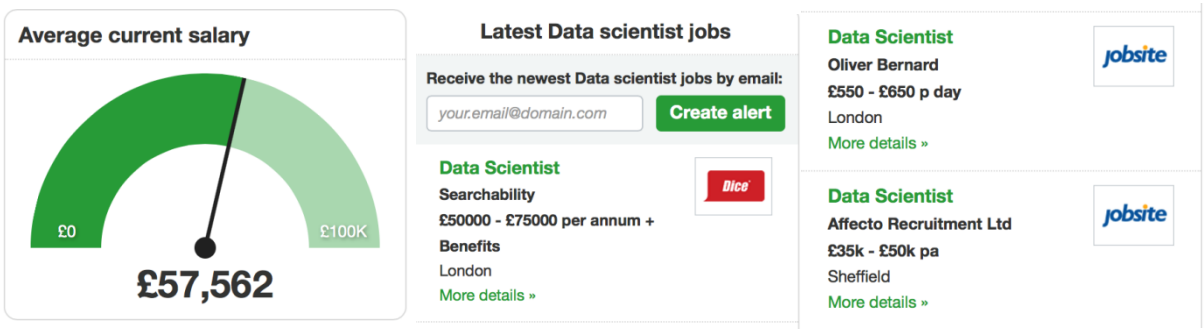


Figure 16 – The average data science salary in the UK is £57, 562

- The average Data scientist salary, as shown in Figure 16, is £57,562. This is 78.7% above the national average advertised salary of £32,207.
- The top companies hiring for Data scientist roles are Amazon, ASOS and Expedia.
- Year-on-year pay for Data scientist jobs has gone up 1.4% year-on-year, compared to an annual change of -1.9% for all jobs.
- In the last 24 hours 62 new Data scientist job ads were added
- Meanwhile, the number of Data scientist job ads is 65.6% higher than last year, with 1,788 current vacancies.
- Most Data scientist job ads are for IT Jobs and Scientific & QA Jobs.

Data science in large and small businesses and organisations in the UK

The pattern of job searches show that there is a strong demand for data scientists across a wide range of industries. This suggests that many large organisations (*e.g.* banks and financial organisations, telecom companies, supermarket and retail store chains, travel companies, airlines, advertising agencies, tech companies, etc.) have strategic knowledge of the value of data science and already have data science groups or departments with high levels of technical knowledge and skills embedded in the business.

However, 95% of businesses in the UK are Small and Medium Sized Enterprises (SMEs – see Table 6) where micro and small enterprises are unlikely to have in-house data science specialists (the exception being companies set up to give data science consultancy). In 2016 in the UK there

were 5.5 million businesses with 99% being SMEs employing 0-249 people, and 5.3 (96%) million micro-business employing 0-9 people. In 2015 there were 383,000 business births and 252,000 business deaths [13].

Table 6 – Definitions of Small and Medium Sized Enterprises (SMEs) [14]

Company category	Staff headcount	Turnover	or	Balance sheet total
Medium-sized	< 250	≤ € 50 m		≤ € 43 m
Small	< 50	≤ € 10 m		≤ € 10 m
Micro	< 10	≤ € 2 m		≤ € 2 m

The European Commission's SME Performance Review estimates the Gross Value Added of SMEs as €473 billion or 49.8% of the UK economy [15]. This means that, although 99% of UK business are SMEs, up to half of the UK economy is generated by large organisations.

2.3 Market Needs Analysis in Portugal

The analysis provides a glance on the training needs most required by industry in Portugal. The methodology used has considered the job positions currently offered by companies on data science and big data fields.

The research started with the selection of the most popular job-seeking websites in Portugal:

- itjobs.pt
- indeed.pt
- sapo.emprego.pt
- net-empregos.com
- empregosonline.pt
- expressoemprego.pt

We then searched for the data scientist open positions. Required hard and soft skills as well as the geographical distribution of the demanding companies were identified. The data were organized in graphs in order to better show and communicate the outcomes.

We report that the data sample is about 44 positions from January 2017 to May 2017. This number suggests that a wider analysis could be carried out in future initiatives.

The discussion carried on within the Portuguese team highlighted the relevance of some results that need to be shared:

2.3.1 Soft Skills Analysis

The results, reported in Figure 17, show that, like already seen in other countries, communicational and relational skills are the most valued skills when companies are looking for Big Data Data Scientists. The ability to work in a team is also considerably important (20%), since the Data

Scientist must be able to work not with just one team but to interact with several teams from many areas of the organization.

As a curiosity, some companies require the soft skill ‘Discretion’, since some data that the Data Scientist will have access may be sensitive.

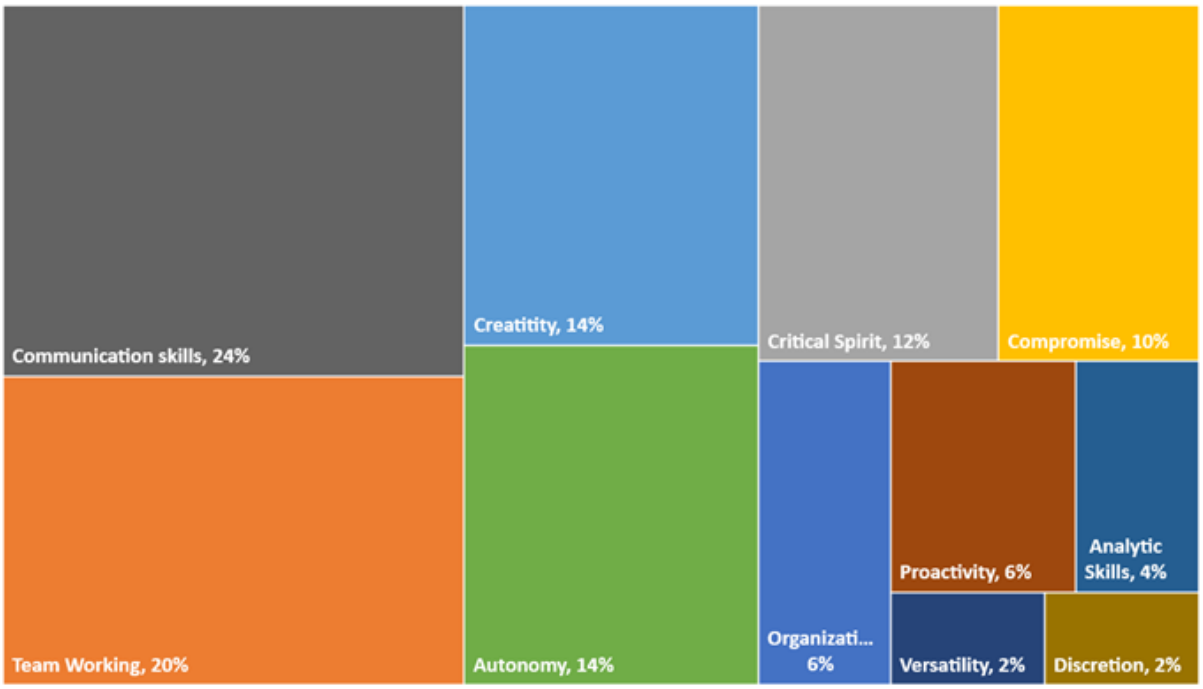


Figure 17 – Soft skills chart

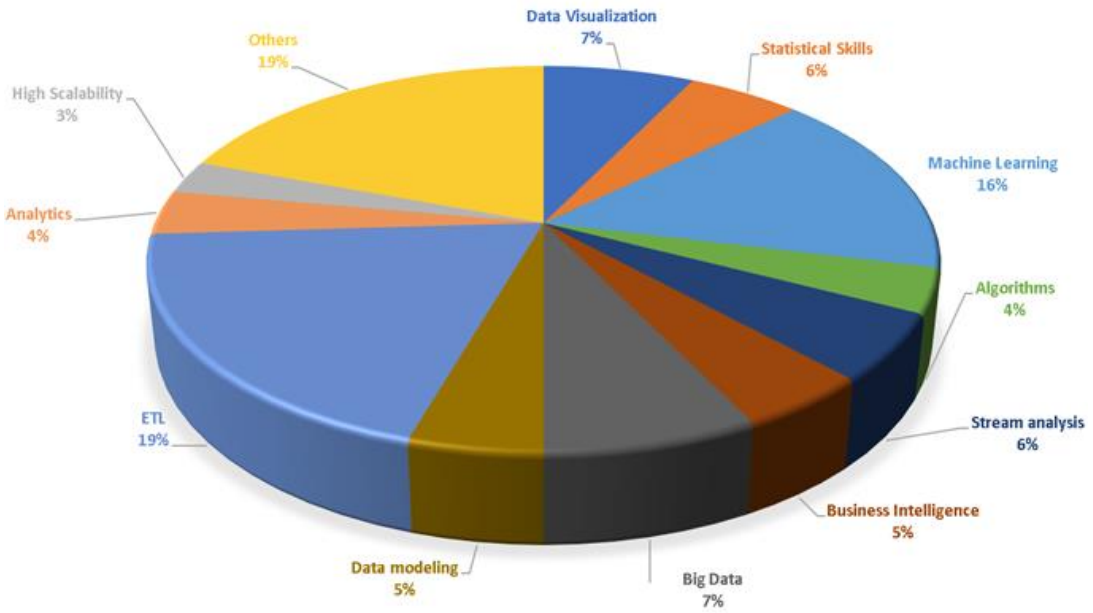


Figure 18 – Competencies diagram

Figure 18 shows the technical needs that are being required by the companies to the candidates. The most required technical competences for the proposed job offers are machine learning (16%), ETL (19%), data visualization (7%) and big data (7%). These technical competences must be considered in the development of training courses.

2.3.2 Hard skills analysis

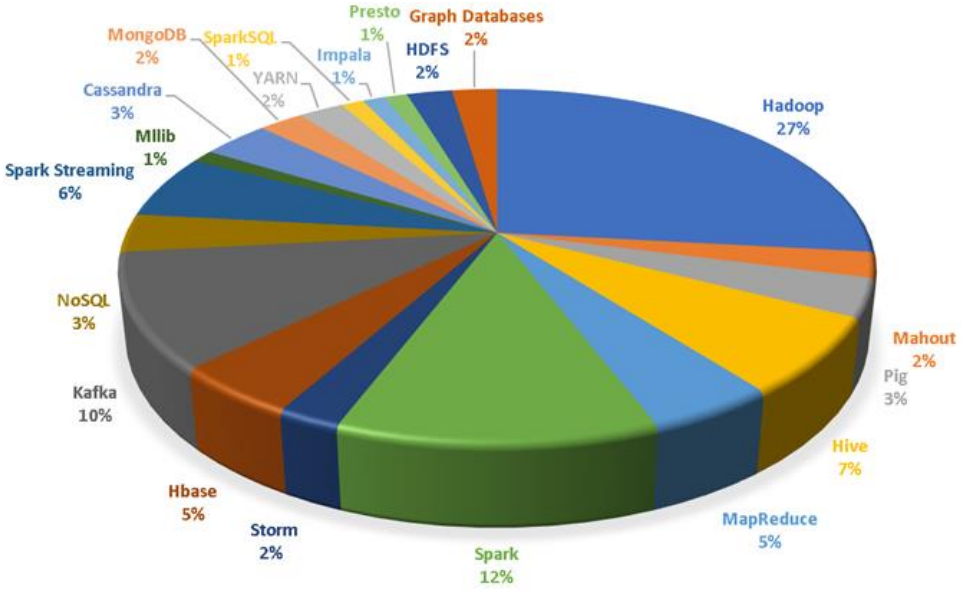


Figure 19 – Hard skills: technologies

Figure 19 presents the results of the required software technologies competences by the companies. Technologies as Hadoop (27%), Spark (12%) and Kafka (10%) have been the most required by the employers. Being a good suggestion in the development of training courses and activities.

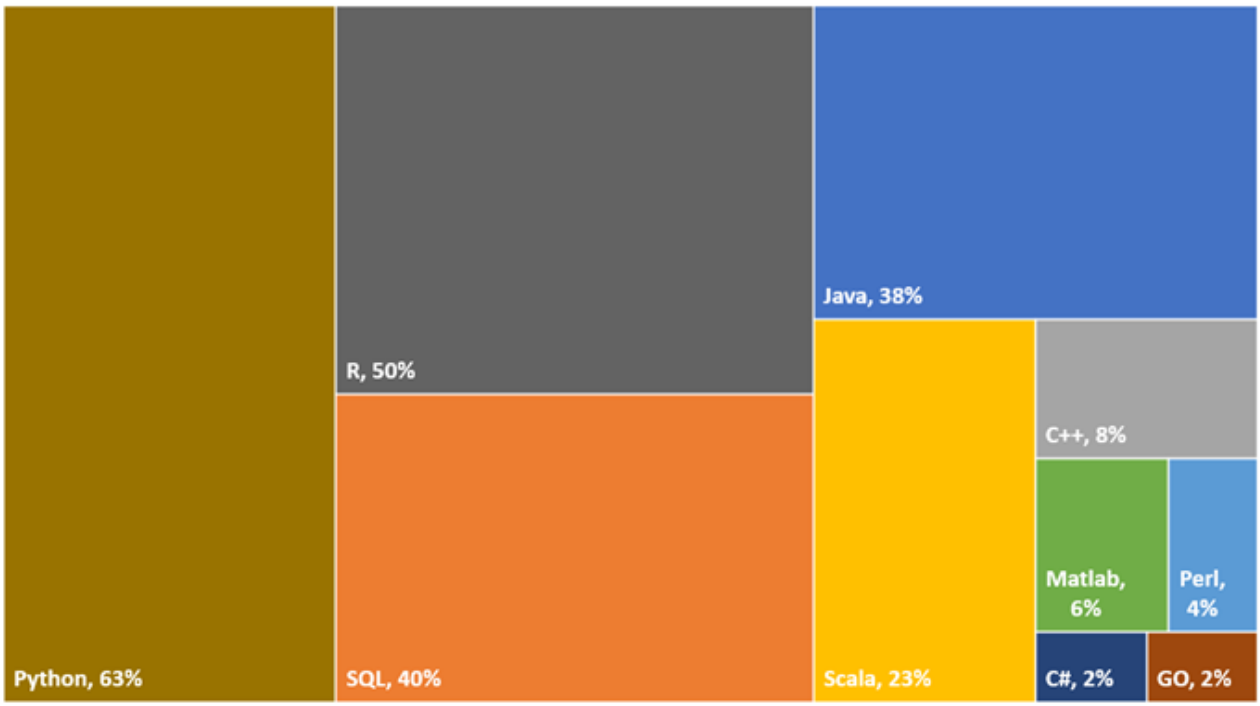


Figure 20 – Hard skills languages chart

The results in Figure 20 show the needs of the Portuguese companies for experts with skills in Python (63%), R (50%), SQL (40%) and JAVA (38%) among others. This data should be taken in consideration not only for the development of training courses but also by the candidates.

Hard skills hubs

Table 7 shows the categories created by MAISIS and IPB researchers to provide an easier way of reading the obtained results. They are shown in Figure 21.

Table 7 – Hard skills hubs

HARD SKILLS	CATEGORY
Java; Python; Scala; C++; C#	Programming Languages
R; MATLAB	Integrated Programming and Analysis Tools
Azure, Google Cloud Platform	Cloud Computing
SQL, NoSQL	Database Management System

Hadoop, MapReduce	Big Data
Kafka , Spark, Storm	Distributed Stream Processing
GO, Docker, Kubernetes	DevOps – Development and Operations

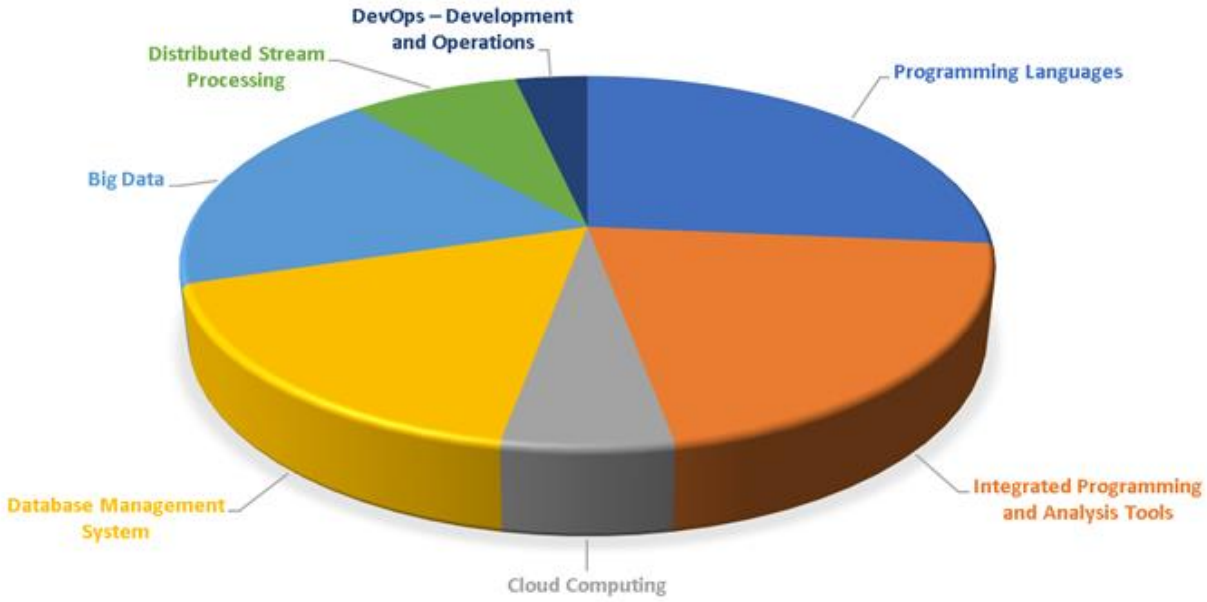


Figure 21 – Hard skills grouped by hubs

2.3.3 Geographical Distribution

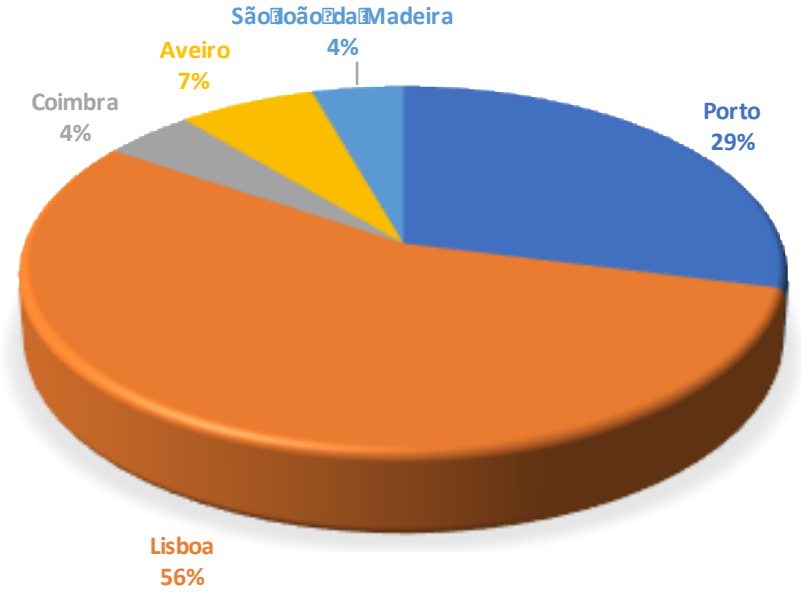


Figure 22 – Job offer geographical distribution

The job offers location underline a clear distinction between the main cities, as Lisbon and Porto, and the others, As shown in Figure 22, most of the opened job offers are located in Lisbon (56%) and Porto (29%). This is a significant datum to be questioned and that can be useful for policy suggestion at national level.

2.4 Market Needs Analysis in Slovenia

The analysis wants to provide a glance on the training needs for data scientists most required by the industry in Slovenia. The standard methodology takes into account the job positions currently required by companies in the data science and big data fields. However, this unfortunately yielded little to no data. Further analysis of this problem resulted in the fact that a lot of Slovenian data science and big data job positions are still clouded in the massive amounts of application programming, business analysis, team management or in the combination of the three job positions.

To combat this problem, we searched for existing Slovenian data scientists and cross-referenced their personal skills with the few existing open data science job positions. Finally, we used an existing analysis [16] to get a better grasp on the subject. This approach yielded a slightly bigger data pool and gave us the basis for our market analysis [17].

We split the required skill set into soft and hard skills a data scientist should have, to successfully land and keep a data science job in Slovenia.

2.4.1 Soft Skills Analysis

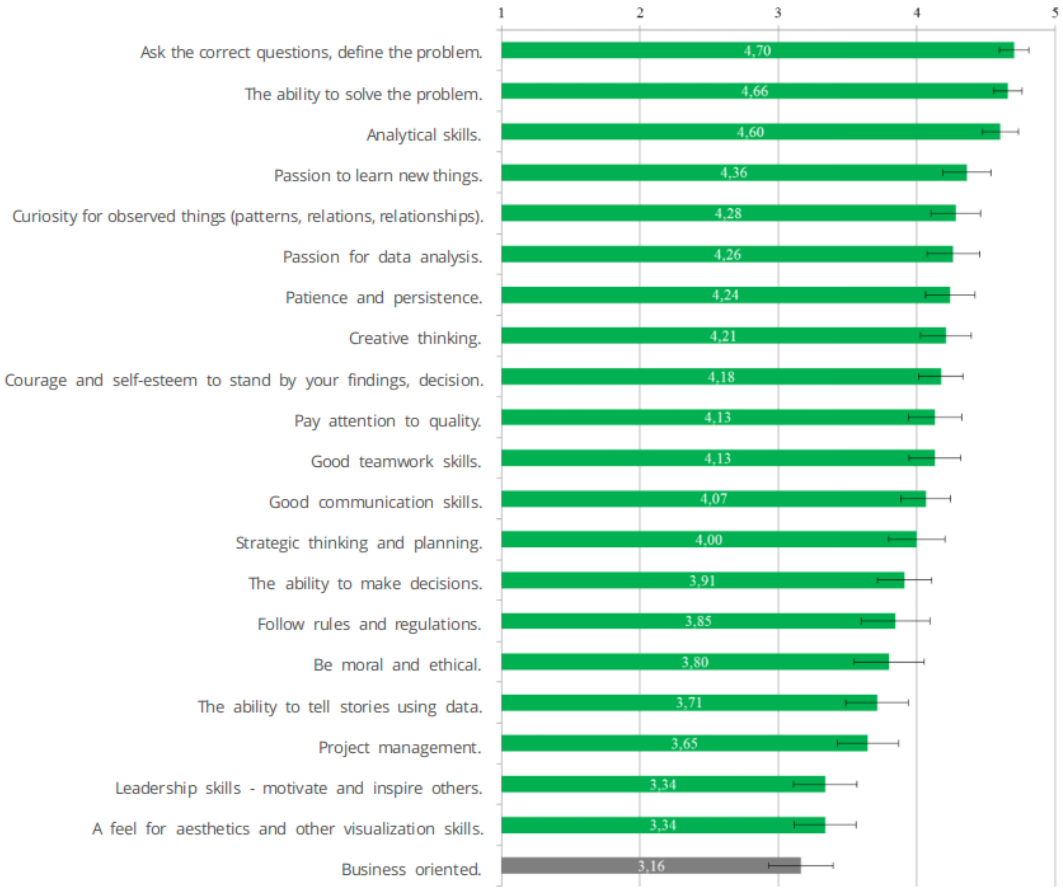


Figure 23 – Chart showing the importance of soft skills for a data scientist [16]

Results show that data scientists should be highly driven self-reliant creative people with good communicational and presentation skills. They should be able to ask the right questions and correctly define the problem, be able to solve the problem or otherwise have the eagerness to explore new technologies to solve the problem. Highly sought after skills are good teamwork and leadership, to solve bigger problems. Data scientists also need to efficiently present their work to other experts and non-experts in the company, which takes patience, perseverance and a positive, make-it-happen, lead-by-example attitude. To round up the soft skill set, these people should be able to work well in a team and cooperate with all the departments in the company.

Figure 23 – Chart showing the importance of soft skills for a data scientist Figure 23 shows the chart of soft skills ranked by existing Slovenian Data Scientists on a scale from 1 – not important to 5 – very important. We gathered the graph from [16] and it can be seen that it fits well with our smaller analysis above.

2.4.2 Hard Skills Analysis

Results show that the Slovenian companies are mostly looking for people who are expert programmers in PYTHON, SQL and R with a strong technical and analytical background (e.g. data processing, data mining, machine learning, optimization, statistics and so on). Former experience in the field is a must-have and any open source projects or other references are “very desirable”.

Educational background should not be ignored. There is a wide range of possibilities, from Computer Science and Mathematics to Business Economics. All of these fields can produce good data scientists as long as each individual is prepared to learn a few things on its own. Other required knowledge includes programming languages such as Java, C#, Python libraries (Tensorflow, Theano, Keras, Caffe, Pandas), JavaScript, Node JS and PHP. Further knowledge of RESTful APIs, Cassandra, MySQL, NoSQL, MongoDB and of course cloud computing (e.g. AWS, Microsoft Azure and so on) is also sought after.

Figure 24 shows us the chart of hard skills ranked by existing Slovenian Data Scientists on a scale from 1 – not important to 5 – very important. We gathered the graph from [16] and as they correctly observed in the thesis, there is a significant under evaluation of modern Data Science tools and techniques among the Slovenian Data Scientists. There could be many reasons for this, however the awareness of this “lack of importance” can be a great motivation and a good success metric for projects like Da.Re. We could aim at increasing the value of these proven modern data science methods and tools among the Slovenian Data Scientists.

The combination of our small research and the analysis done by Grobelnik gives us a comprehensive list of important suggestions on what to include in the curriculums as training courses and activities.

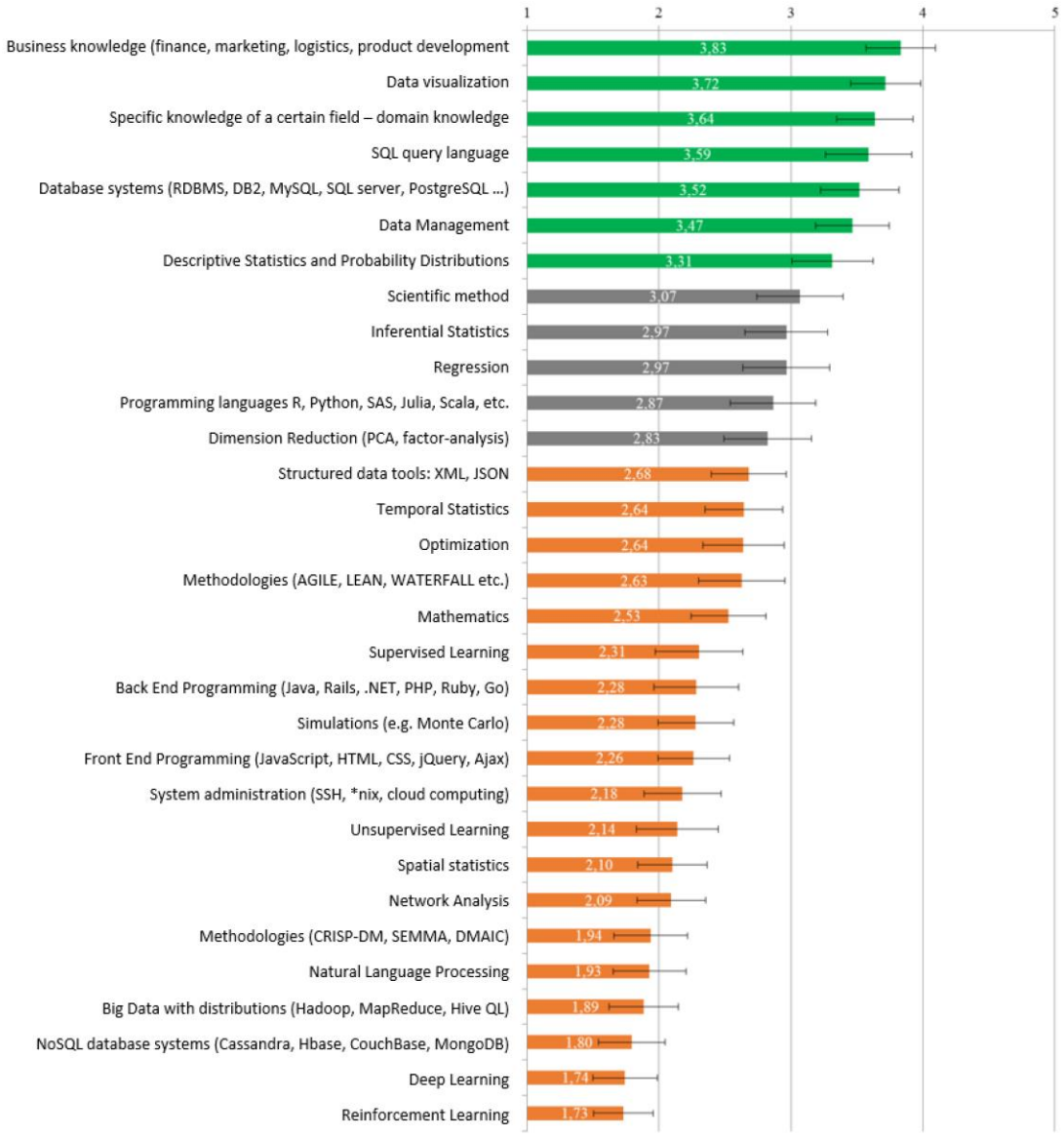


Figure 24 – Chart showing the importance of hard skills for a data scientist [16]

2.5 Market Needs Analysis in Serbia

In recent years in highly developed countries all over the world there is a huge increase of interest for learning highly-required knowledge about big data management. However, in Serbia there is still a less-developed awareness about the role that data science and big data analytics can have for companies, especially for SMEs.

The market needs analysis in Serbia started as in the other Countries, but the number of both advertised and existing job positions in data science found was too low to make a significant analysis.

However, the Nissatech partner, thanks to its network and experience developed in this specific field had the opportunity to interview a few organisations in Serbia that are already partially active

in data science. Questions made were about the type of professional figures wanted and foreseen in the future by the selected IT companies. The survey was put in a general context with a specific attention to data science. The results of these interviews are reported in Chapter 4.

Chapter 3. Existing Training Paths Analysis

This chapter reports on the surveying work for Activity 1.2 and partially Activity 1.1 carried out in Italy, UK, Portugal, Slovenia and Serbia with the objective of taking a snapshot of the study programmes related to Data Science and/or Big Data Analytics available at Higher Education Institutions in those Countries.

The main motivation for this work was the acquisition of a sufficiently detailed picture of the existing official training resources in the project Countries in order to have a base of knowledge on which the effectiveness and innovativeness of the pilot course to be proposed by the project in Phase 2 could be discussed and then evaluated.

3.1 Methodology

The analysis started by discussing among the Consortium the kind of information to retrieve. The partners agreed on focusing on active study programmes in academic year 2016/2017 offered by public or private higher education institutions recognized by the Countries authorities. It was also agreed that the major source of information would be the web sites of the programmes as well as official databases about educational offer in each State.

European Qualifications Framework

We used the European Qualifications Framework (EQF) [18] to classify the existing training paths found across the different Countries involved in the Da.Re. project. This was a convenient way to proceed in order to easily compare, qualitatively and quantitatively, the collected data.

We mainly focused on EQF levels 7 and 6, according to the intended target students/professionals of the Da.Re. project. Let us briefly introduce the kind of programs at each level.

- EQF level 8, PhDs: these are research-oriented programs. Any training needs at this level are usually provided at levels 7 and 6, discussed below.

- EQF level 7, Masters: there is a variety of formats for Master programs in the different Countries. This varies from 12-month residential courses in the UK to 24-month residential courses in Italy, Portugal, Slovenia and Serbia. Even in the UK there is a great variety of master courses with some universities offering M.Phils lasting 2 years at a level between taught masters and PhD by research. Also some UK engineering degrees take 4 or 5 years with the student being awarded a master. Most of the other Countries adhere to the Bologna process (3 years bachelor + 2 years master + 3 years PhD). In Italy, beyond the regular 24-month Masters there are further 12-month programmes that are post-master specializations of regular Master programs. This happens also at EQF level 6. These specializations belongs to the same EQF level of the regular programs.

- EQF level 6, Bachelors: they teach generic skills including finding and synthesizing information, writing essays and reports, a basic level of mathematical and statistical

analysis and some soft skills such as team working. Beyond this they teach domain specific knowledge by discipline such as sociology, economics, physics and informatics. The level of generic skills varies between disciplines and individuals.

- We decided not to consider initiatives that can be classified in EQF level 5 or lower.

Template

A common template to be filled for each programme was defined and is shown in Table 8.

Table 8 – Template used for gathering information about the programmes

Field	Subfield	Description	Value
Title (National Language)		Full title of the training path	
Title (English)		Full title of the training path (official translation if available)	
Explicit reference to DS / BDA?		Does the title explicitly mention Data Science and/or Big Data Analytics? Yes or No	
	Name of the sub-track / Data Science or Big Data Analytics class(es) (English)	In case you answered "No" to the main question, please specify the name of the subtrack or of the class(es) related to Data Science and/or Big Data Analytics within the full training path	
Training Language		Please specify the language in which the training path is taught	
Website		Please specify the specific website dedicated to the training path	
Organisation(s)		Details of the organisation providing the learning path	
	Full legal name (National Language)		
	Full legal name (Latin characters)		

	Department or Area	Please specify the particular department or structure or area within the organization that provides the training path	
	Country		
	Region		
	City		
	Website	Please specify the general website of the Organisation	-
	Type of Organization	What kind of organization is offering the training path? Possible alternatives: HEI (Higher Education Institution), Research Institution, (Small/Medium/Large) Enterprise, Collaboration between HEI and Enterprise, and so on (open field)	
	Is the organisation a public body?		
	Is the organisation a non-profit?		
		Add more lines if there are other organisations	
Official Location		Please specify the official site of the training path	
	Country	Specify if different from that of the Organisation	
	Region	Specify if different from that of the Organisation	
	City	Specify if different from that of the Organisation	
	Address		
Periodicity		Please specify if the training path is being repeated over years or if it is a one-time initiative: Repetated or One-time	

	Starting Year	Please specify the first year in which the training path was offered in case of Repeated Periodicity, otherwise the year of the One-time initiative	
Duration			
	ECTS (full)	Please specify the total number of European Credit Transfer Systems credits associated to the training path (1 ECTS = 25 hours of work, including lectures, labs, practices and individual study)	
	Years (full)	Please specify the number of years expected to complete the training path. In case of months, use decimals.	
	Hours (full)	Please specify the total number of hours of works expected to complete the training path	
	ECTS (DS / BDA)	Please specify the total number of European Credit Transfer Systems credits dedicated to Data Science or Big Data Analytics within the training path (1 ECTS = 25 hours of work, including lectures, labs, practices and individual study)	
	Hours (DS / BDA)	Please specify the total number of hours of works dedicated to Data Science or Big Data Analytics within the training path	
Cost			
	National Currency	Please estimate the maximal total costs for a student to attend the full training path (not including living costs)	
	EUR	Please estimate the maximal total costs for a student to attend the full training path (not including living costs)	

EQF level		Please specify or estimate (using EQF descriptors) the European Qualifications Framework level of the training path. References: https://ec.europa.eu/ploteus/sites/eac-eqf/files/leaflet_en.pdf and https://ec.europa.eu/ploteus/en/content/descriptors-page	
Name of the Degree or Certification Awarded (National Language)		Please specify the full official name of the Degree and/or the kind of Certification that the Organisation will award to the student at the completion of the training path	
Name of the Degree or Certification Awarded (English)		Please specify the full official name of the Degree and/or the kind of Certification that the Organisation will award to the student at the completion of the training path	
General Description of the training path (English)			
	Text	Please provide a short text generally describing the training path (better if taken from official website or official source of information)	
	Keywords for identifying key skills	Please, extrapolate from the given text (or define by yourself) the keywords that are relevant for the topics of Data Science / Big Data Analytics (please use capitalised phrases divided by commas)	
General Description of the Job Title(s) associated to the training path (English)			

	Text	Please provide a short text generally describing the Job Title(s) that is(are) expected to be the output of the training path (better if taken from officiel website or official source of information)	
	Keywords for identifying key skills	Please, extrapolate from the given text (or define by yourself) the keywords that are relevant for the topics of Data Science / Big Data Analytics (please use capitalised phrases divided by commas)	
General Description of the Learning Outcomes of the training path (English)			
	Text	Please provide a short text generally describing the Learning Outcomes of the training path (better if taken from official website or official source of information)	
	Keywords for identifying key skills	Please, extrapolate from the given text (or define by yourself) the keywords that are relevant for the topics of Data Science / Big Data Analytics (please use capitalised phrases divided by commas)	
Contact Person(s)		Please provide a person (or more people in several rows) which can be contacted for obtaining further details about the training path, as those requested for the second part of the survey	
	Name		
	Organisation Role	Role of the person in the Organisation	

	Training Path Role	Role of the person in the management of the training path	
	Email		

The entries of the template were defined with the objective of analysing the following aspects:

- EQF level and specificity of the training program with respect to the topics of Data Science and/or Big Data Analytics. The specificity was determined mainly by the presence of specific keywords in the name of the degree or in the name of an official sub-track. In some cases, in Portugal and Slovenia, also the presence of specific modules within the list of suggested study plans was considered, this requiring a more specific search beyond the entries of the template.
- Geographical distribution in the specific Country.
- Disciplinary areas offering the programmes. This was determined by looking at the departments offering the programs and partially by the general description.
- Costs of the program.
- Moreover, general textual description, job title textual description and general learning outcomes text description were collected for each program. They were fundamental to understand with a better detail the characteristics of the training path also with respect to the skills of the intended professional figure that the course want to provide.
- Finally, contact information for further investigation, possibly in the next Phases of the project, was collected.
- All the partners initially adopted the data collection approach using the template. Then, the large numbers of programmes present in the UK showed that the level of granularity of the data given by the template was inappropriate for the UK. A different approach was taken, as reported in the relative section below.

Teaching methodologies

Another fundamental aspect of training paths to be considered is the teaching methodologies that are used in the courses, which can be uniform throughout a program or can be combined in different ways. This aspect is important to consider for the Da.Re. project because it can affect a lot the innovativeness of the pilot course that will be proposed in Phase 2 of the project.

A brief description of methodologies found in the analysis follows:

- Classical: usually composed of lectures, classes and seminars, laboratory sessions with periodic homework and end-of-course examination. Usually the assessment counts towards certification.

- E-Learning: this can be very variable. It ranges from self-made teaching videos on YouTube, which can be very good and also very poor, to high quality multimedia certificated education. For example the UK Open University does most of its teaching online providing supported open education. For most courses students are allocated a tutor who will hold periodic online face-to-face tutorials, answer student questions online and by phone, mark the students assignments and give periodic tutorials at study centres across the UK. Courses are designed and produced by teams of highly qualified academics, editors, graphic designers, video producers and educational software designers. Typically, a 300 hours course (12 ECTS) can cost in the order of 1,000,000 pounds to produce, with the expectation the course will be studied by 5 to 10 thousands students. Presentation of Open University courses costs in the order of thousands of pounds for staff time, printing and supporting online and other teaching infrastructures.

- MOOCs: Massive Open Online Courses provide a wide range of training and education in all areas including Data Science. Typical MOOCs involve 2 to 8 hours of study per week and may last up to 10 weeks, i.e., between 4 to 40 hours. They mostly assume study skills equivalent to EQF level 6. The quality varies depending on the authors. The providers of MOOCs include Coursera [19] and edX [20] in the USA and Future Learn [21] based in the UK. Courses are usually available online to all Countries. A typical MOOC will have a registration of thousands of students but attendance may decrease rapidly to hundreds or less.

- Boot camps: The term “boot camp” originates from the camps that the U.S. Navy and Marines ran to train their recruits, but is now used for any short and intensive training program[1]. Boot camps regarding topics of Data Science and/or Big Data Analytics are also popular in the UK. With the surge in popularity of data science among institutes and individuals, boot camps in the field are being run to quickly teach people the skills they need to make a career for themselves in data science. The demand for these courses on “data science bootcamp” during 2011-2017 has increased rapidly over the last three years and the origins of the commercial concept of boot camps for data science can clearly been seen as the USA, see Figure 25.

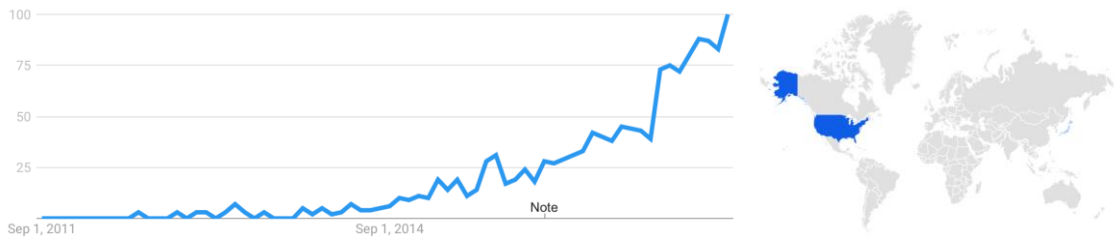


Figure 25 – Searching for Boot Camps and geographical area

[1] http://www.dictionary.com/browse/boot-camp?s=t – 25/08/17

Most boot camps are commercial and employment-oriented. In Italy, a “data science bootcamp”-like, BigDive [2] initiative is a training program to boost the technical skills people need to dive into the big data universe, to extract value and to create impact.

It is important to report that in all the involved Countries, but UK, the collected training paths are taught in the classical way using frontal lectures, exercises, individual reading and individual/group projects.

In UK, in addition to the classical teaching methods, there is a platform with a large offer of MOOCs about data science. There are also other platforms based in other Countries that offer this kind of MOOCs. However, it has to be said that these courses are available on the web independently of the Country of the learner.

3.2 Training Path Analysis in Italy

In the following we summarize the results of the survey in the academic year 2016/2017 in Italy. The source data, in the form of filled templates, can be retrieved annex 1.

3.2.1 Presentation of Italian programmes

We derived a dataset containing 23 entries about Italian Higher Education degrees of European Qualifications Framework (EQF) level 6 and 7. Table 9 shows the entries, the university offering them and the type of the programme.

Table 9 – Training path in Italian universities

N	University	Code
1	Bologna	BSc-SC
2	Bologna	MSc-C
3	Firenze	MSc-SC
4	Genova	BSc
5	Genova	MSc-C
6	L'Aquila	MSc-C
7	Luiss	BSc-SC
8	Milano Bicocca	BSc-SC
9	Milano Cattolica	BSc-SC
10	Milano Politecnico	BSc-SC
11	Molise	BSc-SC
12	Pisa	MSc
13	Pisa	MSc-SC
14	Roma Sapienza	Msc-SC I

[2] http://www.bigdive.eu/about/

15	Roma Sapienza	Msc-SC II
16	Roma Sapienza	MSc
17	Roma Tor Vergata	MSc-C
18	Roma Tor Vergata	MSc-SC
19	Siena	MSc
20	Torino	MSc-SC
21	Torino	MSc
22	Venezia	BSc-C
23	Venezia	MSc-C

Code	**Explanation**
BSc	Bachelor Degree
BSc-C	Curriculum within a Bachelor Degree
BSc-SC	Post-bachelor Specialization (about 60 ECTS)
MSc	Master Degree
MSc-C	Curriculum within a Master Degree
MSc-SC	Post-master Specialization (about 60 ECTS)

The chosen categories reflect the typical Italian scenario. Universities offer Bachelor Degrees (180 ECTS, 3 years, EQF level 6) and Master Degrees (120 ECTS, 2 years, EQF level 7) together with post-bachelor or post-master specialization programs that have a total workload of about 60 ECTS (one year). Note that the post-bachelor and post-master programs are improperly called “Masters” in Italian; this can create confusion in a European or international contexts, thus we decided to call them “post-bachelor” and “post-master” specialization program in this document. These programs are to be considered at the same EQF level of the ones they are specialization of.

We collected the programs that are mostly focused on Data Science and/or Big Data Analytics, i.e., the ones in which the name of the degree explicitly mentions these keywords, and those in which this happens for a significant sub-track (or curriculum) while the name of the degree is a general one.

The geographical distribution of these courses is depicted in Figure 26:

Figure 26 – Map of Italian universities offering paths in Data Science or Big Data Analytics

Figure 26 also shows that the majority of them is located in the northern and central part of the Country. The unique southern university offering a path in Data Science or Big Data Analytics is the University of Molise located in Campobasso.

In Italy there are 97 universities recognised by the Ministry of Education and Research [22]. Only 15 of them (15.5%) offer a course (at some level) in Data Science or Big Data Analytics.

By these data we can claim that the topics of interest are currently lightly covered in Italy. Certainly, in the last few years, there was a reaction to the global attention on the importance of competencies in data analytics not only in (big) companies, but also at the university level. Some of the major universities started to offer training opportunities in Data Analysis. We expect a growth of offers in the next years.

Figure 27,Figure 28 andFigure 29 show graphs that recaps the Italian situation about training paths.

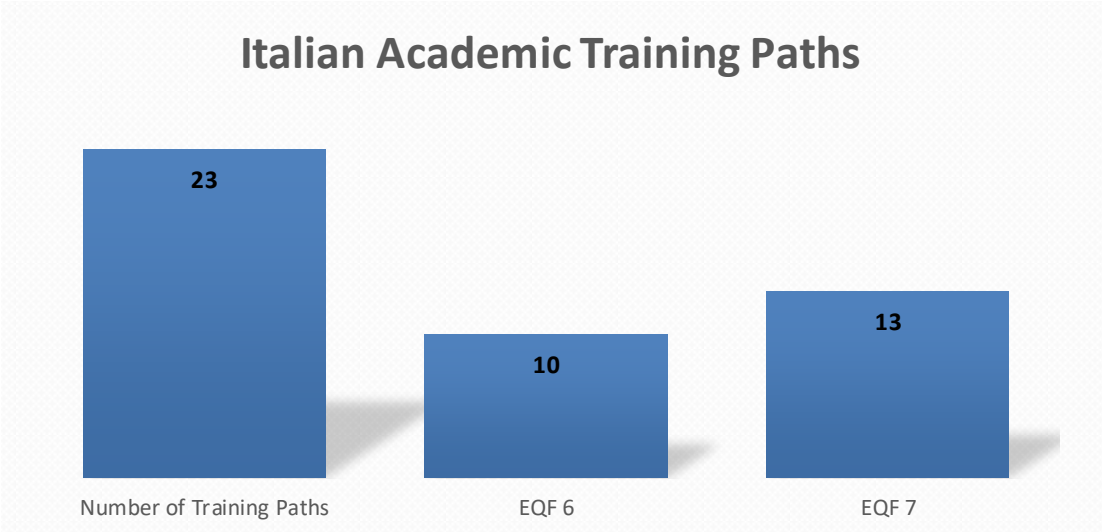


Figure 27 – Italian Academic Training Paths (Master and Bachelor)

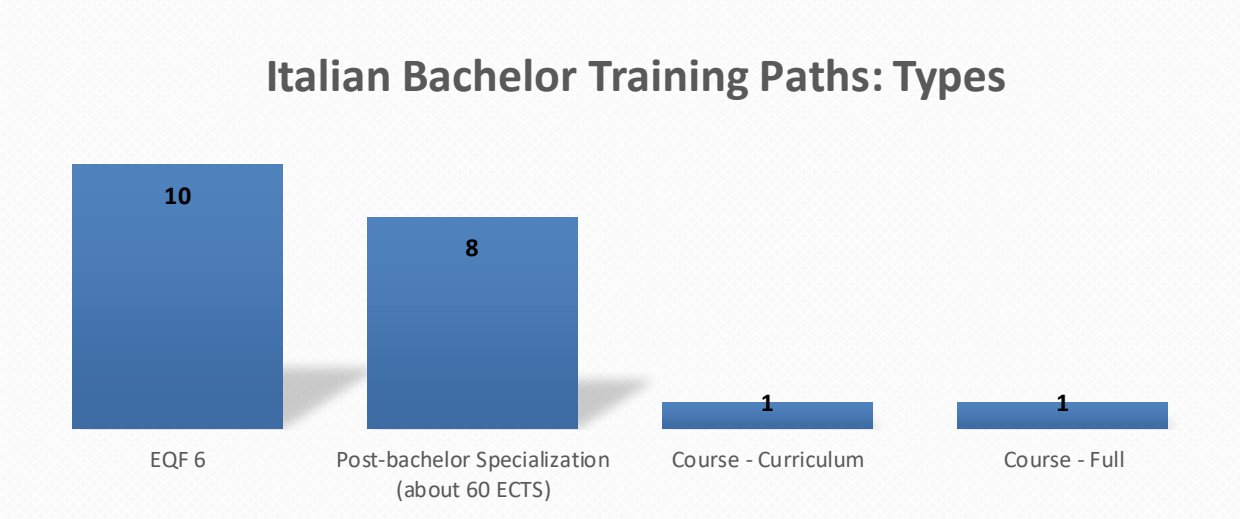


Figure 28 – Bachelors and Post- Post Specialization in Italy

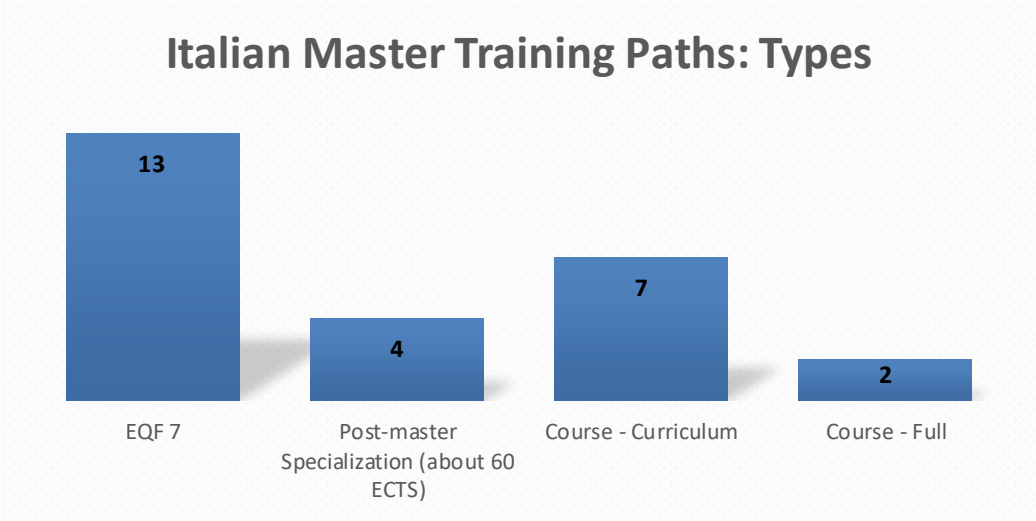


Figure 29 – Master and Post- Master Specialization in Italy

At EQF level 6, as shown in Figure 28, the majority of training paths are post-bachelor specializations. Only one Bachelor degree exists that explicitly focus on the topics of the Da.Re. project and another one has a curriculum within. The situation is different at EQF level 7, Figure 29, where the majority of training paths is composed by tracks within more general Master degrees. Still there is a fair number of specialization courses.

We can also report that on a total of 912 EQF level 6 degrees in scientific areas active in Italy only 10 (1.097%) are related to the target topics. Moreover, on a total of 989 of EQF level 7 degrees in scientific areas active in Italy only 13 (1.314%) are related to the target topics [22]. These numbers confirm the current light coverage of data analytics competences and skills in Italy.

There are currently also some PhD courses (EQF level 8) in Italy that have Data Analysis and/or Big Data Analytics within their topics. We collected the following list of 6 active programs in 2016/2017:

- University of Bari. PhD course in Big Data Analytics. Web: http://www.di.uniba.it/~ceci/micFiles/courses/bigdata/bigdata.html
- IMT School for Advanced Studies Lucca. PhD course in Economics, Management and Data Science. Web: https://www.imtlucca.it/phd/2016-17/economics
- University of Verona. PhD course in Computer Science, sub-track “From Big Data to Big Multidimensional Data: Models, Issues, Challenges”. Web: http://www.univr.it/main?ent=ava&cs=635&id=432&scuoladott=14
- University of Torino. PhD course in Computer Science with Data Science modules. Web: http://dott-informatica.campusnet.unito.it/do/home.pl

- University of Genova. PhD course in Social Sciences with Data Science modules. Web: http://www.disfor.unige.it/didattica/post-laurea/dottorato

- University of Ferrara. PhD course in Engineering Sciences with Data Science modules. Web: http://www.unife.it/studenti/dottorato/corsi/riforma/engineering

There are also 4 PhD programs that will be activated in 2017/2018:

- University of Camerino. Interdisciplinary PhD scholarships of Data Science. Web: https://www.unicam.it/sites/default/files/bandi/2017/06/2017%20Call%20for%20admission%20for%20Doctoral%20Degree%20programs.pdf

- University of Torino. PhD course in Modeling and Data Science. Web: http://dottorato-mds.campusnet.unito.it/do/home.pl

- School of Advanced Studies Sant'Anna of Pisa jointly with Scuola Normale Pisa, University of Pisa, Italian National Council for Research (CNR), IMT School for Advanced Studies Lucca. PhD course in Data Science. Web: https://www.santannapisa.it/en/news/presentation-new-joint-phd-program-data-science-friday-june-16-noon-scuola-normale-pisa-sala

- University of Bologna jointly with ISI Foundation, Italian National Council for Research (CNR), Polytechnic of Milano, Italian Institute of Technology (IIT), Italian National Institute for Nuclear Physics (INFN). PhD course in Data Science and Computation. Web: http://www.unibo.it/it/didattica/dottorati/2017-2018/data-science-and-computation

3.2.2 Training paths by discipline

Interesting information emerging from our data is also the academic areas that offer courses in Data Science or Big Data Analytics. The summary is shown in Figure 30.

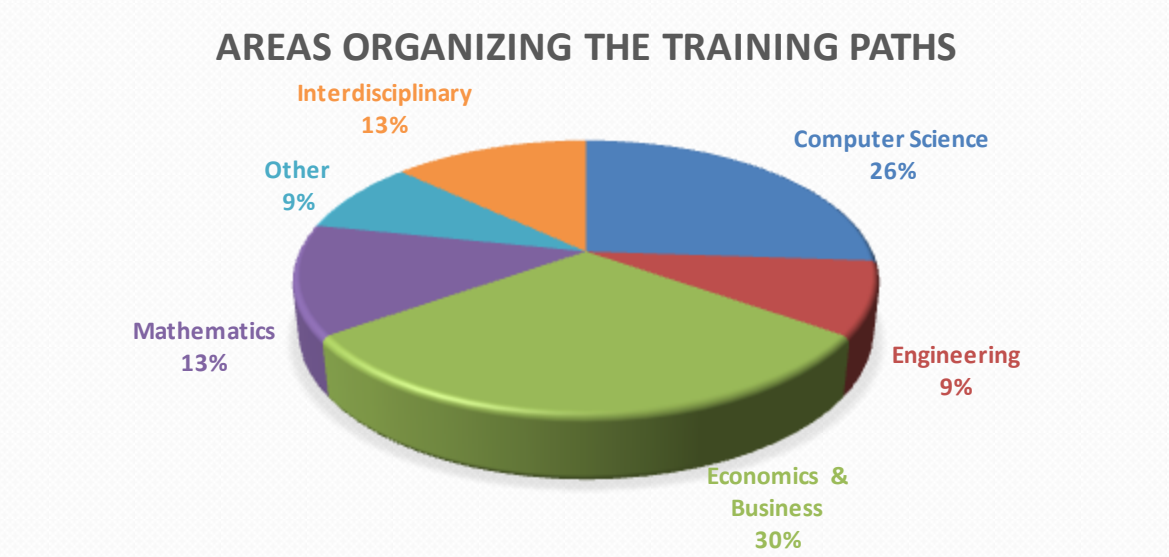


Figure 30 – The main Academic areas involving in organizing the Italian training paths

The main result is that in Italy 30% of the programs are offered by departments of *Economics or Business*. *Computer Science* and *Engineering* areas, taken together, cover the 35%. Courses offered by departments in which the engineering disciplines, not only informatics, are prevalent were classified as *Engineering*. This is related to the fact that in Italy two distinct sectors exist for informatics: computer science (mostly in science departments) and informatics engineering (in engineering departments). The *Interdisciplinary* category corresponds to situations in which the program is offered by collaborations between the main other areas. The category *Other* is used when the program is offered by other areas (in collaboration or not) that are not Economics and Business, Computer Science, Engineering or Mathematics.

It is interesting also to analyse how these areas are distributed with respect to the type of the degrees (see Figure 31) . The following two histograms show this distribution for the EQF levels 6 and 7.

Figure 31 – Italian training paths with respect to Bachelor (top) and Master (bottom)

It is worth noticing that all (but one) programs offered by the Economics and Business area are Post-bachelor specialization course. The remaining one is still a specialization, but at the master level. On the contrary, the other areas are more oriented to offer more structured programs that last generally longer. Moreover, we observe that the Mathematics area is the only one that tends to offer full degrees dedicated to data analytics.

3.2.3 Costs analysis

Finally, we report two histograms, Figure 32, about the cost per year of the found training paths.

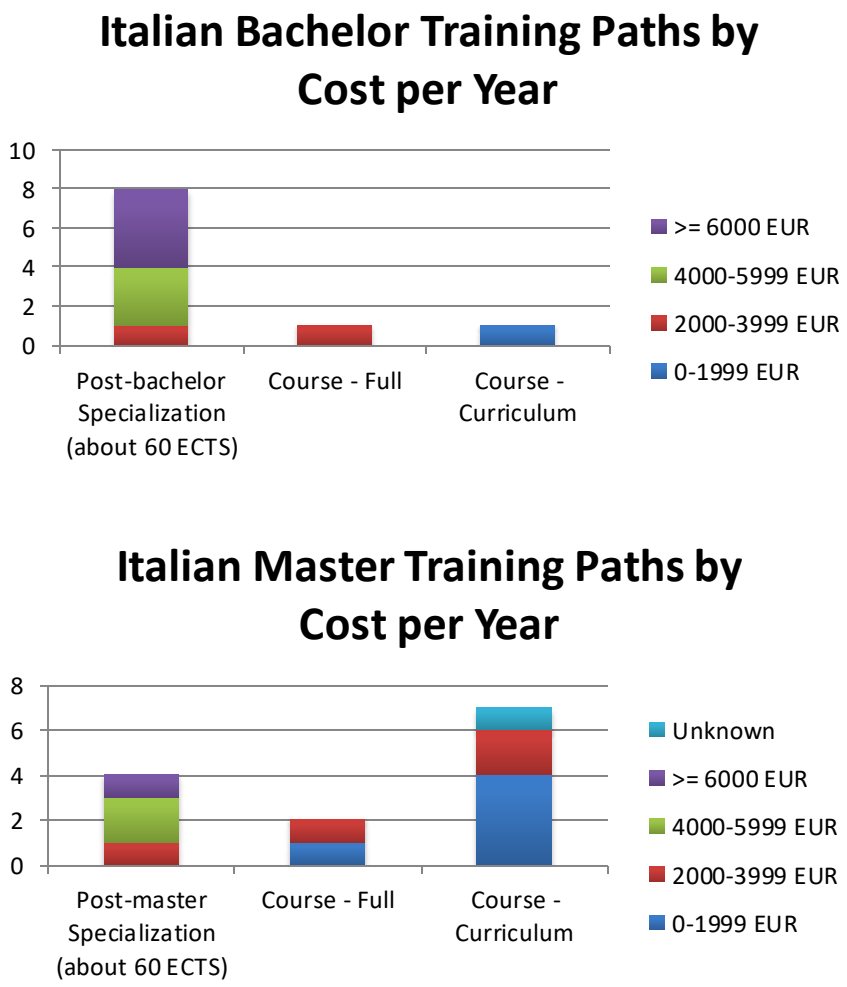


Figure 32 – Cost analysis for Bachelor (top) and Master (bottom)

It is clear from the graphics that both the post-bachelor and the post-master specialization degrees are the most expensive, while the classical structured bachelor and master degrees have a lower cost. In particular, the large majority of the most expensive courses belong to the area of Economics and Business.

3.3 **Training Path Analysis in UK**

As shown in Figure 6 the UK has the highest number of data scientists in Europe. In this section it will be shown that over half of the universities in the UK provide masters courses in data science. This reflects the independent and entrepreneurial nature of British universities, which have no direct government control. In the UK universities are independent organisations enabled by their legal charters to award degrees and other qualifications. What individual universities can do is controlled only by a regulatory framework that aims to maintain quality but British universities cannot be micromanaged by the government. For example, British universities decide their own

staffing policies and may create professorships as part of investment programs responding to perceived 'market' needs.

The view of education as a market rather than a public service has emerged in the UK over a number of decades as successive governments of all colours have tried to increase the proportion of the population with University degrees while trying to reduce the public funds used for the purpose. Undergraduates in the UK pay fees in the order of € 10,000 p.a. and many end their degrees with debts to the government owned Student Loan Company in the order of € 30,000 and € 50,000 which are, in principle, repaid through the tax system as graduates earn above threshold amounts as their careers develop. In practice some do not earn above these thresholds and their student loans are never repaid.

The position is different at masters level where students typically fund their studies without government loans. What we observed in the UK is a market in which there is a great industrial and commercial demand for well-trained data scientists and the universities which are able to supply one-year masters graduate and make money. Within this market it seems that individuals are willing to investing themselves with fees typically in the order of € 10,000 and living costs for a year of another € 10,000. In the UK it is likely that this lucrative market for the universities is close to saturation, with competition and quality being high.

3.3.1 Presentation of UK Programmes

Research Methodology

Initially the research proceeded with a web search on 'data science courses UK'. This produced a list of many courses at British universities as illustrated in Figure 33.

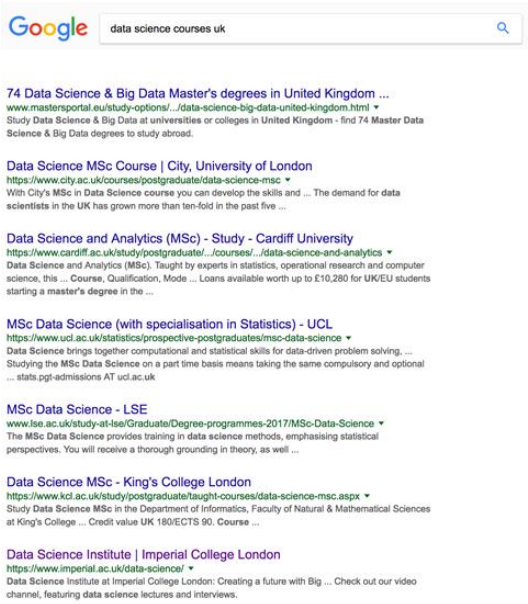


Figure 33 - A web search on 'data science courses UK' gave many results

Initially we completed the Da.Re. template in Table 10.

Table 10 – Template used for gathering information about the programmes

Field	Subfield	Description	
Title (National Language)		Full title of the training path	MSc Data Analytics
Title (English)		Full title of the training path (official translation if available)	MSc Data Analytics
Explicit reference to DS / BDA?		Does the title explicitly mention Data Science and/or Big Data Analytics? Yes or No	yes
Training Language		Please specify the language in which the training path is taught	English
Website		Please specify the specific website dedicated to the training path	http://www2.warwick.ac.uk/fac/sci/dcs/admissions/postgraduate/da/
Organisation(s)		Details of the organisation providing the learning path	
	Full legal name (National Language)		University of Warwick
	Full legal name (Latin characters)		University of Warwick
	Department or Area	Please specify the particular department or structure or area within the	Department of Computer Science

		organization that provides the training path	
	Country		UK
	Region		Warwickshire
	City		Coventry
	Website	Please specify the general website of the Organisation	http://www2.warwick.ac.uk
	Type of Organization	What kind of organization is offering the training path?	HEI
	Is the organisation a public body?		Yes
	Is the organisation a non-profit?		Yes
	Address		Coventry CV4 7A
Duration			12 months
	ECTS (full)	Please specify the total number of European Credit Transfer Systems credits associated to the training path (1 ECTS = 25 hours of work, including lectures, labs, practices and individual study)	JHJ estimates 220 days x 8 hours =1760 hours = 704 ECTS
	Years (full)	Please specify the number of years expected to	1

		complete the training path.	
	Hours (full)	Please specify the total number of hours of works expected to complete the training path	1760
	Hours (DS / BDA)	Please specify the total number of hours of works dedicated to Data Science or Big Data Analytics within the training path	1760

Cost	**National Currency**	Please estimate the maximal total costs for a student to attend the full training path (exclude living costs)	UK and EU students: £7,780 (GBP) Non-EU students: £17,500 (GBP)
	EUR	Please estimate the maximal total costs for a student to attend the full training path (include living costs)	UK and EU students: €9,200 Non-EU students: €21,300
EQF level		Please specify or estimate (using EQF descriptors) the European Qualifications Framework level of the training path.	7
Name of the Degree or Certification Awarded (National Language)		Please specify the full official name of the Degree and/or the kind of Certification that the Organisation will award to the student at the	MSc Data Analytics

		completion of the training path	
Name of the Degree or Certification Awarded (English)		Please specify the full official name of the Degree and/or the kind of Certification that the Organisation will award to the student at the completion of the training path	MSc Data Analytics

	Text	Please provide a short text generally describing the Job Title(s) that is(are) expected to be the output of the training path (better if taken from official website or official source of information)	As one of the most employable disciplines across modern industries, an advanced qualification in data analytics is the ideal way for talented scientists to distinguish themselves. The research focus of our course means our students develop with a rigorous understanding of principles alongside a breadth of technical skills, making them suited to science and industry. Aside from industry, many of our graduates go on to pursue research careers,
	Keywords for identifying key skills	Please, extrapolate from the given text (or define by yourself) the keywords that are relevant for the topics of Data Science / Big	Data scientists, data technology specialist; financial modelling

		Data Analytics (please use capitalised phrases divided by commas)	
	Text	Please provide a short text generally describing the Learning Outcomes of the training path (better if taken from official website or official source of information)	Research methods; Data mining; Foundations of data analytics; Options: High performance computing; algorithmic game theory; image and video analysis; advanced computer security; social informatics; natural language processing; spatial methods and practice in urban science; urban data - theory and methodology
Contact Person(s)	**Name**		Dr Adam Chester
	Organisation Role	Role of the person in the Organisation	Associate Professor
	Training Path Role	Role of the person in the management of the training path	Director of Postgraduate Studies
	Email		apc@dcs.warwick.ac.uk

However, after completing the templates for a few universities it became apparent that to do so for all the hundreds of courses in the UK would be overwhelming while at the same time the detailed level of granularity of the data is inappropriate for the UK. For this reason we took another approach.

There are websites that bring together information on the courses given at UK universities and in the first instance we used these to compile a list of about 100 courses on data science, mostly one-year masters, given by British universities. On checking this list we found some courses that had been omitted. In the end we made a list of the 150 UK universities shown in Table 11, taken from [23], and searched each individually for data science courses.

Table 11 - The 150 Universities in the UK [23]

University of Aberdeen	Buckinghamshire New University, High Wycombe
Abertay University	University of Cambridge
Aberystwyth University	Canterbury Christ Church University
Anglia Ruskin University, Cambridge	Cardiff Metropolitan University (formerly University of Wales Institute Cardiff)
Arden University, private, distance learning (head office in Coventry)	Cardiff University
Aston University, Birmingham	University of Chester, Chester and Warrington
Bangor University	University of Chichester
University of Bath	Coventry University, including CU Coventry, CU Scarborough and CU London
Bath Spa University	Cranfield University
University of Bedfordshire, Luton and Bedford	University for the Creative Arts
University of Birmingham	University of Cumbria, Carlisle (main campus)
Birmingham City University	De Montfort University, Leicester
University College Birmingham	University of Derby
Bishop Grosseteste University, Lincoln	University of Dundee
University of Bolton	Durham University, Durham and Stockton-on-Tees (Queen's Campus)
The Arts University Bournemouth	University of East Anglia, Norwich
Bournemouth University	University of East London
BPP University	Edge Hill University, Ormskirk, Lancashire
University of Bradford	University of Edinburgh
University of Brighton	Edinburgh Napier University
University of Bristol	University of Essex, Colchester and Southend-on-Sea
Brunel University, Uxbridge and London	University of Exeter
University of Buckingham	

Falmouth University
University of Glasgow
Glasgow Caledonian University
University of Gloucestershire, Cheltenham, Gloucester and London
Glyndŵr University, Wrexham
University of Greenwich
Harper Adams University, Newport, Shropshire
Heriot-Watt University, Edinburgh and Galashiels
University of Hertfordshire, Hatfield
University of the Highlands & Islands
University of Huddersfield, Huddersfield & Barnsley
University of Hull, Hull and Scarborough
Imperial College London
Keele University, Staffordshire
University of Kent, Canterbury and Medway
Kingston University
University of Central Lancashire, Preston and Burnley
Lancaster University
University of Leeds
Leeds Beckett University
Leeds Trinity University
University of Leicester
University of Lincoln, Lincoln, Hull, Riseholme and Holbeach
University of Liverpool

Liverpool Hope University
Liverpool John Moores University
University of London
Birkbeck, University of London
City, University of London
Royal Central School of Speech and Drama
Courtauld Institute of Art
Goldsmiths, University of London
Heythrop College
Institute of Cancer Research
King's College London
London Business School
London School of Economics and Political Science (LSE)
London School of Hygiene and Tropical Medicine
Queen Mary, University of London
Royal Academy of Music
Royal Holloway, University of London, Egham
Royal Veterinary College
St George's, University of London
School of Oriental and African Studies (SOAS)
University College London (UCL)
London Metropolitan University
London South Bank University
Loughborough University

University of Manchester
Manchester Metropolitan University
Middlesex University, London
Newcastle University
Newman University, Birmingham
University of Northampton
Northumbria University, Newcastle upon Tyne
Norwich University of the Arts
University of Nottingham
Nottingham Trent University
The Open University, Milton Keynes
University of Oxford
Oxford Brookes University
University of Plymouth
University of Portsmouth
Queen Margaret University, Edinburgh
Queen's University Belfast
University of Reading
Regent's University London
The Robert Gordon University, Aberdeen
Roehampton University, London
Royal Agricultural University, Cirencester
University of Salford
University of Sheffield

Sheffield Hallam University
University of South Wales, merger of University of Wales, Newport and University of Glamorgan
University of Southampton
Southampton Solent University
University of St Andrews
University of St Mark & St John, Plymouth
St Mary's University, Twickenham
Staffordshire University, Stoke-on-Trent, Stafford and Lichfield
University of Stirling, Bridge of Allan
University of Strathclyde, Glasgow
University of Suffolk, Ipswich, Bury St Edmunds, Great Yarmouth, Lowestoft
University of Sunderland
University of Surrey, Guildford
University of Sussex, Falmer and Brighton
Swansea University
Teesside University, Middlesbrough and Darlington
University of Ulster, Belfast
University of the Arts London
Ulster University, Coleraine, Jordanstown, Magee and Belfast
University of Law
University of Wales (will merge with UWTSD in 2017)
University of Wales, Trinity Saint David (UWTSD)

University of Warwick, Coventry
University of the West of England, Bristol
University of the West of Scotland, Paisley, Hamilton, Ayr & Dumfries
University of West London, Ealing and Brentford
University of Westminster, London

University of Winchester
University of Wolverhampton
University of Worcester
University of York
York St John University

British universities are widely distributed across the UK, as shown in Figure 34.

Figure 34 - British universities are widely distributed across the UK

In the UK universities can be established by private or public organisations. Over the last two decades many former collages have gained university status, and a number of privately controlled university organisations have been established. In principle, anyone can set up a university if they can fulfil the legal conditions.

For each of the universities in the list above a search was done on the courses they provide in data science. The results of this research are given in Table 12 below.

Table 12 - University courses on data science given in the UK

Course Provider	Course name	Level	Cost
Aberystwyth University	Data Science	MSc	£6,950.00
Arden University	Data Analytics (lots of courses)	MSc	£8,550.00
Aston University	Business Analytics	MSc	£11,950.00
Bath University	Data Science	MSc	£9,500.00
Bedfordshire University	Data Science	BSc	£9,250.00
Birkbeck University	Advanced Computing Technologies	MSc	£8,475.00
Birkbeck University	Data Analytics	MSc	NA
Birkbeck University	Cloud and Data Technologies (Part Time)	PGCert	£2,825.00
Birkbeck, University	Data Science (Part Time)	MSc	£4,250.00
Birmingham City University	Big Data Analytics	MSc	£7,900.00
Bournemouth University	Applied Data Analytics	MSc	£6,000.00
Bradford University	Big Data Science and Technology	MSc	£7,750.00
Brighton University	Data Analytics	MSc	£6,120.00
Bristol University	Advanced Computing-ML, Data Mining, HPC	MSc	£10,200.00
Brunel University	Data Science and Analytics	MSc	£9,500.00
Cardiff Metropolitan University	Data Science	MSc	£7,000.00
Cardiff University	Data Science and Analytics	MSc	£8,500.00
Cardiff University	Data Science and Analytics (Part Time)	MSc	£4,250.00
Cardiff University	Computational and Data Journalism	MSc	£8,500.00
Cardiff University	Computer Science and Informatics	PhD	NA
Chichester University	Data Science and Analytics	MSc	£8,000.00
City University London	Data Science	MSc	£9,000.00
City University London	Health Informatics	MSc	£9,000.00

Coventry University	Data Science and Comptutational Intelligence	MSc	£7,374.00
Cranfield University	Environmental Data Science	MSc	£7,800.00
De Montfort University	Data Analytics	MSc	£5,350.00
De Montfort University	Business Intelligence Systems and Data Mining	MSc	£5,350.00
Derby University	Data Science	BSc	£9,250.00
Derby University	Big Data Analytics	MSc	£12,240.00
Dundee University	Data Science	MSc	£7,450.00
East Anglia University	Knowledge discovery and Data Mining	MSc	£7,300.00
East Anglia University	Knowledge discovery and Data Mining (Part Time)	MSc	£3,650.00
East London University	Data Science	MSc	£8,100.00
East London University	Professional Doctorate in Data Science	Prof Doc	£24,300.00
Edge Hill University	Big Data Analytics	MSc	£4,860.00
Edge Hill University	Data Science	BSc	NA
Edinburgh Napier University	Data Science	MSc	£4,500.00
Edinburgh University	Informatics	MSc	£9,100.00
Edinburgh University	High Performance Computing With Data Science	MSc	£10,800.00
Edinburgh University	Data Science	PhD	NA
Edinburgh University	Data Science	MSc	£12,300.00
Essex University	Data Science and Analytics	BSc	£9,250.00
Essex University	Data Science	MSc	£6,250.00
Exeter University	Data Science (Professional)	MSc	£26,400.00
Exeter University	Policy Analytics	MSc	£7,500.00
Glasgow Caledonian University	Big Data Technologies	MSc	£5,000.00

Glasgow University	Data Analytics	MSc	£7,250.00
Glasgow University	Data Science	MSc	£7,700.00
Goldsmiths University	Data Science	BSc	£9,250.00
Greenwich University	Big Data and Business Intelligence	MSc	£9,250.00
Greenwich University	Enterprise Systems and Database Administration	MSc	£9,250.00
Heriot Watt University	Data Science	MSc	£5,200.00
Imperial College	Business Analytics	MSc	£27,000.00
Kent University	Advanced Computer Science (Cloud & Big Data)	MSc	£6,500.00
Kings College London	Data Science	MSc	£9,450.00
Kings College London	Big Data in Culture and Society	MA	£9,450.00
Kingston University	Network and Data Communications	MSc	£6,400.00
Lancaster University	Data Science	MSc	£8,000.00
Leeds University	Data Science and Analytics	MSc	£8,500.00
Leeds University	Advanced Computer Science (Data Analytics)	MSc	£10,000.00
Leeds University	Consumer Analytics and Marketing Strategy	MSc	£10,500.00
Leeds University	Business Analytics and Decision Science	MSc	£10,500.00
Leeds University	Health Informatics	MSc	£7,500.00
Leeds University	Health Informatics	PGDip	£5,000.00
Leeds University	Business Analytics	BSc	£9,250.00
Leicester University	Data Analysis for Business Intelligence	MSc	£9,730.00
Liverpool John Moores Univ	Data Science	MSc	£6,700.00
Liverpool John Moores Univ	Data Science	BSc	£9,250.00
Liverpool University	Big Data and High Performance Computing	MSc	£6,000.00
Liverpool University	Big Data and HPC with a year in industry	MSc	£6,000.00
London Metropolitan	Data Analytics	MSc	£8,100.00

London School of Economics	Data Science	MSc	£25,944.00
London South Bank University	Data Science	MSc	£8,300.00
London South Bank University	Computing Science and Informatics	PhD	£8,600.00
Loughborough University	Cyber Security and Big Data	MSc	£10,000.00
Manchester Metropolitan Univ	Data Science	MSc	£7,560.00
Manchester University	Health Data Science	MSc	£9,000.00
Manchester University	Health Data Science	PGCert	£3,600.00
Manchester University	Health Data Science	PGDip	£7,200.00
Middlesex University	Data Science	MSc	£4,500.00
Middlesex University	Visual Analytics	MSc	£4,500.00
Newcastle University	Cloud Computing for Big Data	MRes	£4,540.00
Newcastle University	Cloud Computing for Big Data	PgDip	£3,025.00
Northumbria University	Information Science (Data Analytics)	MSc	£3,111.00
Northumbria University	Information Technology and Data Science	BSc	£9,250.00
Nottingham Trent University	Data Analytics for Business	MSc	£10,000.00
Nottingham University	Data Science	BSc	£9,250.00
Oxford Brookes University	Data Analytics for Government (Part Time)	MSc	£7,200.00
Plymouth University	Data Science and Business Analytics	MSc	£6,500.00
Portsmouth University	Data Science and Analytics	BSc	NA
Portsmouth University	Data Analytics	MSc	£5,200.00
Queen Mary University	Big Data Science	MSc	£8,700.00
Queen Mary University	Big Data Science with Industrial Experience	MSc	£8,700.00
Queen Mary University	Internet of Things (Data)	MSc	£8,700.00

Queen Mary University	Business Analytics	MSc	£8,700.00
Queen Mary University	FT IoT (Data) with Industrial Experience	MSc	£8,700.00
Queens University Belfast	Data Analytics	MSc	£5,500.00
Regent's University London	Digital Marketing and Analytics	MSc	£17,300.00
Robert Gordon University	Data Science	MSc	£6,750.00
Royal Holloway University	Data Science and Analytics	MSc	£10,400.00
Salford University	Data Telecommunications and Networks	MSc	£7,380.00
Salford University	Data Science	MSc	£7,380.00
Sheffield Hallam University	Big Data Analytics	MSc	£6,400.00
Sheffield Hallam University	Business Analytics	BA	£12,750.00
Sheffield University	Data Analytics	MSc	£10,970.00
Sheffield University	(Eng) Data Communications	MSc	£10,970.00
Sheffield University	Data Science	MSc	£8,000.00
Southampton Solent University	Data Analytics Engineering	MSc	£7,750.00
Southampton University	Marketing Analytics	MSc	£11,500.00
Southampton University	Business Analytics and Management Sciences	MSc	£11,500.00
Southampton University	Data Science	MSc	£9,000.00
St Andrews University	Data-Intensive Analysis	MSc	£7,500.00
Staffordshire University	Data Science	BSc	£9,250.00
Staffordshire University	Big Data	MSc	£7,200.00
Stirling University	Data Science for Business	MSc	£6,200.00
Strathclyde University	Advanced Computer Science with Big Data	MSc	£6,000.00
Strathclyde University	Information Management	MSc	£6,000.00
Strathclyde University	Data Analytics	MSc	£9,500.00
Strathclyde University	Data Analytics	BSc	£9,250.00

Sunderland University	Data Science	MSc	£4,750.00
Sunderland University	Data Science (Part Time)	MSc	£4,770.00
Surrey University	Data Science for Health	BSc	NA
Surrey University	Biomedicine with Data Science	BSc	NA
Surrey University	Business Analytics	Msc	£11,500.00
Sussex University	Advanced Computer Science	MSc	£9,500.00
Swansea University	Health Data Science	MSc	£6,250.00
Swansea University	Health Data Science	PGCert	£2,100.00
Swansea University	Health Data Science	PGDip	£4,150.00
Ulster University	Data Science	MSc	£5,290.00
University College London	Data Science (specialisation in Computer Science)	MSc	£11,800.00
University College London	Data Science (specialisation in Statistics)	MSc	£11,800.00
University College London	Data Science and Machine Learning	MSc	£11,800.00
University College London	Data Science for Research in Health & Biomed	MSc	£9,560.00
University College London	Web Science and Big Data Analytics	MSc	£11,800.00
Warwick University	Big Data and Digital Futures	MSc	£8,970.00
Warwick University	Big Data and Digital Futures	PGDip	£6,250.00
Warwick University	Big Data and Digital Futures	PGCert	£3,523.00
Warwick University	Data Analytics	MSc	£8,170.00
Warwick University	Data Science	BSc	£9,250.00
West England University	Data Science	MRes	£7,000.00
West Scotland University	Big Data	MSc	£3,400.00
Westminster University	Big Data Technologies	MSc	£9,500.00
Westminster University	Business Intelligence and Analytics	MSc	£9,500.00
Wolverhampton University	Data Science	BSc	£9,250.00

From this table it can be seen that UK universities provide:

- 111 masters (MSc) courses (EQF Level 7) on data science; and
- 13 bachelors (BSc) courses (EQF Level 6).

Typically, the fees for masters courses are €10,000 for one year and the fees for bachelors courses are €30,000 for three years.

The position for doctoral research (PhD) in the UK is more complicated. Generally doctoral student positions in the UK depend on entrepreneurial activity by the supervisor to fund the PhD studies, which costs about £75 K over three years including stipend and fees. The motivations of supervisors vary, but generally they are pursuing their own research programme branded under an umbrella project name. For example, there could be a Centre for Archaeological Forensics where an academic specialised in using modern data science methods to address historical criminal cases. In this case the PhD would probably go under the name of 'Archaeological Forensics' rather than 'Data Science'. Certainly many UK universities have centres of excellence in data science. Of 150 UK universities about one third have world-class research in most disciplines, about one third have world-class research in many disciplines and about a third have some or no world-class research. Of the first and second groups it is likely that their various departments graduate at least five to ten PhDs per year in subjects relating to data science, suggesting that each year about five hundred to one thousand people are trained in data science related areas to doctoral level.

Massive Open Online Courses (MOOCs)

Table 13 shows a number of Massive Open Online Courses available on the UK-based MOOC platform Future Learn [21].Typically such courses take 4 – 8 hours per week and last for 4 – 6 weeks, i.e. 16 – 48 study hours.

Table 13 - MOOCs offered by the UK-based FutureLearn MOOC Platform

Future Learn	Data Science
Future Learn	Machine Learning for Data Science
Future Learn	Various Courses in Data Science
FutureLearn	Learn to Code for Data Analysis
FutureLearn	Big Data Analytics
FutureLearn	Big Data Mathematical Modelling
FutureLearn	Big Data: Data Visualisation
FutureLearn	Big Data: Statistical Inference and Machine Learning

FutureLearn	Big Data: from Data to Decisions
FutureLearn	Introduction to R for Data Science
FutureLearn	Fundamentals of Data Science
FutureLearn	Data Science
FutureLearn	Data Science and Analytics
FutureLearn	Data Science

We have examined these MOOCs and studied some as students. Their quality varies but we can make the following observations:

- many of these MOOCs address a general audience and assume a relatively low level of prior knowledge and proceed at a leisurely pace
- many lack a serious professional approach where students are expected to follow the course to a tight schedule with well-defined and significant work to be done by the student
- some of the technical courses give poor instructions for downloading and using software causing students to lose a lot of time and often to drop out.
- the certification available is useful and can contribute to a user's CV. Some rigorous assessment is possible in examination conditions using Pearson's international network of examination centres. However, there are opportunities for new kinds of certification at programme level.
- MOOCs have great potential for communal learning that is not being fully used.
- MOOCs have great potential for peer evaluation and peer-mentoring.

The Open University partner has a lot of experience in the area of MOOCs and will use it for managing the activities of Phase 2 of the Da.Re. project.

Boot Camps

Commercial 'boot camps' provide more professionally oriented data science education than most MOOCs, but are expensive with courses lasting last a few weeks or months costing up to £15,000.

Boot camps offer more obviously hands-on training than academic courses and have potential attraction to employers by being focused on applications rather than theory.

Figure 35 shows that there are many boot camp offerings in the UK.

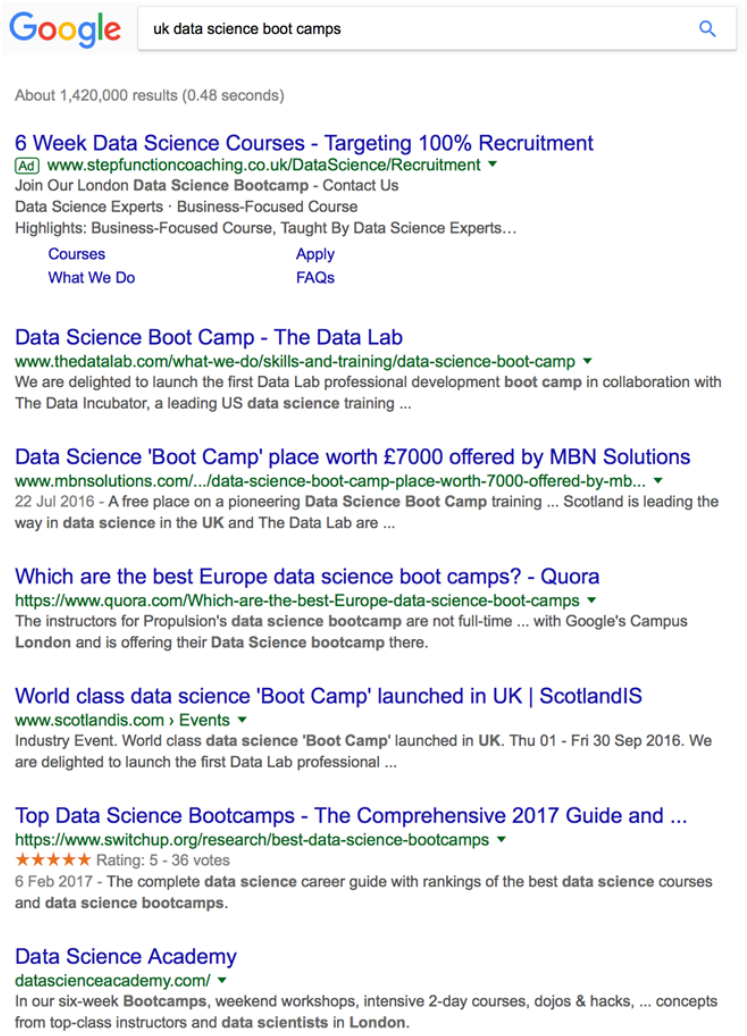


Figure 35 – A Google search for data science boot camps in the UK

3.4 Training Path Analysis in Portugal

In the following we summarize the results of the survey in the academic year 2016/2017 in Portugal. The source data, in the form of filled in templates, can be retrieved from Annex 1.

3.4.1 Presentation of Portuguese programmes

To perform the analysis of the existing training paths in Portugal, it was necessary to make an intensive research in the different education schools websites. With the gathered information, it was possible to create a dataset with the different training paths that exist in the Country and consequently build the following conclusions.

In a general overview of the existing training paths, see Figure 36, there are in total 25 training paths, being 3 of level 6 in the EQF, 20 of level 7 in the EQF and 2 of level 8 in the EQF scale. This means that the existing training paths are mainly focused on the Master level of education.

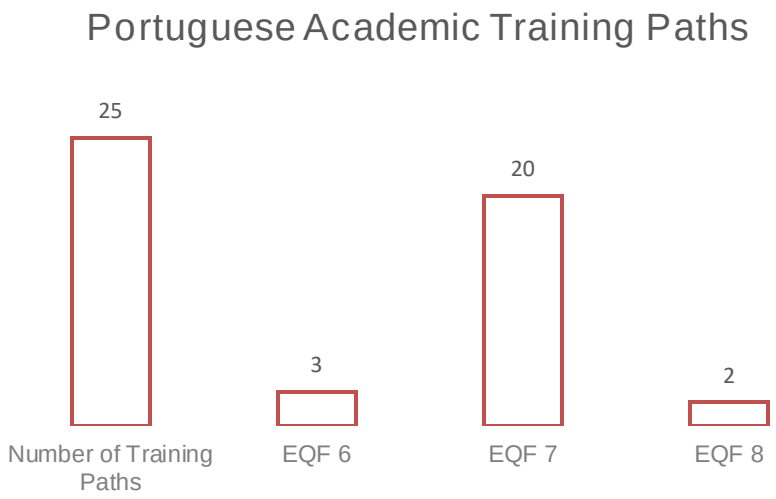


Figure 36 – General overview on existing Portuguese training paths

Bachelor Training Paths

In terms of Bachelor training paths, in Figure 37, there are two types, specialization courses and courses – curriculum. From the three bachelor training paths one is a specialization course and two are courses – curriculum.

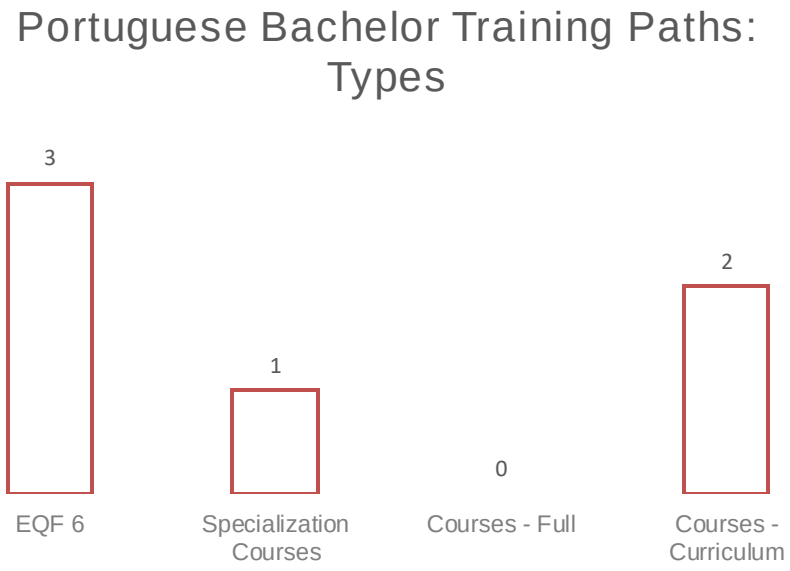


Figure 37 – Bachelor training path in Portugal

Master Training Paths

In Figure 38 we reported about the Master training paths that exists in Portugal: of the existing twenty, six are specialization courses and fourteen are courses curriculum.

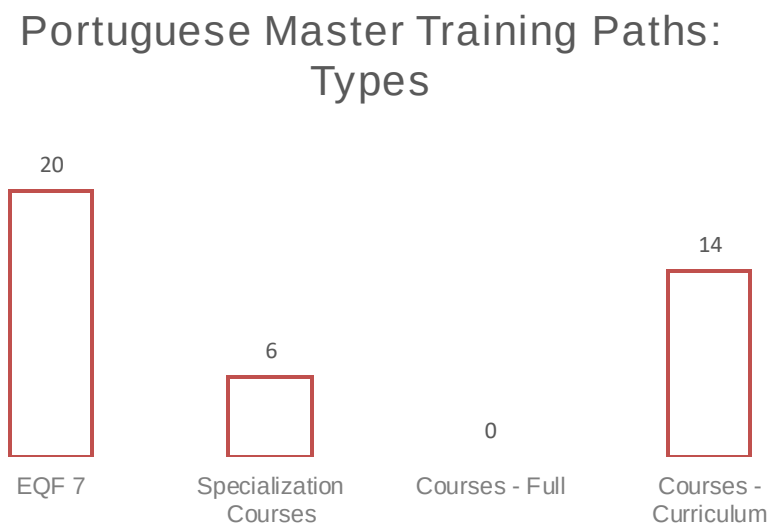


Figure 38 – Master training paths in Portugal

PhD Training Paths

As it is reported in Figure 39, there are two PhD training paths and those are specialization courses.

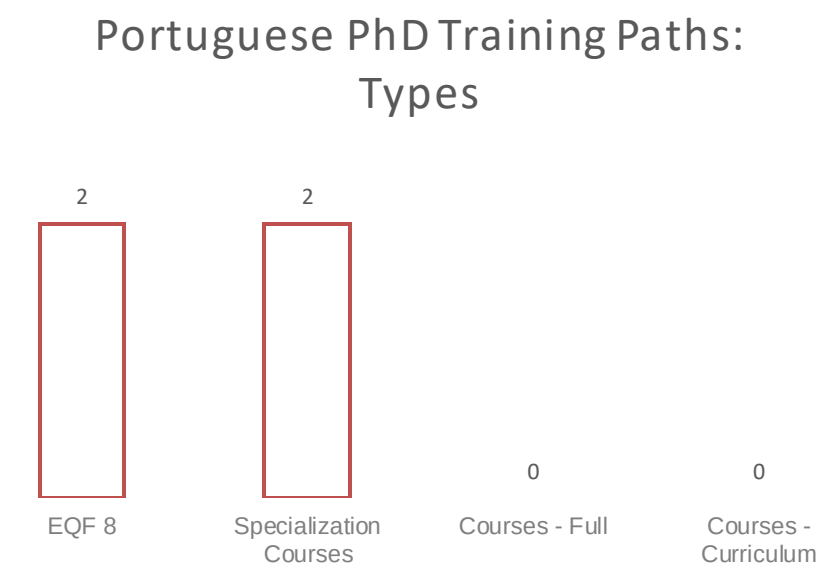


Figure 39 – Ph.D. training paths

3.4.2 Training paths by discipline

The different training paths are included in several areas, such as, Computer Science, Mathematics and Engineering, among others. In Figure 40 is possible to concluded that the area that covers most of the training paths is the Computer Science (72%).

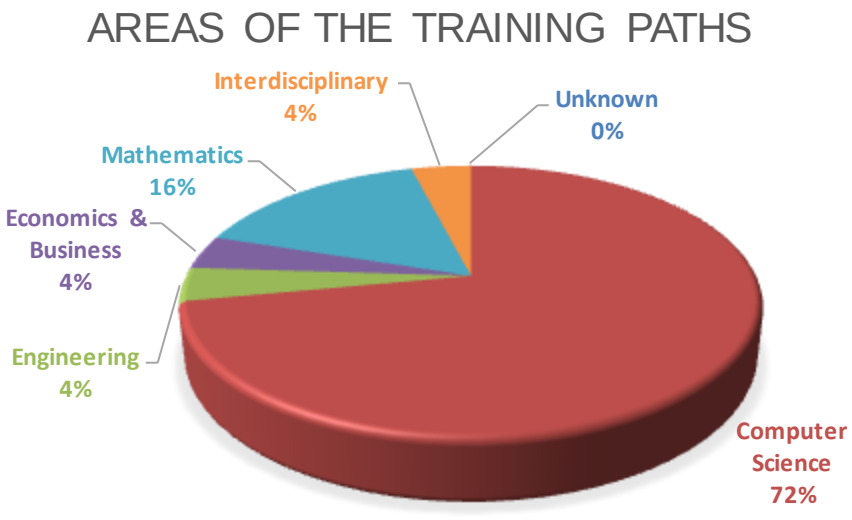


Figure 40 – The main Academic areas involving in organizing the Portuguese training paths

3.5 Training Path Analysis in Slovenia

In the following we summarize the results of the survey in the academic year 2016/2017 in Slovenia. The source data, in the form of filled in templates, can be retrieved from Annex 1.

3.5.1 Presentation of Slovene Programmes

This analysis of the training paths in Slovenia includes all the state accredited training paths of EQF levels 6, 7 and 8. There is a total of 309 training paths in the academic year 2016/17 that are held by the 3 state funded universities (University of Ljubljana, University of Maribor, University of Primorska), the private University of Nova Gorica and some "stand-alone" faculties.

The geographical distribution of these courses is depicted in Figure 41 :

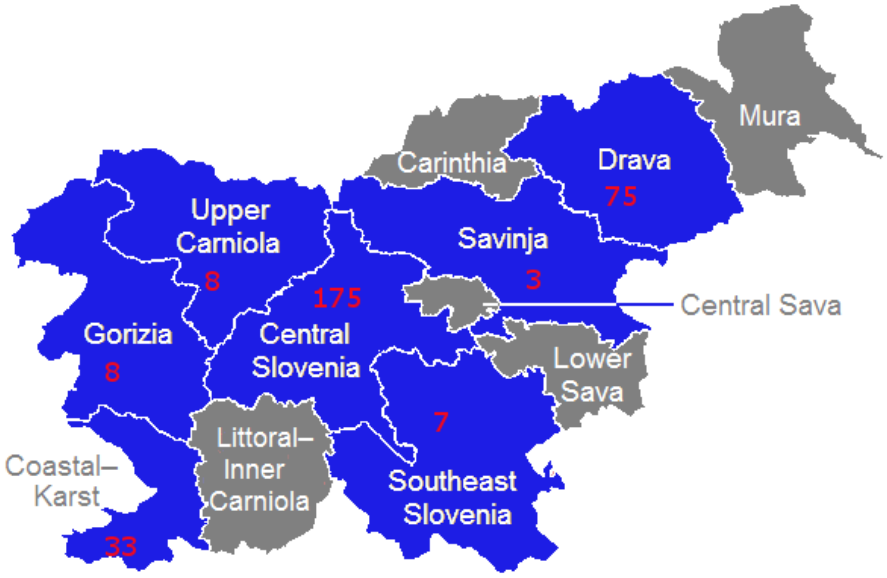


Figure 41 – The map showing the distribution of training path in Slovenia

Out of the total 309 training paths, 41 have at least some elements of Data Science, meaning that at least 10% of the courses that make up the training path can be considered Data Science courses. Hence, this analysis concentrates on those 41 training paths. Figure 42 represents a general overview of the 41 "Data Science" training paths by EQF level.

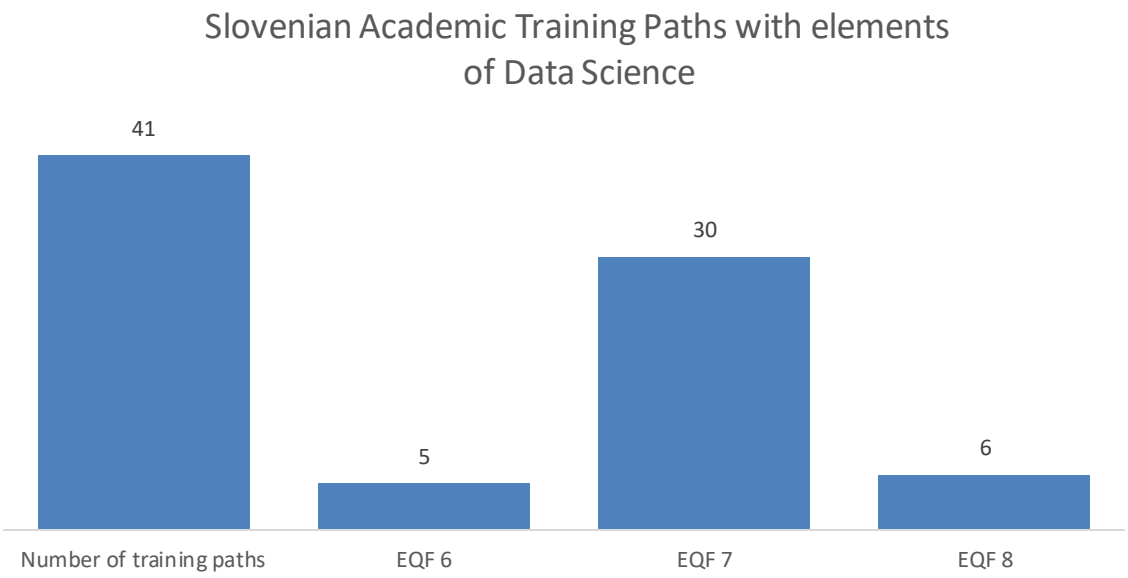


Figure 42 – General overview of Slovenian training paths

The vast majority (30 courses – 73%) of the training paths (with elements of Data Science) in Slovenia is of EQF level 7. This fact is not surprising since all bachelor and master courses fall in this category. There are 5 (or 12%) of the training paths that are EQF level 6 – these are specialization courses, and 6 (or 15%) of EQF level 8 training paths – the PhD courses.

3.5.2 Training paths by discipline

Analysing the Slovenian training paths by discipline (see Figure 43) we can see that 15 (37%) of them are Computer Science courses, 15 (37%) are interdisciplinary (but these courses are all Computer Science with either mathematics, electro technics or engineering as the second discipline), 5 (12%) are engineering courses, 5 (12%) are electro technics courses and there is 1 (2%) course for which we could not determine the discipline.

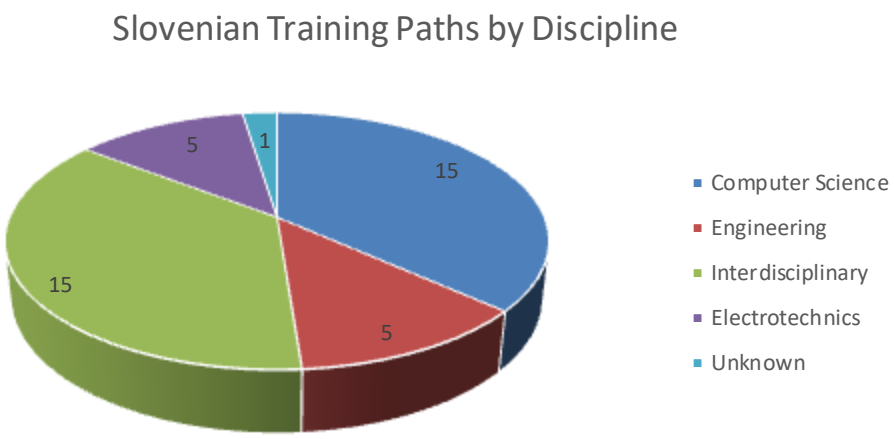


Figure 43 – The main Academic areas involving in organizing the Slovenian training paths

By combining the "EQF level" and the "by discipline" figures we get the Figure 44 that represents the 41 training paths by discipline and EQF level.

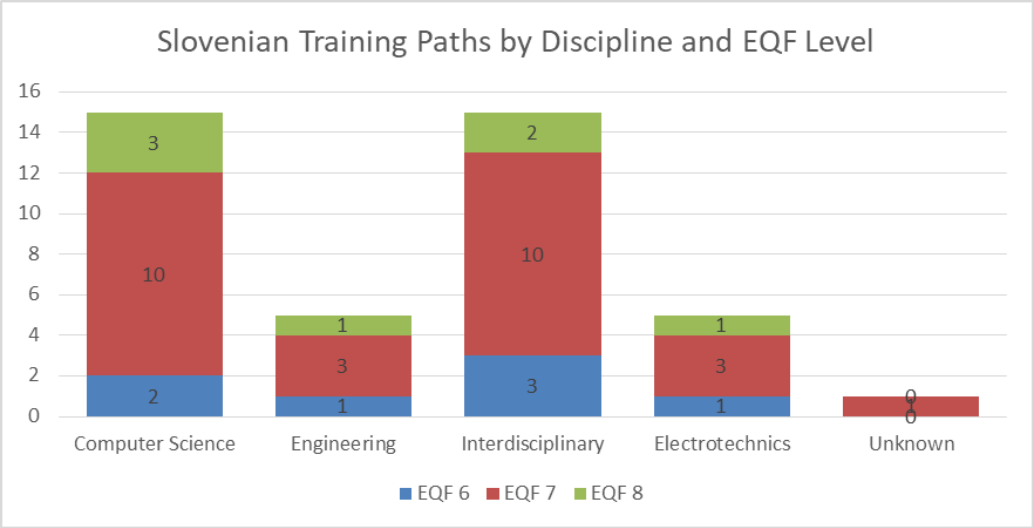


Figure 44 – Combination of Discipline and EQF levels

From this analysis, we can conclude that for Slovenian training paths with Data Science elements there is a predominance of computer science courses, either "pure" or interdisciplinary, followed by engineering and electro technics.

3.5.3 Costs analysis

In Slovenia, all Bachelor and Master study programmes are free of charge for students that decide to enrol "full-time". The cost for PhD programmes ranges from 2,500 to 5,000 EUR per study year.

3.6 **Training Path Analysis in Serbia**

In this subsection, we provide the analysis of the situation in the higher education regarding Data Science in Serbia.

The analysis was based on the:

- analysis of the offering of the higher education institutions in Serbia,
- material collected in the survey reported in Chapter 4 and
- various discussions with the professors responsible for the topics related to Data Science.

As in the case of the market need analysis in Serbia, presented in Section 2.5, the situation with the higher education is complex, since many activities are on-going and it is not clear what will be the best models for the education.

3.6.1 Presentation of Serbian programmes

Three ways of educating Data Scientists in Serbia:

- Within specialized CS courses
- Within courses in Applied mathematics
- Within courses in economy and management

Largest universities in Serbia support all three ways. Courses specialized for Data Science are still immature.

Our analysis of the existing offering showed that there are three institutions offering courses dedicated to Data Science, two of them belong to state universities (Faculty of Sciences, University of Novi Sad, Faculty of Technical Sciences, University of Novi Sad) and one to the private university (Faculty of Technical Sciences, Singidunum University).

In the following text we provide details about the programs:

- Faculty of Sciences, University of Novi Sad (state)
  - Master program “Applied Mathematics – Data Science“
    - https://www.pmf.uns.ac.rs/studije/studijski-programi/primenjena-matematika-nauka-o-podacima/
    - 1st and 2nd semester are common

- there are a lot of topics in mathematics (advanced)
- 3rd semester specialization in:
    - Data Analytics, Machine learning, Optimization, Signal Processing
    - HPC, Big Data

- Faculty of Technical Sciences, Singidunum University (private)
    - Bachelor program Software and Data Engineering – Data Science
        - http://tf.singidunum.ac.rs/
        - 1st and 2nd year Software Engineering
        - 3rd and 4th years, specialized:
            - Artificial Intelligence
            - Applied Probability and Statistics
            - Mobile Application Development
            - Introduction to Machine Learning
            - Data Visualization
            - Deep Learning with GPU programming
            - Mining Social Networks
            - Doing Business with Data Science
            - Mining of Massive Datasets

- Faculty of Technical Sciences, University of Novi Sad (state)
    - Bachelor program: Information Engineering, 4 years
        - http://www.ftn.uns.ac.rs/n1554983877/informacioni-inzenjering
    - Master program: Information and Analytical Engineering, master, 1,5 year
        - (http://www.ftn.uns.ac.rs/n479272485/informacioni-i-analiticki-inzenjering) i

- Master program: Information Engineering, master, 1 year
    - (http://www.ftn.uns.ac.rs/17606715/informacioni-inzenjering)

There are also two courses which are not dedicated to data science, but they contain lectures/topics related to data science and can be used as additional material for the education. They are:

- Faculty of Organisational Science (FON), University of Belgrade
    - Information Systems and Technologies
        - http://www.fon.bg.ac.rs/studije/master-i-specijalisticke-studije/master-studije/informacioni-sistemi-i-tehnologije/
- Faculty of Mathematics, University of Belgrade
    - http://www.matf.bg.ac.rs/cp/7/izborni-modul-racunarstvo-i-informatika/

Offering in the higher education in the domain of Data Science is in the early phase of the development. Only a couple of the courses dedicated to the data science has started recently. It is still unclear how they will close the gap between the demand (from industry) and the supply (from the universities). The problem is that the students are not completely clear about the value of the courses dedicated to the data science, comparing to general computer science courses, which give a broader IT knowledge. This is one of the very important issues to be clarified.

We argue also that the opportunities for the education of Data Scientist in this way in Serbia will go beyond the borders and will be well received in the Western Balkans. Figure 45 illustrates the opinion of the IT companies about the development of the data scientists comparing to the countries in the region. Around 60% are not sure about the answer, whereas “only” 8.82% thinks that the education in Serbia is better than in the other countries in the region.

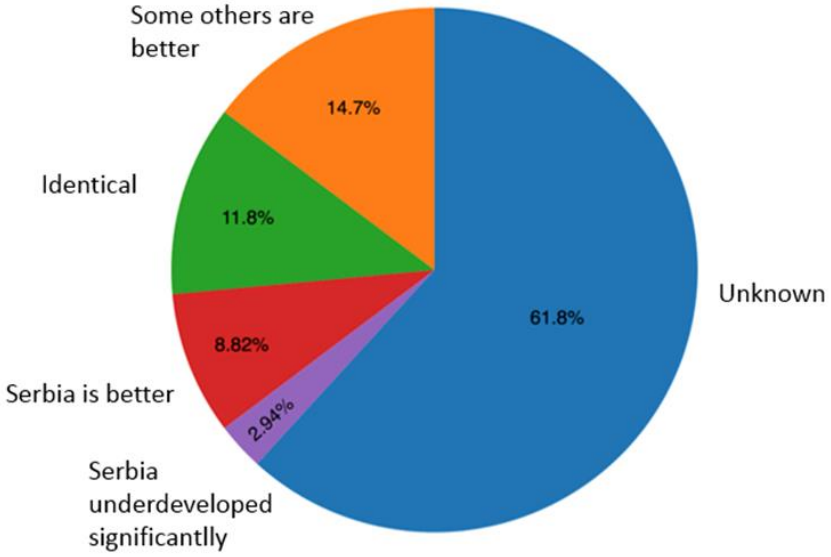


Figure 45 – Development Serbia with respect to the region

This analysis has shown that the industry need for the high professional personal cannot be so easily satisfied with the current efforts in the higher education. The main issue is the need for specific knowledge in data science, which cannot be obtained “only” by transforming some of the existing courses (especially master ones) in the data-science flavoured courses. The most important reason is the need for an industry involvement and support in order to ensure that the knowledge and skills obtained in the lectures will be applicable in an industry setting.

Chapter 4. Training Needs Analysis and Participatory Events Results

This chapter reports on the results of Activity 1.3. The Da.Re. project aims at raising the attention of the stakeholders in order to improve the impact and the results of the activities thanks to their participation.

The main target for the partners was to draft a list of needs to be satisfied with the organization of the Pilot Course for Data Scientist using both the online research and the direct involvement of the members of the community to be created around the Da.Re. initiatives.

The creation of a group of stakeholders started from a phase of study of the environment related to the main theme of the project: **Data Science**.

The world of data can be studied using two different points of view: the **Academia** and the sector of **Private Organizations**. The Science of Data is developing quickly, in the last five years has become a crucial topic in every business activity but at the same time it is a wide theme that requests a deep evaluation of the training methodologies and the outputs to be obtained.

For this reason, the Consortium decided to manage the training need analysis dividing the stakeholders in two main groups, targeting **Professors and Students** for the Academia, **Big Companies and SMEs** for the Private Organizations.

All the five countries involved in the project adopted a specific approach, according to the results obtained with the research made on the Training Paths and the Market Needs already illustrated in the previous chapters.

The approach chosen by the different Partners has been shared with the whole Consortium in order to optimize the actions and the effect of the dissemination activities, while the different results have been used for making the best conclusions to be used for shaping the Da.Re. pilot course.

In Italy, the presence of Confindustria helped orienting the activity to the Private Sector. The Portuguese Partners focused on the involvement of the Academia, while the Partners in Slovenia, Serbia and UK differentiated their approach according to their specific environment and available networks.

In the following paragraphs is contained a description of the specific actions implemented in the different countries.

4.1 Training Needs Analysis in Italy

Following the results obtained with the desk research on the profile of the data scientists' skills in Italy, **the Italian Partners focused their activities on the involvement of the SMEs**, the only category with a lack of information to be collected through the web.

The category of SMEs is the most widespread in Europe and especially in Italy SMEs almost represents the total amount of companies [24].

With the aim of collecting the highest quantity of information regarding the training needs to be addressed by the Pilot Course, Loccioni, University of Camerino and Confindustria worked for creating different opportunities for the involvement of small companies.

In particular, Confindustria Marche Nord gave to the Partners the opportunity to organize **two different events**, the first took place on May 26th 2017 in the content of the Confindustria Market Place Event, the second, the first official Italian Participatory Event of the Da.Re. Project, took place on the 4th of July 2017 at Confindustria, in Ancona.

Marketplace Day

Figure 46 – Website homepage of Maketplace Day

The Marketplace Day, Figure 46, is the main networking event in the Marche Region, with more than 300 companies, 90 startups and other innovation players such as Universities. The Marketplace Day wants to be a hub of relationships and promoter of the culture of innovation, as witnessed by the conference on Industry 4.0 and Open Innovation organized during the event.

The event represented a great opportunity to highlight the activities of Da.Re. international consortium: in the central part of the building dedicated to the event, in the IoT thematic area, entrepreneurs have been invited to find out more about the role of Data Scientist and the benefits that the company can derive from their involvement in the business, having the opportunity to meet some of them personally to satisfy doubts and questions.

The University of Camerino used its stand to tell how companies can transform data into strategic value for their business.

The knowledge provided by Confindustria regarding the companies invited to the event allowed the Italian Partner to better organize a strategy for the involvement of the potential participants.

This first direct approach to stakeholders has been considered as a good moment to be used for **evaluating the level of awareness of the Italian small companies on the potential of the science of data**: the partners organized a survey to be administrated both online and on paper in order to follow this purpose.

The interested participants had the possibility to use the following link:

https://docs.google.com/forms/d/e/1FAIpQLSebI9KPOLJtCf5dt-GbWI4WNZYnqw0kHqFhVLW2n21PtUW5Ng/viewform?usp=sf_link

The link could be used with the tools offered by the University of Camerino (Laptops, Tablets, and Smartphones) or with any other kind of device connected to the web.

The survey has been created to classify the participants according to their level of knowledge on the Science of Data.

The questionnaire has been proposed to the participants in the Italian language.

The first stage of the **survey** asks for the email of the participant and for the following 5 entries to be selected from a multiple choice:

- **Sector** of the participant
- **Kind of data available** in the participant company
- Which kind of **purpose** is linked to the collection of data in the participant company
- Who is the **person in charge of collecting the information** in the participant company
- Who is the **person in charge of analysing the data** in the participant company

The second stage contains 2 more detailed questions:

- The request to have an **example** of the case in which the participant company analyzes its data
- A specific declaration on the purpose aimed with the data analysis made in the participant company (**reporting vs prediction**)

The conclusion of the survey gives a classification of the participant divided in 3 different **categories of Data Scientist**:

- Beginner
- Intermediate
- Expert

Figure 47 – The final page of the survey

Each of the participants involved in the survey during the Market Place Day obtained a card with the result achieved with its participation to the questionnaire, Figure 47, and the invitation to connect to the Da.Re. Community with the participation to the Participatory Event. See Figure 48 for an example of the card.

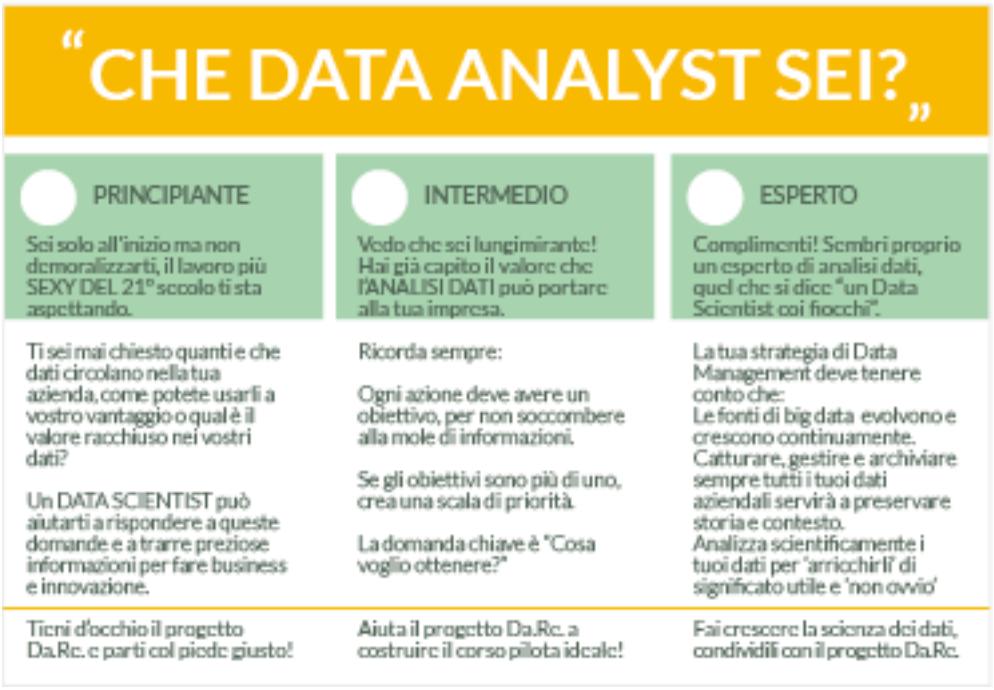


Figure 48 - The invitation to the Da.Re. Participatory Event

The link to the survey is still open, it can be used to increase the number of participants and it can still be considered as a valid dissemination tool.

Until now, 13 people fill the survey. The survey considers as beginner, who has still not a specific method or team to collect data in the company and where data are usually analysed not by a

dedicated team or a technician but is checked only by the entrepreneur. The intermediate profile instead, has already awareness about data collection and analysis but data are investigated to obtain a report of the current situation of business processes and / or functional areas and decide on operational / strategic solutions based on the analyzes carried out. Finally, the expert profile works in a company where there is a specific team of data scientist who analyses data to be able to forecast future behaviors or results and that the proposed action according to the results of the analyzes carried out would be implemented autonomously. From Figure 49, the 38% of profiles outlined are beginners, the 31% are intermediate and expert. Moreover, from the survey it emerged that, refer to Figure 50, data collected are analysed for the 39% by the entrepreneur, for the 23% by an administrative profile and for 38% by a data scientist.

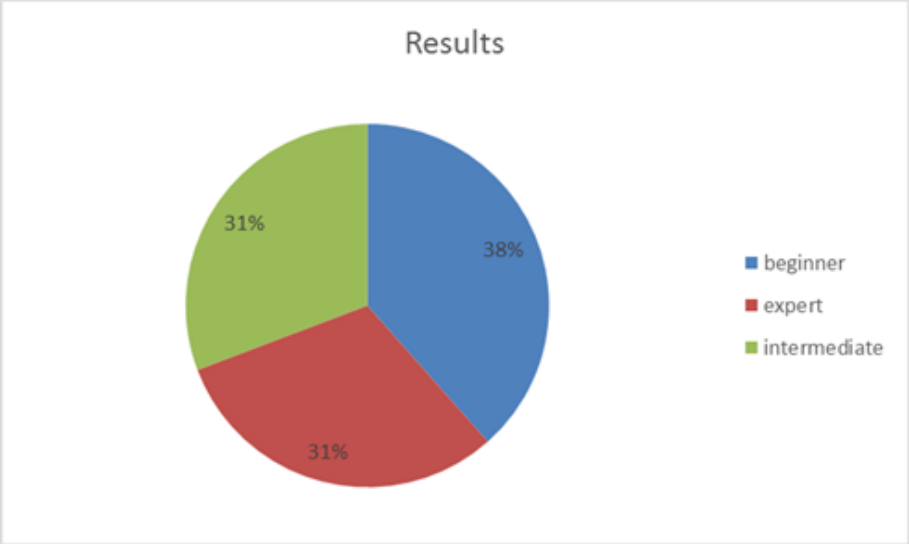


Figure 49 – General results

Figure 50 – Who takes care of your data?

From Figure 51, the most relevant commodity sectors in the area where the survey is submitted are business services for enterprises, ICT and Fashion companies with respect to Advanced equipment and technologies companies, services to people and communities, education, measuring systems for quality control. Moreover, as it reported in Figure 52, the 54% of data collected are related to production, the 31% to commercial data and 15% to administrative information.

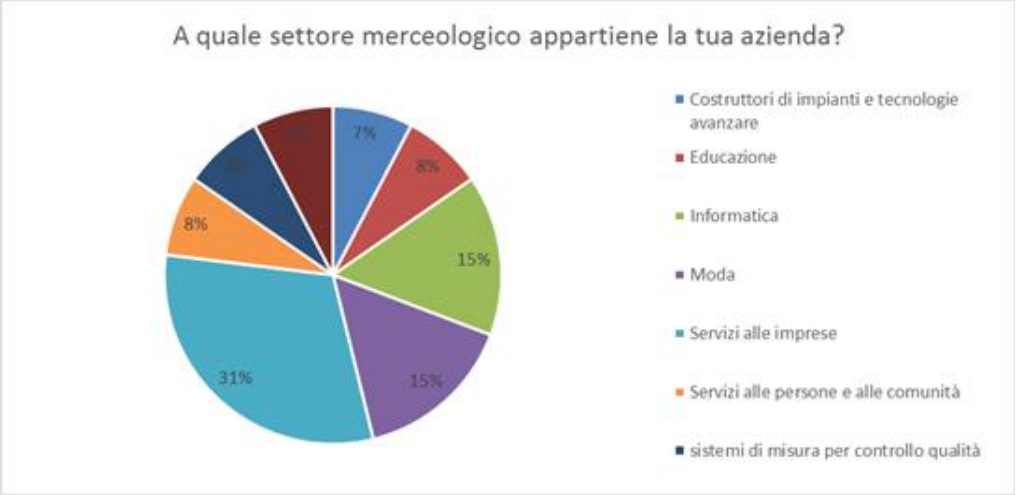


Figure 51 – Commodity sectors

Figure 52 – What kinds of data do you have?

Figure 53 – What do you use your data for?

Finally, in Figure 53, we can see that the 69% of data collected are analyzed to strategic decision making.

According to the feedbacks obtained by the representatives of the University of Camerino the event was big, with a high number of companies from a very large set of sectors.

The number of people that accepted to fill in the form was relatively little with respect to the people present.

Probably, the main reason for the result was the kind of interactions that could be established. It was difficult to focus on the project topics and to attract the attention of the potential audience. The general aim of the project could be explained but a very few guests asked for deepening the information with the presentation already prepared.

The first trial represented by the participation to the Market Place underlined the early stage of the activities reached by the Da.Re. initiative compared to the necessity to give information to the potential stakeholders.

The opportunity of the Marketplace will be probably taken in more advanced steps of the project.

Figure 54 – A moment of the discussion during the participatory event

Participatory Event

The Participatory Event, in Figure 54, focusing on the business opportunities that data analysis can open to companies, has been designed to respond to certain **features**:

- To have participants geographically located in the territory of the **Marche Region**;
- To have a small number of participants who can be close to the themes of data analysis and their use, so as to **stimulate** their **direct involvement as stakeholder**;
- To organize a brief event that could provide **practical examples** to participants, but also useful stimuli to solicit interaction with project partners.

Through the support of Confindustria, Loccioni and University of Camerino, 30 companies selected to provide a practical feedback to Da.Re's goals, have been identified.

The workshop lasted for 2 hours and the participants received information about real case studies in which data were used for the valorization of business activities, in different areas and in different ways and quantities.

The event provided a short speech dedicated to the **dissemination** of Da.Re.'s themes, a special **storytelling experience** based on Talete (identified as the first data scientist in history) and the **practical case studies** proposed by Vittorio Di Tomaso (H-Farm) and Daniela Isidori (Loccioni).

The Storytelling has introduced to the participants the topic of data analysis starting from Talete. His ability to observe reality and to link apparently unrelated information, guaranteed him wealth and success in business. Then, the audience had the opportunity to understand how it is possible to move from data to knowledge, through the story of case studies with different amounts of data and applied to completely different worlds, from automotive to large-scale retail trade, through the human health and energy efficiency of smart grids.

Finally, Vittorio Di Tomaso made his experience available to coordinate the works of the workshop, involving the audience to talk about the future and solicit public interaction.

It was possible to find that **higher quantity of data means higher quantity of predictions**.

17 Companies and 8 Students attended the meeting and gave to the project partners the possibility to make some new deductions:

- There is a **little awareness** by small companies on how data could help improve their business
- There are **difficulties for entrepreneurs** to identify in which areas of their business data analysis can be useful
- There is the willingness to participate to the discussion on data scientists' training paths but not yet the capability to help with the creation of the modules of the course
- There is the **need of training** for entrepreneurs
- The possibility to involve companies specialized on "big data" in support of SMEs is seen as a good opportunity

The Participants highlighted a wide variety of situations and levels of expertise on the Data Science, in particular the need to dedicate the project not only to those who present an advanced level of knowledge and awareness on data potentials, but also to those just at the beginning of the pathway.

Through the Participatory Event, Da.Re. partners started a relationship that will continue in the coming months, online and face-to-face.

4.2 Training Needs Analysis in UK

In the UK the requirement for data scientists trained to masters level and beyond **appears to be satisfied** by the many university courses that are currently offered. However **we have identified a gap** in the education and training provision, **namely that of the "bridge person"**.

Let a data science bridge person be defined to be someone who understands the market or public-facing mission of an organization, and who also understands enough of the theory and practice of data science to communicate with highly trained and/or highly experienced specialists.

Given the added value that data science can bring to organizations, bridge persons could help create a lot of added value for the 95% of micro-businesses in the UK.

Bridge Persons in large and small organizations

During this research, a meeting was held with members of a data science department embedded in a marketing organization. The interview took place with the data science director and a business manager responsible for customer-facing marketing projects within the company. From the business manager's perspective, their role consists of working with the client to elicit requirements and work up possible marketing campaigns, both responding to the client's ideas and bringing in ideas from their organization. Once the outline campaign has been agreed it is the role of the business manager to coordinate the various specialized services, as depicted in Figure 55, within the organization to produce an integrated service for the client. In this case, the specialized services could include making and placing advertisements with A – television, B - print media, C – social media, and so on.

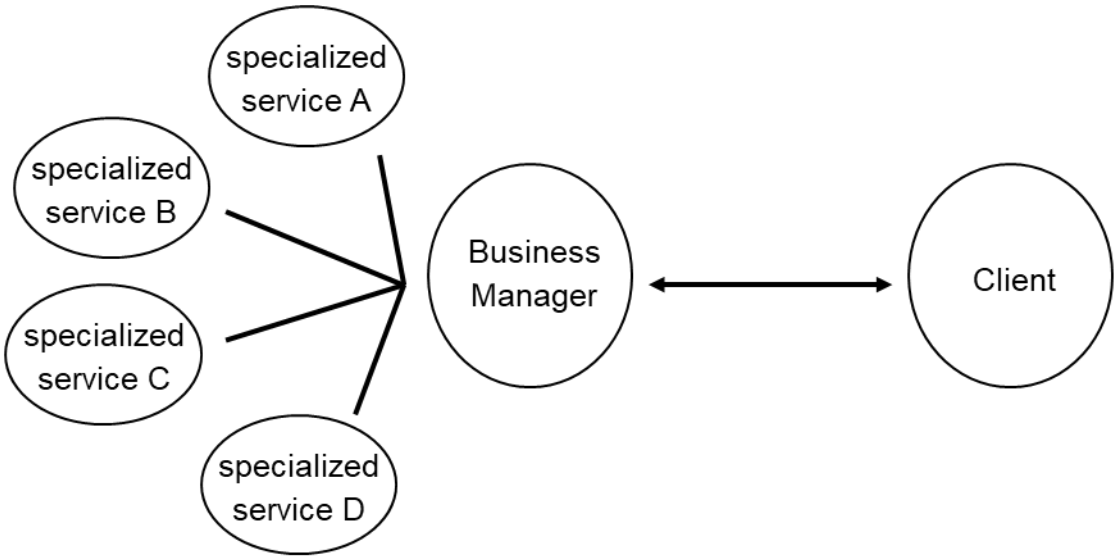


Figure 55 – Bridge Person

Analytics provide an essential service for marketing campaigns, allowing the customer to evaluate the effectiveness of the campaign and the added value it has created. Clients will have their own data sources and, sometimes, their own in-house analytics capabilities.

Generally clients will want to know if marketing campaigns have been 'successful' where some part of this relates to quantified data such as increased footfall, increased revenue, and so on.

Some aspects of success may be more qualitative, such as the perceived ethical status or trustworthiness of the organization. For example, the recent Volkswagen emissions scandal has damaged the brand and, no doubt, a great effort is underway to restore public trust – a concept that may depend on a range of indicators.

A question asked at this interview was to what extent data analytics is fundamental during the design of marketing campaigns, and to what extent it comes later. For example, knowledge of the many kinds of data and analytic techniques available could lead to campaigns designed around what is measureable, rather than measurements being made on a more constrained set of possibilities after the campaign is designed.

During the interview it appeared that within this organization it is known that embedding data analytics in the design of campaigns is the ideal, in practice the implementation is patchy. This reflects the **inevitable gap between the "communication planners"** who provide the flair and imagination essential to designing innovative and successful campaigns that engage the target audience, **and the "analysts"** who use technical skills to collect and process data.
Anecdotally, the business manager volunteered that she would like greater knowledge of what is possible in data science to make her campaigns more productive and more useful for the client.

Bridging the gaps in small and large organizations

It is conceivable that large organizations also have communication gaps between market-facing parts of their organizations and in-house data science capability, when it exists. Thus, large organizations may also have need for a data science bridge person.

Thus, **the hypothesis is that data science bridge persons are required by both small and large businesses and organizations**, as illustrated in Figure 56.

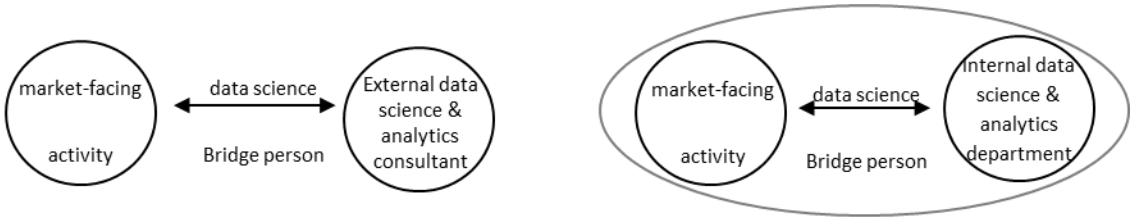


Figure 56 - On the left, no internal data science capability while, on the right, in-house data science capability

4.3 **Training Needs Analysis in Portugal**

With the goal of promoting and disseminating the Da.Re project, the Portuguese participatory event was conducted. Furthermore, and since all the interested stakeholders were present at the event, one of the major expected outcomes from the event was also to get feedback about what would be the course expectations from the participants perspective.

The Portuguese partners, namely IPB and MAISIS, developed the participatory event on the 29th of May of 2017 whose name was “Re-imaginar a educação na ciência dos dados”. For this event, an invitation brochure was developed, see Figure 57, where key information was depicted namely the session participants panel. This invitation was sent to both academia and industry using the internal mailing lists.

Figure 57 – Portuguese Participatory event invitation

The session started by Paulo Leitão (IPB) giving a short introduction of the Da.Re project, followed by a round table discussion. This round table was composed by 4 persons each one having a different perspective from the market needs and personal/domain expectations. Therefore, a professor whose work is closely related with data science, an engineer from a software company

that offers data science solutions to their clients, a company CEO that is in need to have more knowledge from its internal data and a PhD student have composed the round table.

Some photos, depicting the Portuguese participatory can be seen in Figure 58,Figure 59 Figure 60.

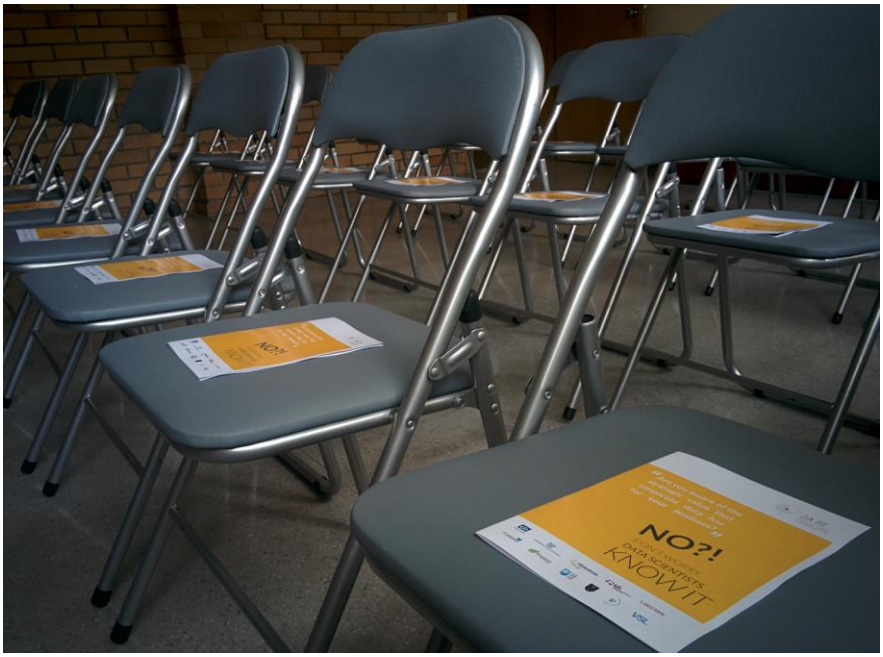


Figure 58 – Participatory event preparation

Figure 59 – Participatory event round table

Figure 60 – Participatory event round-up and final discussions

From the discussion, it is possible to conclude, in a summarized way, several important factors:

- The need to make analysis on data is present at our daily life and everybody is doing this (although may times without being aware of this);
- There's no solution "one fits all" and people may find different needs from their personal experience and/or needs;
- "Old fashion" software tools could play a valuable solution in the data science domain, e.g., Python or R;
- Companies are in urgent need for collecting knowledge from the collected data;
- Industry is aware that this will help them to better improve their production processes, e.g. improve the maintenance procedures and forecasts;
- Undergraduate students are not very much aware of the "data science";
- On the other hand, graduated students are aware of the importance of having skills in this domain;
- Both agree that a more practical and hands-on course would produce better results as also would catch more their attention.

As a general conclusion, students might not be aware of the importance of data science, at least, until the moment they are introduced to the topic. On the other side, companies are aware of the importance of getting valuable knowledge out of their data, which may imply a near future market

need explosion. On the offer side, both universities and software companies starting to offer solutions compatible with data science, although not at the desired pace.

4.4 Training Needs Analysis in Slovenia and Serbia

As a result of the first research activities made in the different Countries we concluded that Slovenia and Serbia represent the countries in which **Data Science is still a very narrow theme**. It is **not yet considered as an independent topic** but as a set of skills that must be collected from different training paths, such as computer science, mathematics, economics and management.

The private companies and universities interviewed were selected from a very small and specific group who helped to complete the research illustrated in the previous chapters.

After sharing the first experiences collected in the other countries Da.Re. Consortium decided to **differentiate the approach in Serbia and Slovenia postponing the direct involvement** of the wider part **of the stakeholders**.

The idea is to **organize different events during the second phase of the project**, dedicated to the creation of the Pilot Course: the feedbacks of the stakeholders will be used to refine the modules created and the approach given to the course.

However, the Nissatech partner, thanks to its network and experience developed in the specific field of data science, had the opportunity to interview a few organisations in Serbia that are already partially active in that field. In the following we report about this survey.

Data Science Development Directions in Serbia

A goal of this analysis is to indicate possible development directions of academic education in the area of data science, as well as advantages that can be deployed in the academic education system of Serbia.

The analysis is based on:

- the recent survey among managers of the Serbian IT companies [25]
- the communication with different industries and
- own experience in big data.

The analysis showed that although the need for well-educated data scientists in the traditional industry is low, the need for consuming data analytics services is increasing (as a part of the global trend). This is the reason why this analysis is mainly focused on the need for data analytics expertise of companies that are providing different (mainly IT) services.

Therefore, the analysis is related to the current state and needs of Serbian software companies for knowledge and professionals from the data science area. In the following text we provide the main findings from the analysis.

The analysis is based on the information collected from the companies active in IT services provision, structured as presented in Figure 61. Most of the companies are software development companies (31.5%), followed by software consulting companies (20.5%)

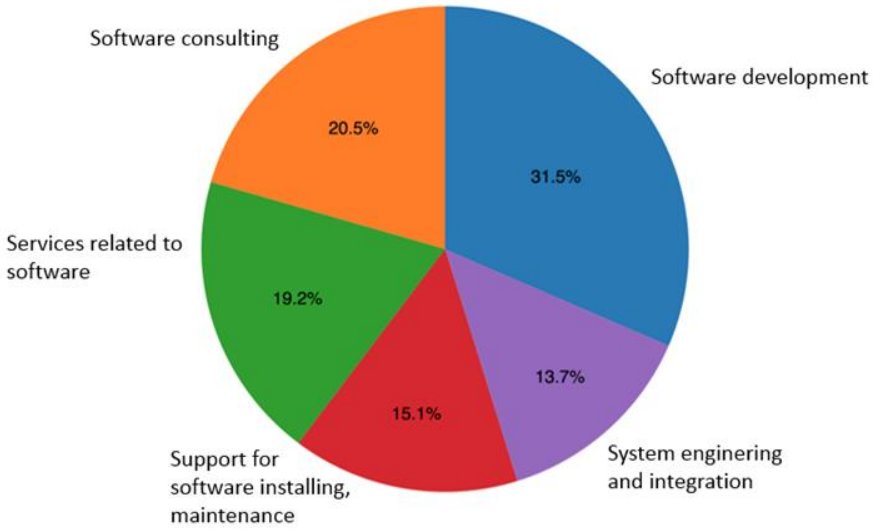


Figure 61 - Domains, which the interviewed companies are active in

Figure 62 summarizes the need for IT professionals in general in the selected companies. Main conclusion is that software developers (general and specialized) are missing (22.9% and 21.9%). Data Scientists are required in 10.5%.

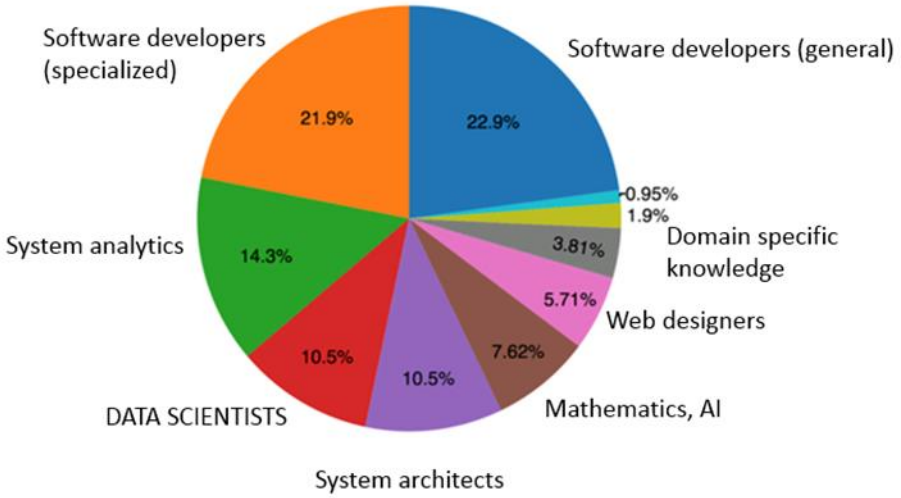


Figure 62 - Need for IT professionals

Figure 63 illustrates the status of the implementation of the Big Data activities. It is interesting that 26.5% + 20.6% companies already realized or have started Big Data activities and 20.8% have planned to do it.

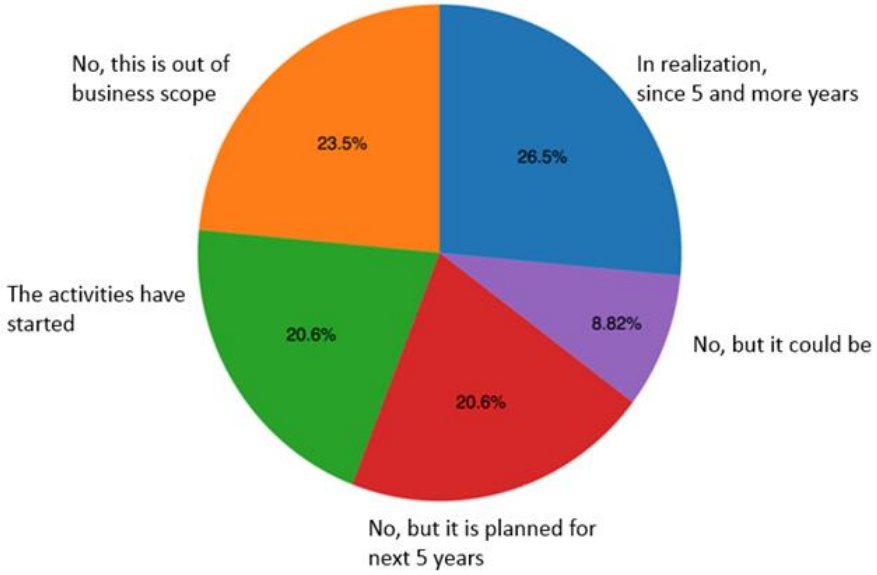


Figure 63 - Activities in Data Science / Big Data

Figure 64 illustrates the situation regarding the availability of human resources for Data Science. In 28.5% companies such resources exist already, in 32.4% the development of human resources is in progress.

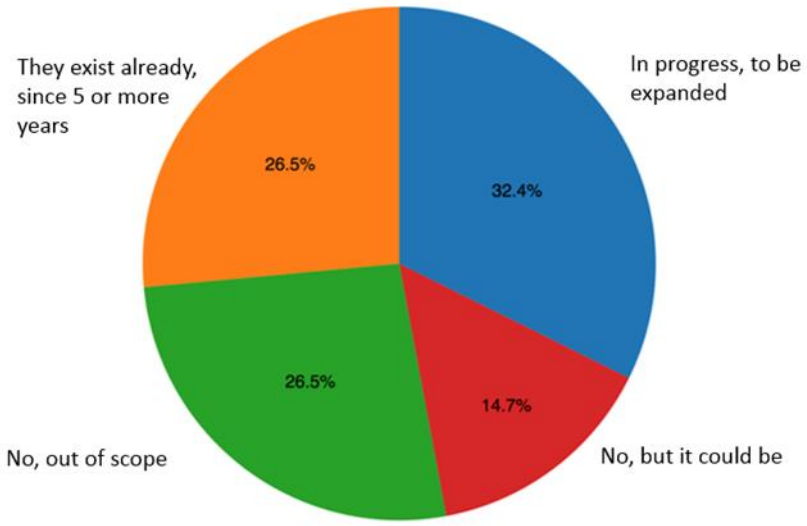


Figure 64 - Human resources in Data Science

Figure 65 illustrates the plans for acquiring new personal for Data Science. Around 30% of the companies is clear about the need/strategy, but not about the amount. Another 30% is clear about the amount (50, 20, 5).

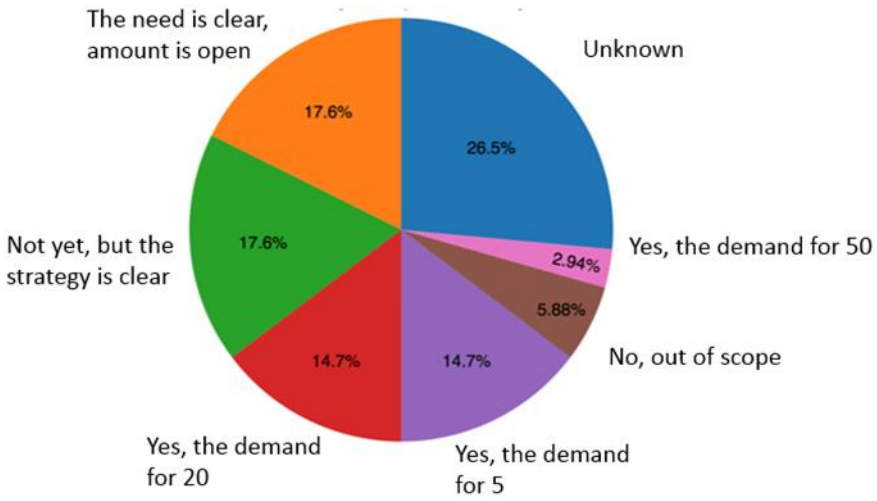


Figure 65 - New personal for Data Science (plan)

Figure 66 reflects the problem related to the availability of the professional personal. Almost 30% of companies thinks that such personal will be available in a couple of years, whereas 20.6% thinks that the demand for professional personal is much bigger than offer.

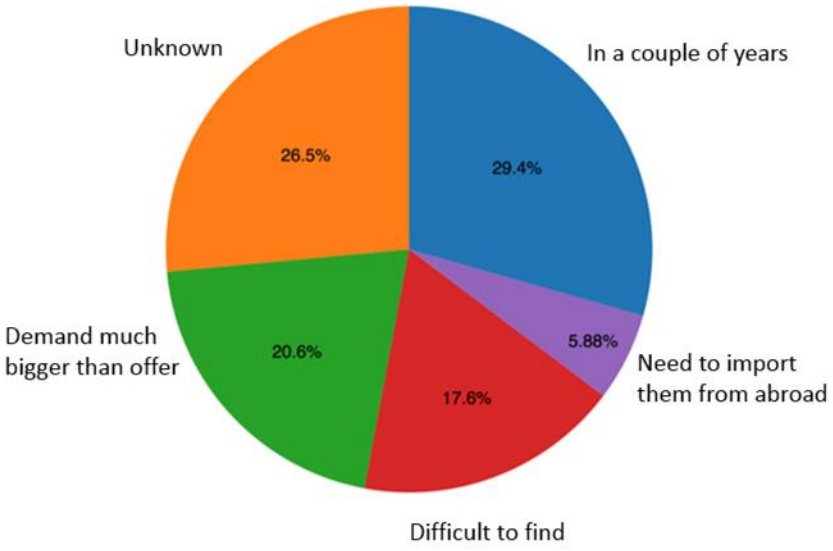


Figure 66 - Availability of professional personal

Figure 67 explains the distribution of the application domains. The most widely used is the decision making domain (18%), followed by the financial domain, insurance, banking (17.1%) and telecommunication, internet (15.3%).

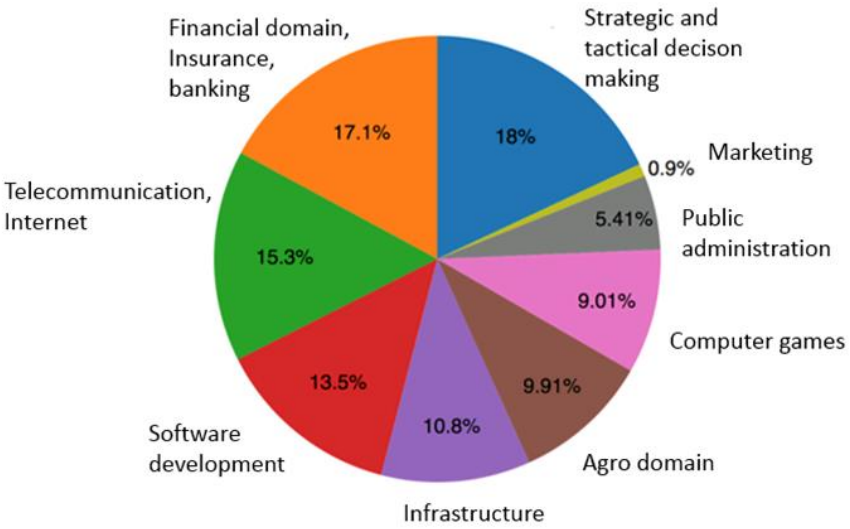


Figure 67 - Application domains for data science

Finally, in the following figures we illustrate the part of the analysis presented in Section 3.6 related to the development of the Data Science in Serbia.

Figure 68 illustrates the awareness of the development of the Data Science in Serbia. Around 44% thinks that the development is in an early stage and 26.5% finds that there are no significant results in this development.

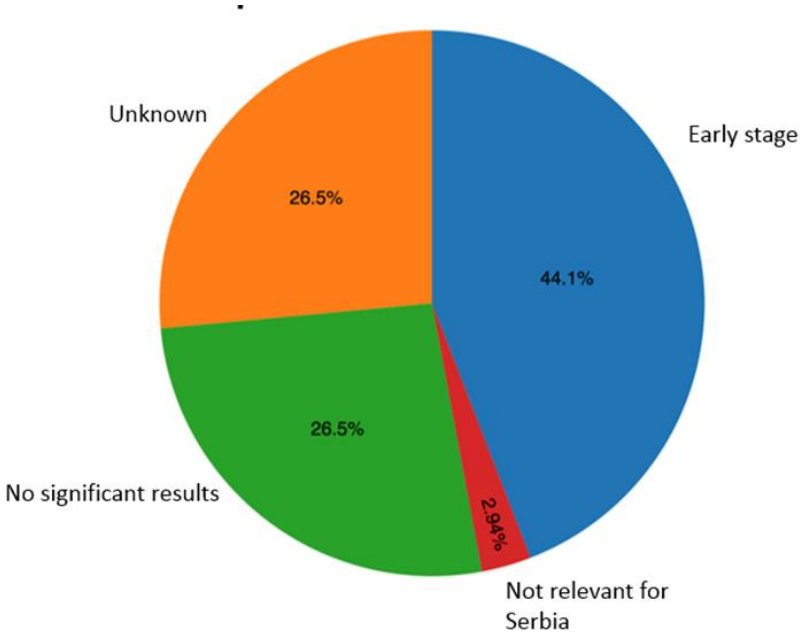


Figure 68 - Development of DS in Serbia

Figure 69 illustrates the status of the education for Data Science. Most of the companies (40.5%) find that there is a possibility for being educated through other courses offered by some faculties, which indicates a lack of the proper educational programs in this domain. In addition, around 36% finds that there is no systematic support for this type of education.

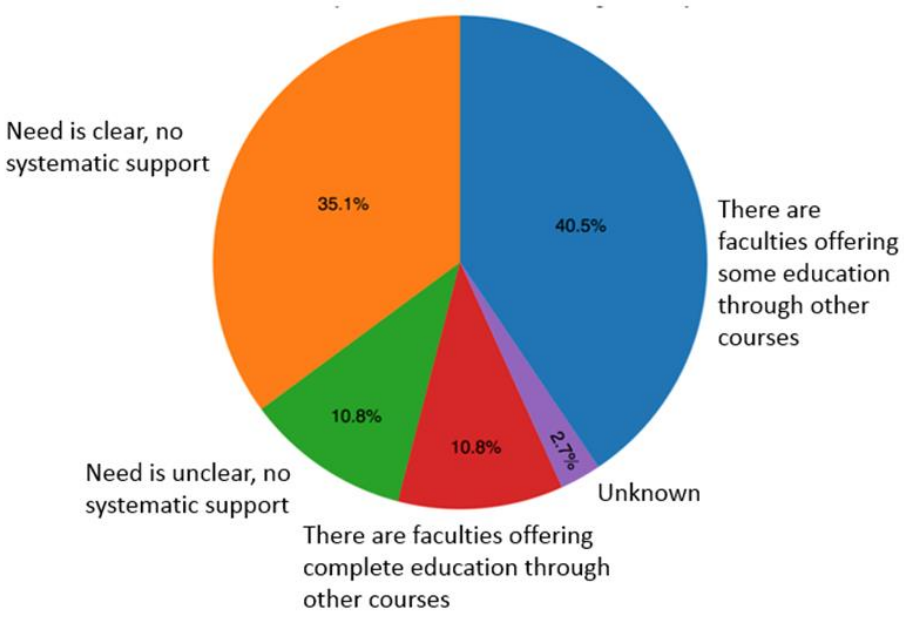


Figure 69 - Status in education for DS

Since the IT development in Serbia is in an expansion, it is to be expected that general IT professionals could cover some of the needs for Data Science skills. Figure 70 illustrates this situation through the opinion about how existing IT professionals can cover the need for data science skills. About 47% are not sure that existing courses will bring the required expertise, whereas 20.6% finds that there is a need for specialized courses.

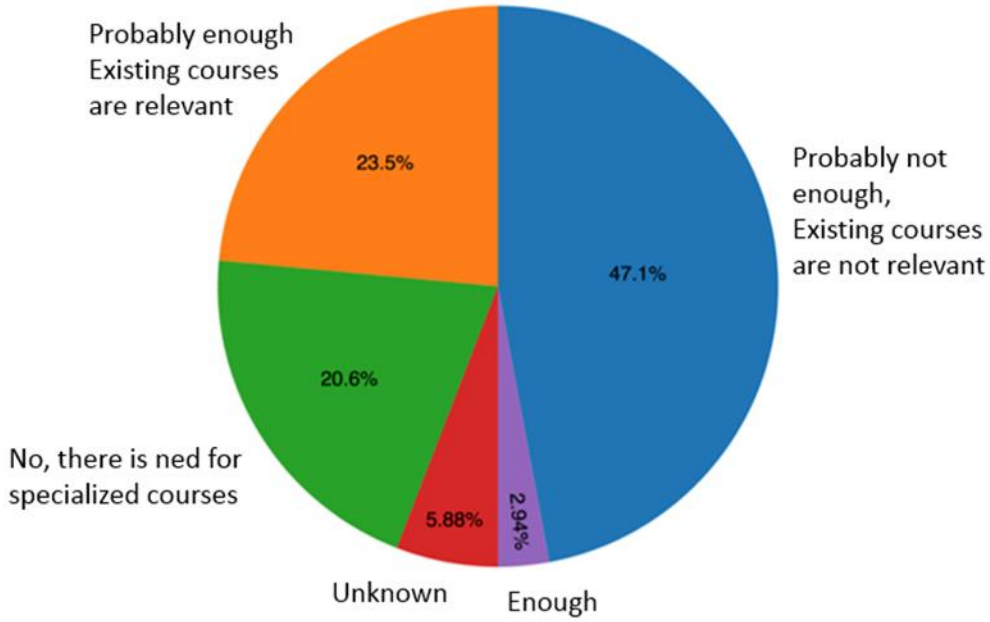


Figure 70 - IT professionals covering DS

Closing the Gap

Based on the performed survey we can see that there is a gap between the need and offer for the education in data science. We can define three ways of closing this gap:

1. Specialized courses (Universities)
2. Learning by doing (Industry support)
3. Self-education (On-line courses)

Each of those have some advantages and disadvantages, we discuss briefly:

1. Specialized courses (provided by Universities)

- Advantages
    - Systematic support

        - Part of official education process
    - Drawbacks
        - Relevance/Actuality of the knowledge
        - Too many “theoretical” lessons
        - No really applicable knowledge

2. Learning by doing (Industry supported)

    - Advantages
        - Practical knowledge
        - Industrial experience in problem solving
        - Efficient learning process
    - Drawbacks
        - No systematic support in learning
        - Biased learning
        - Less focus on basic theory

3. Self-education (provided through on-line courses)

    - Advantages
        - Very fast basic knowledge
        - Efficient learning (only relevant material)
        - Problem-oriented
    - Drawbacks
        - No systematic support in learning
        - Usually without foundation
        - Not easy reusable

We assume that the best model is a hybrid approach consisting of Specialized courses (Universities) and Learning in a practical/industry environment. In this way it will be ensured that

the methodology of learning will be satisfied and the required practical experience will be collected properly.

Chapter 5. Comparative Analysis and Conclusions

A comparative analysis and matching of the results obtained in the previous phases was carried out and is presented in the following sections.

5.1 Market Needs Comparison

The general scenario that comes out from our analysis regarding the market needs for the figure of data scientist is very heterogeneous among the involved Countries.

This variety is immediately evident from the number of samples taken in to account for the analyses. The situation was similar for Italy and Portugal, in which 40 and 44 job advertisings, respectively, were found. In UK, the number of offers available was in the order of thousands, so a limited representative set was used to extract the required skills. The situation was different in Slovenia, where the number of advertised jobs was extremely limited.

Finally, the analysis performed in Serbia showed a typical early stage scenario similar to regional areas of other project Countries in which there are no job advertises for data scientists. In Serbia the analysis of the market needs could not be done due to this lack of basic data. Instead, a survey was conducted by interviewing individual experts selected among mainly IT companies. It can be said that the need for the professionals in the domain of Data Science has started to increase, primarily due to a strong demand from some IT service provider companies, which see a huge market potential in this domain.

It is worth noticing the big gap between the two extremely different situations of UK, where the figure of data scientist is highly defined and economically recognised, and of Serbia, where the awareness about data science itself is still limited. Italy and Portugal have a similar situation with a fair number of requests for data scientists. Finally, the situation in Slovenia can be considered intermediate between Serbia and Italy/Portugal. However, the scenario shown by the analysis has to be considered simply a snapshot of the particular period in which the search was performed. It is evident that, apart from UK where there is maturity, the state of affairs is continuously evolving with a high increasing of attention, jobs and educational opportunities in all the Countries.

It also emerged from the available data that the type of company that usually search for a data scientist works in the ICT sector. Another sector of some relevance is that of insurance companies.

5.1.1 Soft skills

The range of soft skills required for data scientists that came out from our analysis reflect a fast-moving industry with rapidly evolving technologies, where teams work together intensely to create solutions to new or difficult problems, and where data scientists need excellent communications skills to interact with their colleagues and clients. Generally data scientists should be highly motivated and self-driven.

In particular, it emerged that in Italy the most desirable soft skills are to communicational and relational aspects, followed by problem solving attitude and the ability to work in teams. In the UK the situation is similar, but team working capability is on top of the requirements. In Slovenia the analysis was done using not directly job seeking websites due to the absence of explicit reference to the figure of data scientist (still referred to as more classical professional figure). However, using the approach explained in Chapter 2, the results present a scenario similar to the one in Italy and UK with more attention to the problem definition and solving abilities.

5.1.2 Hard skills

The range of hard skills is very wide ranging from generic skills such as computer programming, in Python, R or Java, and using open source tools to highly specific skills such as designing distributed databases using specific industry standard software such as Hadoop, Mapreduce or Sparks. The wide range of tools available and their rapid evolution means that data scientists must be willing and able to learn them rapidly at their own initiative.

Some of the hard skills of data scientists require a high level of theoretical and technical knowledge in mathematics, statistics and computation, e.g., machine learning, natural language processing. Typically this knowledge assumes a degree (European Levels 6, 7, or 8) in a numerate subject such a mathematics, statistics, physics, engineering, and so on.

Some of the hard skills required for data scientists require considerable experience of practical applications, taking at least two, often five or more years to gain.

5.1.3 Geographical distribution

The analysis gave also information about the geographical distribution of the offered jobs. The offers are mostly located in big city areas such as Milan and Rome in Italy, Lisbon and Porto in Portugal and London in UK.

Outside big city areas, the scenario is different. While in the UK the number of jobs (in this case) is still high outside the London area and well distributed across the Country, this is not true in Italy and Portugal. Here the offers outside the big cities are still concentrated around other cities, Bologna, Torino and Trieste in Italy; Aveiro, Coimbra and São João da Madeira in Portugal. Moreover, large parts of the Countries are not covered at all, for example south Italy and eastern Portugal.

5.2 Training Paths Comparison

The scenario that emerges from the analysis reported in Chapter 3 is, as that emerged in Chapter 2, quite heterogeneous. It is clear that existing higher education on the topics of Data Science and/or Big Data Analytics is at different stages of development in the considered Countries.

5.2.1 Number, type and diffusion of programmes

UK is certainly the Country with the mostly advanced offer, not only with respect to classical university programmes, but also to new paradigms of teaching such as e-Learning, MOOCs and Boot Camps. The majority of British universities run programs explicitly addressing Data Science and/or Big Data Analytics. They are also geographically well distributed as well as the job offering, showing maturity on this field, at least from the academic point of view.

On the contrary in Italy, Portugal, Slovenia and Serbia only classically taught programmes exist and with different shades of focus and diffusion. Some universities in these Countries, not the majority, have started to run courses, certainly stimulated by the attention and promotion that the "new profession" of Data Scientist has recently received over media.

There are common situations but also differences of focus within this group of Countries. In particular, it is worth noticing that Bachelor and Master programs explicitly addressing Data Science and/or Big Data Analytics are very rare in all of them, while most of the collected entries are sub-tracks (curricula) of general subjects such as Computer Science, Mathematics, Economics or Engineering. Moreover, in Portugal, Slovenia and Serbia it was necessary to look further inside the programs, beyond the name of the degree or the name of the sub-tracks, to find single specialization modules addressing the topics of interest. This generally shows an early stage development of the Data Science as a whole discipline in the higher education system of these Countries.

Italy showed a difference with respect to the other Countries with its peculiar post-Bachelor and post-Master specialization courses, for which there is a fair offer of education in Data Science. Since these kinds of programs are very flexible and do not typically require regularity over the academic years, they are an easier tool for universities to "follow the market" and propose innovations that are risky to implement directly in classical established degree courses such as Computer Science/Informatics Engineering, Mathematics or Economy/Business. From the analysis in 3.2 we can also observe that this approach was mainly taken by departments of Economics and Business, showing that in Italy these disciplines are more reactive to the requirements of job market about data science.

For what concerns the EQF level of the collected entries, it is clear form all the examined Countries that the peak on the offer is at EQF 7, Master level. This shows empirically that Data Science and Big Data Analytics are currently considered disciplines that certainly require already acquired basic skills on top of which the development of specific skills can be funded. This is of fundamental interest for the Da.Re. project especially regarding the entry level and the structure of the pilot course that will be defined and delivered in Phase 2.

Regarding the geographical distribution of the courses, it is clear from the results that in most of the considered Countries the location of the offered courses mostly coincides with the locations of the offered jobs in Data Science. Exemplary is the case of Italy, in which collected jobs and training paths are present only (with an exception) in the northern and central part of the Country. Furthermore, it can be reported, in general, a higher offer in areas with big cities.

5.2.2 Training paths by disciplines

Data analysis is by its nature an activity that is performed in a wide range of disciplinary fields. It is fundamental in all experimental sciences, in social sciences and also in several technical disciplines. The recognition of Data Science as a whole discipline by itself has been put forward only recently and there is not a fully accepted definition yet. This is also reflected in the genesis of the offered training paths in the higher education system.

The analysis performed in the project Countries showed that different classical disciplinary areas propose Data Science programs and that there is a certain degree of collaboration between these areas. The classical areas involved are Computer Science/Informatics/Engineering, Mathematics and Economics/Business. We can report that Computer Science/Informatics/Engineering is the prominent area given that it provides required hard technical skills on software and hardware tools as well as the algorithmic techniques on which the analytics tools are based. These techniques are also partially supplied by Mathematics, which is another classical area from which data analysis takes theoretical foundations. Statistics was considered as a sub-area of mathematics in our analysis. Finally, there is certainly a contribution from Economics and Business because of the importance of analytics in supporting decisions for business, in creating added value from data and, in general, as investigation mean for financial and economic phenomena.

At this stage of the development of Data Science as a discipline, collaboration among these areas can be considered natural. This was also reflected in our analysis; in particular in Italy and in Portugal, some of the collected courses were classified as Interdisciplinary as they were organized and supported by different departments within the same university. This interdisciplinarity can also be observed in some of the tracked PhD courses, in particular in Italy and UK.

5.2.3 Training paths by costs

Cost of the education in Data Science varies significantly across the considered Countries, mostly depending on the specificity of the higher education systems of the Countries themselves. In Slovenia bachelor and master education at state universities is free while in Italy the cost is different depending on the kind of program, in the range 1,300 EUR – 14,000 EUR per year. In particular, standard bachelor and master courses are generally less expensive than post-bachelor and post-master specialization courses, with a peak of costs in courses of these kind offered by Business and Economics departments.

In UK the cost is more homogeneous than in Italy and it is on average higher being around 10,000 EUR per year at all levels. Costs for boot camps are even higher, while the MOOCs are less expensive, provided that some of them have a unit cost only if certification is wanted. This is mostly related on the specific organization of British universities, as reported in Section 3.3.1.

To sum up, apart from specific situations in the considered Countries, it can be said that the cost for the education of a data scientist is high, but this is balanced by the fact that the average level of salaries for data scientists is generally higher than other positions.

The analysis of the costs is relevant for the Da.Re. project in order to discuss not only the target students and level of the pilot course of Phase 2 but also for future exploitation of the results of the project in certified collaborative programmes at European level. Another related aspect to be discussed is the kind of certification that would be needed for the pilot course and for future exploitations.

5.3 Training Needs Analysis Comparison

The activity of the Da.Re. Consortium on training needs analysis was conducted using different approaches:

- Direct Interviews
- Surveys
- Events

The use of a specific approach was planned according to the state of the art obtained as a result of the research activities made during Phase 1 of the project in the different European countries involved.

The results of the activity made can be summarized in Table 14:

Table 14 – Results of the activity of Phase 1

COUNTRY	RS and SLO	IT and PT	UK
State Of Art	Data science is not yet perceived as a specific topic by the majority of the companies	Data Science is a specific topic only for big companies	Data Science is well known and part of the activity of many companies
Decision Taken	Organize events during the second phase of the project	Events organized involving Academia and Private Sector	Interviews organized to suggest a new profile to be trained
Conclusions	To be taken during the second year of activity	Low level of awareness especially for SMEs High level of interest on training related to the topic Low capacity to evaluate the specific needs for training a	Specific training needs already satisfied by the education system A niche can be found with the “bridge person” able to link the specific data issues with the

		Data Scientist Need to involve companies specialized on Data Management to deal with Data issues to be solved in SMEs	specific domain knowledge The “bridge person” is an interesting figure to be considered by small, bigger, specialized and not specialized companies and can be trained with Da.Re. Pilot Course

The performed research led to the possibility to find a specific need to be covered by the Da.Re. Pilot Course to be organized during the second year of the activities. What is needed by the stakeholders is a person to be trained to help solving the problems faced by the companies using data.

We tried to synthesize the different results obtained in order to use them for creating specific training modules able to help shaping a new generation of Data Scientist. We believe that the mature scenario observed in the UK can be considered a good benchmark on which define and test innovative pathways to re-imagine education in data science. The other Countries are going to develop towards such mature scenario and will still benefit by now from the innovation proposed by the Da.Re. project.

The study of data science education shows that in the UK many needs of industry are already satisfied by university courses, MOOCs and boot camps. Our study of the needs of industry shows that, apart from a high level of technical knowledge, companies need people with data science experience. Thus, a major part of industry’s need in the UK is satisfied by a combination of existing academic courses followed by gaining experience on the job.

We have identified the profile we want to train as the one of a ‘bridge person’. This is a class of professional not well served by existing educational or training offerings. This is where we believe the Da.Re. project can make a significant impact on business, serving both large and small enterprises. Usually, the professionals with a strong knowledge of the domain of their company are not able to understand the importance of the data they are managing. A bridge between the specific domain of the companies and the capability to analyse data is necessary.

Chapter 6. Towards a Pilot Course to Re-imagine Education in Data Science

The following section offers a first introduction to the creation of the Da.Re. Pilot Course. It was produced considering the expected learning outcomes and the possible audience to be involved to attend the course.

6.1 The Data Science Training Business Needs Identified by the Da.Re. Project

As documented in this report, in the UK there is a high level of awareness about data science, with many universities, boot camps and MOOCs providing education for companies and other organisations. In the UK a university master's degree is about 1,800 study hours (45 weeks x 40 hours) and a boot camp about 500 study hours (12 weeks x 40 hours). The planned 150 study hours for Da.Re. means that our project cannot do the same thing as these courses since we have only a fraction of the study time, and our project should not try to do the same thing as these courses because they already amply supply that kind of education. The other countries we studied do not have the same kind of awareness and educational provision as the UK, but we can expect this to change with their universities responding to the increasing demand for data scientists.

Thus while the need for trained data scientists is high in industry in the UK, the existing educational and training available satisfy that need. However our research identified a gap in the provision of data science education that is not satisfied by the universities or boot camps, namely for what we have called the *bridge person*, i.e. someone who combines knowledge of an organisation's business with sufficient knowledge and understanding of data science to 'bridge' between non-technical people in the business with highly skilled data scientists able to add value to the business. We can identify three types of participants for our programme:

- someone already employed in an large organisation with a good knowledge of the business and an awareness of the importance of data science to the business, but without sufficient breadth and depth of technical knowledge of data science. The role of this person is to act as a bridge between their colleagues in the organisation and the data science services (in-house or external) used by the organisation.

- someone already employed in an SME with a good knowledge of the business and an awareness of the potential importance of data science to the business, but without sufficient breadth and depth of technical knowledge of data science. The role of this person is to raise the level of awareness of the potential of data science to the business and to form a bridge between all levels of management and the people able to provide data science services that can add value to the business – for most SMEs this is most likely to be external consultants such as a Big Data Analytics company.

- a person wanting to increase their knowledge of data science without taking a year out for a master's degree or three months out for a boot camp and/or not able to meet the cost of the

education and training. On completing our course such a person could play either of the bridging roles identified above.

- a person attracted to data science but unsure they are suited to it. At the end of our course such a person would have enough knowledge and experience to know if they wanted to go further with their data science studies. In this case our course is bridging the student to a data science career, contributing to the pool of data scientists.

This is the niche identified by Da.Re. where it can have its greatest impact on the ability of European businesses and other organisations to add value through the use of data science.

This suggests that the proposed Da.Re. education in data science implies the need to establish or extend the role of *Chief Data Officer* in companies. Such a role should be a bridge between the top management and data processing department, forcing the development of the big data activities and clarifying their importance for the continuous improvement in companies.

Apart from its 150 hour programme, Da.Re. can provide a service to senior members of companies and organisations such as CEOs, board members, directors and business persons who either do not know about data science or don't have enough time or money to start implementing it.

This project plans to produce a specifically tailored online course or a short MOOC to give an overview of what exists and what is possible and the value that data science can create if used correctly. Within the first hour of our online course (ideally within the first few minutes) we will aim to convince the business person that “Data science can increase the value of your business!” and show how to take the first step towards using data science.

In summary the problem Da.Re. will address is the gap between:

- deep technical knowledge about data science and
- deep expert knowledge about current business operation and development

Our solution is to enable a proper additional education in the data science domain for:

- employees from business who have expert knowledge about business and know little about data science. Career path: they can become Chief Data Analytics Officer (or similar) in the company (after a while)
- graduate students who want to work as a data scientist in business and are motivated to learn new technical topics of the value for their future position. Career path: they can become Chief Data Scientist (after a while)
- senior business people who have little time but want to know how data science can add value to their business, and how to take the first steps towards it.

6.2 **An Outline of the Da.Re. Programme**

The Da.Re. programme has two parts: 80 hours online education followed 70 hours face-to-face education. The idea is that the online education provides students with the technical knowledge and skills needed to do the hands-on training at the two-week 70-hours face-to-face residential school. By combining online and face-to-face education, Da.Re. can combine the best of MOOCs and the boot camp approach to provide new, useful and sustainable data science education in Europe.
The logic of this programme design is that the content of the 80-hours online education will be determined by what students need to know in order to address the case studies in the 70 hour residential part of the programme. The number and length of time for each chosen case study remains to be determined but possible areas already identified include:

- business: modelling and forecasting, modelling production and supply chains
- marketing: designing campaigns, analysing data on sales, footfall, web clicks etc.
- education: analysing data on web clicks, study times, marks gained, study paths
- scientific: analysing large quantities of multidimensional numerical data
- medicine: classification for diagnosis and treatments, statistical analysis
- city planning: modelling & mapping to forecast land use, transport, housing, services

It is too early to specify the detailed content of any online course, but the following things are likely to be needed:

- the common notion of “data lifecycle” / pipeline / methodology/ mind map
- technical issues, e.g. setting up virtual machines in the cloud with generic tools
- programming: use of computer languages, e.g. Python, JavaScript
- databases and query languages: SQL, noSQL
- modelling: types of model, e.g. network models, systems models
- statistics: statistical theory and packages, e.g. SPSS, R
- web design: user interface design, HTML, CSS, front and back end programming
- visualisation: using visualisation tools, graphics, maps GIS

This list again illustrates the trade-off we will have to make between breadth and depth.

6.3 **Prerequisite Knowledge and Background**

By hypothesis our students will know of the existence of data science but they are not expected to have in-depth knowledge of any particular element. Our target participants will have the following characteristics:

- Level 6 education or higher, e.g. a bachelors degree in any subject
- numerate and able to read simple equations, graphs and charts
- literate and able to write reports with illustrative graphics
- good search skills, finding and synthesising information
- interest in patterns of data as they impact on business
- good self-study and time-management skills
- good teamwork skills – a willingness to work with others for the desired outcome

To these can be added

- a willingness to learn and apply data science: people who have heard about data science and assume that it can help in improving business processes, but do not know enough about data science to make decisions

or

- people ready to use data science: knowing more about data science than above, but don't have enough time or resources to start implementing it.

Thus our typical students will have education to level 6 or higher, be seeking a job in industry, or be already employed people in companies (typically SMEs) who have the knowledge of their business domain and will acquire the data science competences of the bridge person.

6.4 **General Educational Principles**

There are some basic general principles that guide education. In particular it is essential that the purpose of education is clearly stated in terms of learning outcomes. It should be known what the student will know or be able to do after the course that they could not do before. Related to this is

the requirement for learning outcomes to be testable, so that it be assessed whether or not a student has achieved a stated learning outcome.

For any education it must be known who is being taught:

- what is the assumed prior knowledge?
- what are the assumed preferred learning styles?
- what is the student motivation?

These questions are discussed in the previous sections.

Also we must know what is being taught:

- what are the objectives and learning outcomes?
- what topics are included?
- what is the depth versus breadth?
- what teaching media will be used?
- how to test that learning outcomes achieved?

Learning outcomes must be written so that they can be tested. To illustrate the relationship between learning outcomes and testing consider the following examples:

e.g. Learning Outcome A1
The student will be able to describe each of the following: Data Ingestion, Adaptation, Exploratory Analysis, Cleaning, Transformation, Windowing, Dimension reduction, Clustering, Inspection, root cause analysis, visualisation, reporting, real-time anomaly detections, and give examples where they are used in data analysis.
This can be tested by a question such as: “When would an online retailer use clustering?”

e.g. Learning Outcome C6

The student will be able to open files of type .xyz in Jupyter Notebooks, find problems and clean the data, write a Python programme to implement a given function, and display the results in various ways (graph, charts, maps).

This can be tested by a student mini-project, e.g. “Use the files data.xls and map.xyz to show an Italian company the potential for selling its product in France, and write a report.”

There are various methods for assessing the objectives and learning outcomes including:

- Multiple choice questions - can give very good results

  questions can include images

  questions can include maths

- Short answer marking – can use machine learning
- Peer marking – students mark each others work

  very good for students – they learn from marking

  good students mark consistently – not poor students

- Expert marking – very 'expensive', e.g. examine a thesis

  but DA.RE. could make a 'community' of 'graduate' markers

6.5 The Design of Online Course Modules

A common misconception is that leading researchers always make good teachers. Excellent subject knowledge is necessary but not sufficient for teaching. Communicating that knowledge effectively requires a different set of skills. Often much thought is required to communicate technical ideas effectively. e-learning involves innovative use of multimedia forms including text, voice, images, animations, videos with moving images demonstrating what is being taught, narrative videos that 'tell the story', interactive graphics, interactive computer software, databases, downloadable texts such as pdf documents, and so on.

Figure 71 illustrates another common misconception - that e-learning involves videoing conventional lectures. Although some e-learning uses this approach it is generally dull and misses the opportunity for web-based presentations to allow multimedia teaching, where the medium can be chosen that best communicates what is being taught. The Da.Re. project will make its online education multimedia avoiding the 'talking heads' approach.

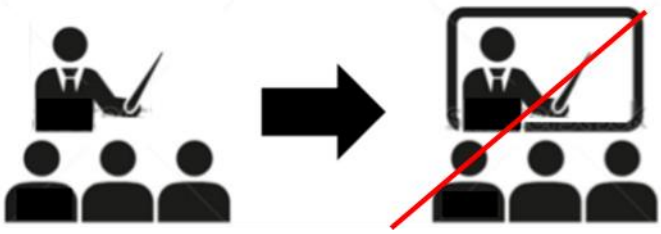
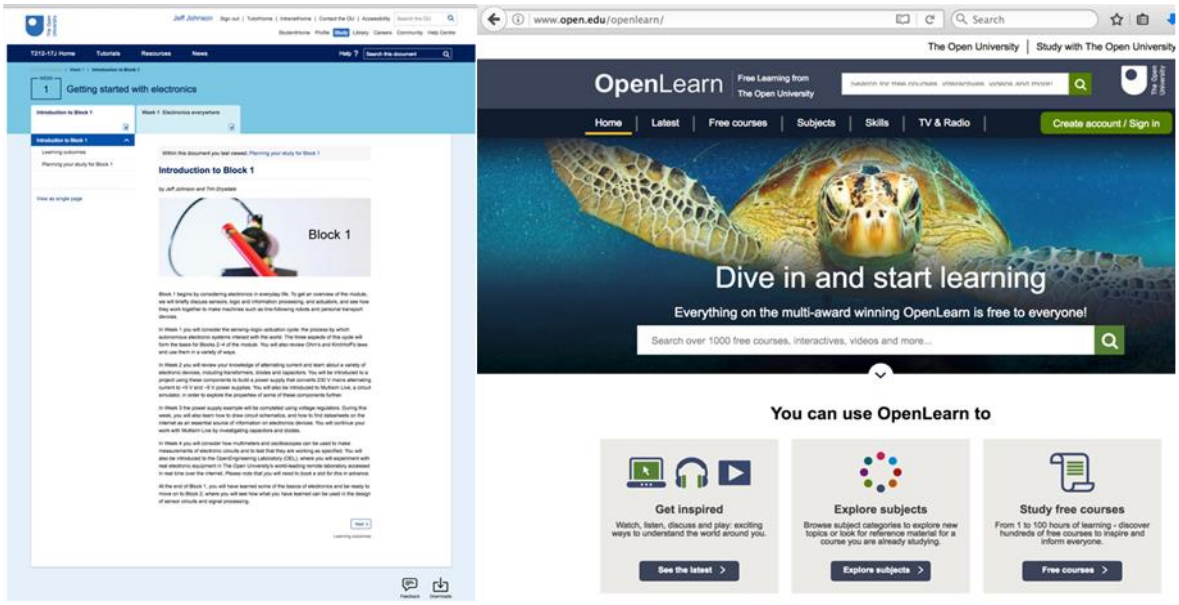


Figure 71 - A common misconception: e-learning ≠ videoing conventional lectures

The various platforms for e-learning include hand-crafted web sites, Moodle-based specialist sites such as that used by the Open University for its VLE (Virtual Learning Environment) (Figure 72) , and commercial or not-for-profit MOOC platforms – (Massive Open Online Course) platforms such as edX and Coursera based in the USA and FutureLearn based in the UK (Figure 72 and Figure 73).



(a) The Open University's VLE (b) The OpenLearn platform that delivers free courses

Figure 72 - Open University course platforms

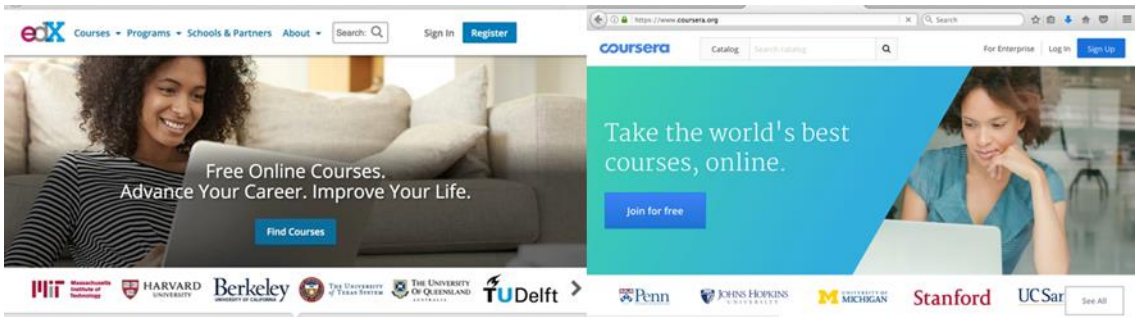
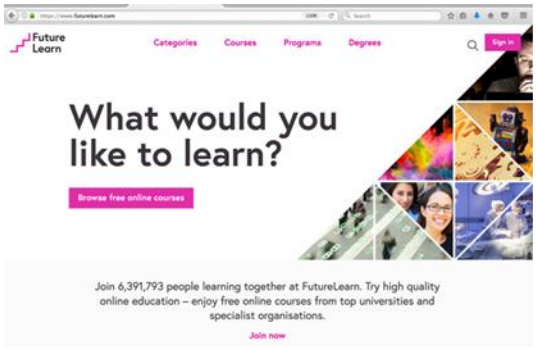


Figure 73 - The edX and Coursera MOOC platforms based in the USA

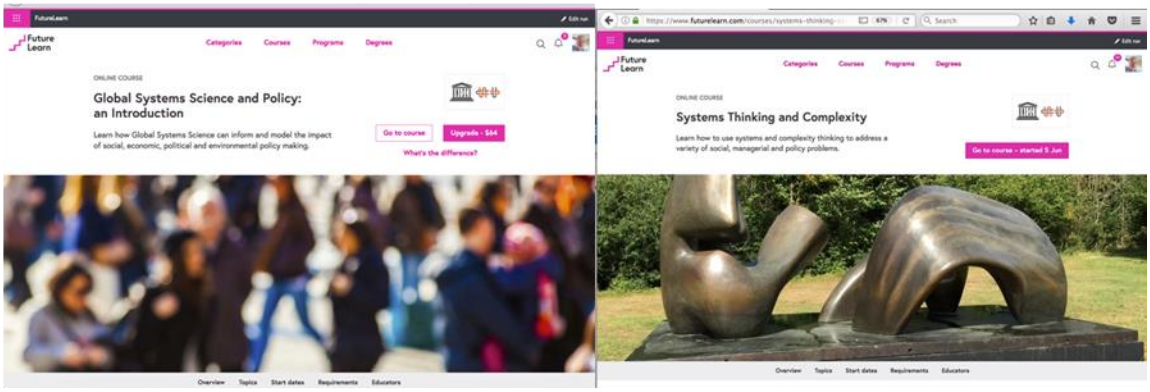
It is proposed that DA.RE. will use the FutureLearn platform for its MOOCs (Figure 74(a)). The Open University in the UK is a member of the UNESCO UniTwin Complex Systems Digital Campus (CS-DC), a consortium of 120 universities worldwide with a mission to provide free technical education in Europe, Africa, Latin America and worldwide (Figure 74b). Since Data Science is an area that it has identified for its courses, there is synergy between CS-DC and Da.Re.. Therefore we will work with CS-DC to produce the Da.Re. online courses. This has the

great advantage that CS-DC has an established relationship with FutureLearn having made two MOOCs that have now been presented many times (Figure 74c). Thus The Open University through its CS-DC connection has the necessary experience to lead the production and delivery of the Da.Re. online education and bring it to successful conclusion.



(a) The UK-based FutureLearn Platform

(b) the CS-DC as a partner for DA.RE.



(c) The UK members of the DA.RE. consortium have already made two MOOCs.

Figure 74 - Da.Re. will use the FutureLearn platform in partnership with the CS-DC

The Da.Re. online courses could have thousands of students worldwide which creates interesting educational opportunities, e.g. students create data as they study and we may be able to use their own data for case studies.

6.6 **Use of Third Party MOOCs**

The number of free MOOCs available on the internet is very large and many of these cover subjects in data science. Our research shows that, whereas some MOOCs would not be satisfactory, others may be useful for the Da.Re. project and give leverage on its funding. Also the

use of third party MOOCs could be important for the sustainability of the Da.Re. programme in the longer term and/or for acquiring some prerequisite knowledge.

6.7 The 70 Hour Face-to-Face Residential Course Module

The participants in the 70 hour face-to-face course will either be sent by business or apply to attend on their own accord.

Students and business personnel who attend the course will be separated into groups, each solving a real problem from industry in the form of a use case.

- Groups can be formed with the use case study in mind.
- Positions in the group (coordination focused data science, programming focused data science, algorithm focused data science, domain focused data science …) should be filled with the students interested in the special data science focus. For example, a business person can take the role of the coordination focused data science, a computer science student could take the role of the programming focused data science.

Use cases will be (at first) gathered from the industry partners of the consortium. Later they will be supplied by the companies sending personnel to be trained (these personnel should be domain experts).

- Primary use cases gathered from industry partners are the basis for all students.
- Special use cases can be adopted with the special interest from the companies.
- Each use case needs a document that guides the student - what to do.

Lecturers will support the groups at solving the problems (including how to use specific tools/algorithms/methodologies specifically prepared for each use-case).

At the residential courses we will create parallel sessions, challenges, etc. as creative and fun activities.

The general learning outcomes of the face-to-face course will be

1. The students have a clear knowledge of the data lifecycle:
   - Data Preparation
   - Data Analysis
   - Data Visualisation.

2. The students will practice, solving real problems that companies face.

3. Students will have the capacity to go beyond the data lifecycle by creating added value to the organisation through their knowledge of its business.

4. The trained student will be able to organise and revise a data lifecycle in an organisation. More precisely, they will be able to identify and select existing and not existing competences in the organisation, create a team and structure the work for going through an established pipeline: 1:Problem Identification; 2: Data Preparation; 3: Data Analysis; 4: Data Visualisation; and 5: Solution.

6.8 **Summary**

Our understanding of how to produce the Da.Re. educational programme is well advanced as we enter our second year.

The 70 hours study over two weeks that students will have face-to-face with their teachers and technical support assistants will be designed to be a rich and fulfilling experience. Students will work in supported teams on a range of case-study examples selected to give experience of different aspects of data science,

The 80 hours online MOOC education that precedes the residential school will be designed to give students exactly what they need for their hands-on analysis of the case study examples. We are well prepared to produce and deliver the MOOCS since we have an experienced partner in this area complemented by a wide range of in-depth technical expertise across the Da.Re. business and academic partners.

By combining online preparation with face-to-face education the Da.Re. project is beginning **to re-imagine education in data science**. Normally at residential courses you don't know what the students already know, but the Da.Re. pathway makes this possible. We believe that the roles of the so-called Bridge Person and Chief Data Analytics Office identified by the Da.Re. project can make a significant impact to improving the application of data science to European businesses and organisations.

BIBLIOGRAPHY

[1] "Big Data Value Association," [Online]. Available: http://www.bdva.eu/. [Accessed August 2017].

[2] "European Data Science Academy," [Online]. Available: http://edsa-project.eu/. [Accessed August 2017].

[3] "Edison - Building the Data Science Profession," [Online]. Available: http://edison-project.eu/. [Accessed August 2017].

[4] "Edison - Data Scientist Profession," [Online]. Available: http://edison-project.eu/edison/data-scientist-profession. [Accessed August 2017].

[5] "Indeed," [Online]. Available: https://www.indeed.com. [Accessed 2017].

[6] "Infojobs," [Online]. Available: http://www.infojobs.it. [Accessed 2017].

[7] "Monster," [Online]. Available: http://www.monster.it. [Accessed 2017].

[8] "Glassdoor," [Online]. Available: http://www.glassdoor.it. [Accessed 2017].

[9] Procorre. [Online]. Available: http://blog.procorre.com/data-scientist-roles-surge-in-first-half-of-2016. [Accessed 2017].

[10] J. H. Johnson, V. Modhavadia and C. Jimenez-Romero, "A survey of data science jobs in the UK," 2017.

[11] Adzuna, "Data scientist salary stats," [Online]. Available: https://www.adzuna.co.uk/jobs/salaries/data-scientist. [Accessed 2017].

[12] O. Hogan, L. Holdgate and R. Jayasuriya, "The value of Big Data and the Internet of Things to the British Economy," 2016.

[13] C. Rhodes, "Business Statistics," 2016.

[14] European Commission, "What is an SME?," [Online]. Available: http://ec.europa.eu/growth/smes/business-friendly-environment/sme-definition_en.

[15] M. Ward and C. Rhodes, "Small businesses and the UK economy," Standard Note SN/EP/6078, 2014.

[16] M. Grobelnik, *Analysis of the knowledge and skills of data scientists in Slovenia (master's thesis), Available in Slovene: http://www.cek.ef.uni-lj.si/magister/grobelnik4991.pdf,* University of Ljubljana, Faculty of Economics, 2016.

[17] B. Horvat; J. Borlinić (Abelium), "Internal Research".

[18] European Commission, "Descriptors defining levels in the European Quallifications Framenwork (EQF)," [Online]. Available: https://ec.europa.eu/ploteus/content/descriptors-page.

[19] "Coursera MOOCs platform," [Online]. Available: https://www.coursera.org. [Accessed 2017].

[20] "edX MOOCs platform," [Online]. Available: https://www.edx.org. [Accessed August 2017].

[21] "Future Learn MOOCs platform," [Online]. Available: https://www.futurelearn.com.

[22] "Cercauniversità," [Online]. Available: http://cercauniversita.cineca.it. [Accessed 2017].

[23] "List of Universities in UK," 2017. [Online]. Available: https://en.wikipedia.org/wiki/List_of_universities_in_the_United_Kingdom.

[24] A. Renda and G. Luchetta, "Dipartimento Politiche Europee," [Online]. Available: www.politichecomunitarie.it/file_download/1570. [Accessed August 2017].

[25] I. Luković and M. Šolaja, "Trends in the academic education in data science in Serbia," in *XXIII symposium Development trends in higher education in Serbia*, Zlatibor, Serbia, 2017.

Annex 1

The Annex 1 contains the whole set of the data collected on the European Training Paths by the project partners.

The Template used for the collection of the information is described in detail in Chapter 3 of the Intellectual Output 1.

The detailed content of the Annex 1 can be downloaded at the following link: http://dare-project.eu/download/