EDISON

Education for Data Intensive Science to Open New science frontiers

Project no. 675419
Coordination and Support Action
Funded by the Horizon 2020 Framework Programme of the European Union

Call identifier: H2020-ICT-2015-1
Topic: INFRASUPP-4-2015 - New professions and skills for e-infrastructures
Start date of project: 1 September 2015 (24 months duration)

Deliverable D4.1 (version 1 of 2)

Strategic market analysis for sustainability potential

Opportunities for EDISON in the Data Science Professionals labour market

Due date:	31/08/2016
Submission date:	24/08/2016
Deliverable leader:	INM

Dissemination Level

☒	PU:	Public
☐	PP:	Restricted to other programme participants (including the Commission Services)
☐	RE:	Restricted to a group specified by the consortium (including the Commission Services)
☐	CO:	Confidential, only for members of the consortium (including the Commission Services)

EDISON project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 675419

List of Contributors

Participant	Short Name	Contributor
Inmark Europa	INM	RAR, AMP, CLR, SG
EGI	EGI	TA
University of Southampton	Southampton	SB
University of Amsterdam	UVA	YD

Change history

Version	Date	Partners	Description/Comments
0.0	06/05/2016	INM	Initial version. Structure and content
0.1	22/07/2016	INM	Complete draft for internal review…
0.2	24/08/2016	INM	Final draft
1.0	25/08/2016	UVA	Final version

Executive summary

The tagline of the EDISON consortium: *"Building the Data Science profession"* describes the efforts to foster growth in the number of Data Scientists or Data Science Professionals (DSP) getting proper jobs in companies, research institutions or public administration bodies. In this emerging labour market, **the EDISON team works to help DSP labour Supply side players -i.e. education, training and certification providers- to improve the employability of DSP, thus addressing the needs of these new professionals and those of their potential employers (the DSP labour Demand side players).**

This report summarises the outputs of a first strategic analysis of the DSP labour market. **It provides insights to design the sustainable exploitation of EDISON results, aiming to maintain the validity of the assets produced in the project beyond 2017, when EU funding of EDISON ends**. This exploitation will comprise different kinds of education, training and certification services and mechanisms together with the technical and investment decisions to deliver them under competitive terms. This report is the first of two editions of the strategic market analysis of the DSP labour market, the second being due at the end of the project. In this edition we present the conceptual definition and the current situation of the DSP labour market, we analyse key features of the demand and the supply sides and **we conclude depicting opportunities and recommendations to exploit EDISON's outputs while closing diverse market gaps explained in this report**.

The general conclusion is that **the DSP labour market is incipient, in its infancy stage and it therefore shows all the distortions which are typical of an emerging market.** Up to now, there is no formally established and widely accepted -at market level- definition of what a DSP should be in terms of value delivered to employers, or the roles and basic mechanisms of the market (roles of the market players, recruitment channels, retribution packages, etc.). This leads to blurred boundaries and misinterpretation on both the demand and the supply sides of the market. There are, indeed, forecasts on future demand growth, and as to what constitutes the DSP labour market; but estimates are still quite imprecise. Furthermore, as part of the digital economy novelty and the lack of clear definitions, it is almost impossible to define the potential size of the DSP labour market today.

However, **ultimately these market gaps offer us big opportunities for exploiting EDISON outputs:**

- The **growing quantitative gap** between the number of DSP needed by innovative sectors and the skilled workforce that established education, training and certification suppliers are currently able to provide. There is a **visible sense of urgency** to fill this gap among sectors such as the Internet economy or the ones widely exposed to the digital transformation. This urgency is quite evident in the USA, where this labour market already shows a higher degree of advance.
- At the same time, there is **still low demand traction for DSP jobs in most sectors of the European economy, in the research ecosystem and in the public administration.** This gap offers a certain **lead-time to EU Supply side players to shape a competitive offering to future DSP.**
- There are **many unmet needs for data driven professionals who can interpret data at all levels, ranging from the clerical level to the executive level**. All roles have to be catered for; and in order to achieve this, data literacy has to be oriented from the start. **There is a clear opportunity to shape the future DSP labour market**. Universities and training centres need to ensure their DSP related courses actually attract students and offer solutions to DSP employers. **There is a chance to build up the attractiveness of the profession**, as companies progressively understand the value of data and will be prepared to reward those able to extract greater value from data.

The **urgency to mitigate the identified gaps and support strategic changes can only be addressed by aligning interests and efforts of DSP Demand and Supply market players as fast as possible**. To date, there is scarcely visible progress in this convergence. **An ecosystem providing the mechanisms to mobilise the DSP space is needed as soon as possible; and EDISON can provide its building blocks. It is the right time to invest efforts and money to put all of us to work together in this endeavour.**

TABLE OF CONTENTS

1 Introduction

1.1 Purpose and content

The EDISON H2020 Coordination and Support Action tagline: *"Building the Data Science profession"* describes our mission to foster future growth in the number of Data Scientists or Data Science Professionals (DSP) occupying job positions defined as such.

Along the project lifecycle, the EDISON team is working to help Demand side players (DSP employers) and especially Supply side players (DSP education, training and certification providers) to close quantitative and qualitative labour market gaps. Within this context, this document gathers the first results of the DSP labour market analysis, one of the foundations for designing plans for the long-term usage ("sustainable exploitation") of EDISON outputs.

Formally, this report delivers the results of Task 4.1. As stated in the DoA: "The purpose of this task is to ensure that EDISON products and services are positioned on the market to fulfil demand needs in a profitable (sustainable) manner", which in turn contributes to project Objective 3 of EDISON CSA: "Develop a sustainable business model…", the sustainability plan being considered one of the solid foundations – three pillars – for the Data Science development: the development of a sustainable business model and the proposal of a (long-term) roadmap for the establishment of the Data Science profession in Europe that will allow sustainable operation of EDISON services and outputs beyond the duration of the project.

This document constitutes the first of two editions of the strategic market analysis deliverable, the second being due at the end of the project lifetime. The content of this report is structured as follows:

- In the introduction, the concept of a strategic market analysis is explained and a DSP labour market model is described. The basic tools to understand how the DSP labour market is structured and how it operates are explained, together with the methodology used to carry out this study.
- In Chapter 2 the current situation of the DSP labour market is analysed, and the roles of DSP in the different market segments and how this market is expected to evolve are reviewed.
- Chapter 3 delineates the demand side of the DSP labour market, looking at the drivers to hire DSP, their employment conditions and the recruitment practices.
- Chapter 4 analyses the supply side of the labour market, that is, the actions carried out by the different players or educational organisations that prepare these professionals to enter the labour market, the characteristics of the educational offerings and the employers' positioning vis-a-vis the potential DSP.
- Chapter 5 presents the conclusions of the analyses in terms of the opportunities and recommendations to exploit EDISON outputs to generate a competitive offering of products and services helping the DSP labour market to develop and consolidate.

1.2 Approach and methodology

The strategic market analysis report is a living document that primarily supports EDISON partners and interested third parties in defining how the project outputs could be used both in commercial and non-commercial settings once the H2020 funding of EDISON comes to an end in mid 2017. This analysis focuses on the so-called "Windows of opportunity"[1], understood as a space of unmet needs among certain demand side groupings of players, due to the failure or inexistence of an adequate offering by supply side players. In other words, it provides insights to formulate the business cases for the future

[1] A primer in Entrepreneurship Chapter 2 – Recognizing Opportunities and Generating Ideas, Prof. Dr. Ulrich Kaiser Institute for Strategy and Business Economics University of Zurich Fall semester 2008

exploitation of EDISON-enabled products and services. These Business cases will be in turn the foundations for the Sustainability Plan, which is the ultimate ouptut of EDISON´s Workpackage 4.

The methodology of this initial market analysis is based on the following approach:

- Socio-economic and market modelling. Profiling the *Supply* and *Demand* sides of the DSP labour market, identifying the roles of the market players and their interdependencies and economic behaviour. This modelling has been framed within the context of the competitive landscape defined by external factors beyond the managerial control of DSP labour market players. Modelling tools have also been used for the diagnosis of the current and prospective evolution of the DSP labour market, for example, well-known analytical tools such as PEST *factors*[2], and *SWOT*[3].
- Desk research and fieldwork techniques for gathering data and information required for feeding and supporting the analysis, synthesis, and modelling of findings and preliminary conclusions.

The main steps for combining data and information into useful inputs for the study have been:

- **Background analysis**: as a first stage, we drafted preliminary statements on the diverse issues under investigation, extracted from the team's previous knowledge on the Data Science space, as well as from literature review.
- These starting points were enhanced and illustrated with insights and real-world data derived from **fieldwork** via qualitative interviews with DSP current and potential employers and educators/trainers performed by EDISON team members. This **primary data gathering** (i.e. from original sources) included:
    - 35 qualitative face-to-face and/or telephone-based in-depth interviews performed by EDISON team members with domain experts and market players. These interviews aimed at complementing, validating and illustrating desk research findings on the diverse issues related to DSP employability. A set of discussion guidelines were devised (included in Appendix 2), adapted to the profile of the Supply and Demand market players. The interviewees comprised a relevant group of managers, researchers, and practitioners involved in leading and/or pioneering DSP employment related initiatives in either private or public sector organisations such as: Research Infrastructures and Institutions, Higher Education Institutions and industry. The list of organisations covered by these interviews is presented in Appendix 2.
    - Participation of team members in Data Science related events e.g. conferences, workshops and research project meetings, which allowed the team to gather insights directly from state-of-the-art presentations and debates among top level researchers, practitioners, and domain experts as well as to identify and contact key interviewees in a "friendly" environment, thus fostering discussion engagements and facilitating further contacts.

- ***Desk Research*** included the collection and review of a variety of labour market published sources such as market research reports, white papers, articles in specialised business and educational publications, policy roadmaps and prospective studies, as well as and statistics, directories and market players' databases. The review of these sources established the framework for the study with a description of DSP labour market trends, socio-economic conditions and structures, competition, market, education and innovation mechanisms.

[2] The PEST analysis. Aguilar, Francis. *Scanning the Business Environment*. New York: Macmillan.

[3] SWOT stands for "Strengths, Weaknesses, Opportunities, and Threats". Strategic planning diagnosis tool, created at the Stanford Research Institute in the 1970´s by a team led by Albert Humphrey

1.3 EDISON DSP labour market model

The first step in the analysis was the definition of a simple model to depict the structure and operations of the DSP labour market. This model helps us clarify the players to be involved as well as the relevant relationships and focal points or market gaps. The initial model is depicted in Figure 1 below:

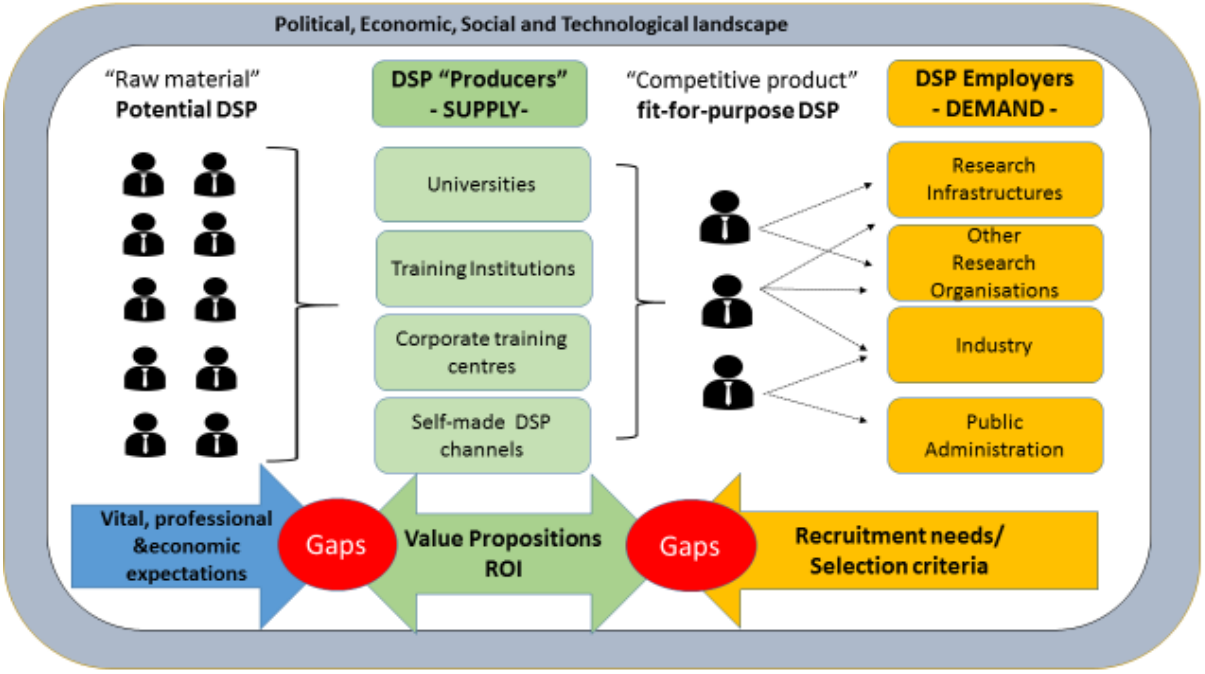

Figure 1 – Modelling the DSP labour market

The basic description of the components of the model is the following:

- **The Political, Economic, Socio-cultural, and Technological Landscape** represents the combinations of external factors (the so-called PEST factors) shaping the creation and development of a market. The analysis of these macro-environment variables[4] or Market Shaping Forces is a pre-requisite for establishing the contextual framework for the analysis of the DSP labour market. The future evolution of landscape trends and facts will largely shape the context for DSP progress towards becoming a relevant professional market category. Our current understanding of the landscape is depicted in section 2.1.
- **The DSP labour market** depicts on the one hand, the Demand side, the "DSP employers", organisations that employ or could employ DSP, which range from the financial sector, telecom and retailers to defence and health; other institutions, such as research organisations, both large Research Infrastructures and smaller research organisations and public administration bodies (e.g. health systems). On the other hand, the Supply side of the market ("DSP producers") are the education, training and certification institutions and organisations active in preparing individuals willing to acquire DSP related skills and to become eligible for DSP related jobs. The Supply side is made up of all organisations that educate or train DSP candidates, and this includes universities and other training centres, and the in-house corporate training options that companies and other organisations offer to their employees, including Massive Open Online Courses (MOOCs), Continuing Professional Development (CPD) and self-driven options. **If the Supply side is to "produce" DSP that successfully respond to what the demand side asks for**, these skilled DSP are to be trained according to the needs of the DSP employers, i.e. becoming **fit-for-purpose.**

[4] In the sense of expressing a lack of managerial control of such variables by DSP labour market players

Therefore, the **Supply side players must systematically address two main gaps** in order to succeed in fulfilling their role in building up a competitive, skilled, fit-for-purpose workforce of DSP:

- **Meeting the needs of the future employers of DSP in both qualitative and quantitative terms**, which they express by means of the job descriptions/requirements and their recruitment mechanisms and selection criteria (the Demand Gap). In order to reduce the qualitative gap, universities and other stakeholders on the supply side have to take all these aspects into consideration when they plan their courses and curricula.
- **Simultaneously, the DSP Supply players must meet the needs and expectations of the future DSP (The "DSP raw material")**, those who are capable and willing to be educated, trained and/or certified in order to fill job positions as DSP in research, industry and the public administration. These mainly consist of university students looking for their professional career development, and by professionals willing to update/enhance their skills to re-launch/re-orientate their professional path. These people have a wide variety of vital, professional and economic expectations, that DSP Supply side players must consider in order to educate/train/certificate them successfully, and this includes helping them find the tracks that are adequate to their capabilities, providing the best possible learning conditions and fitting their educational offer to employers' needs. These economic and personal expectations range from wanting to have a stable job, albeit not the highest salary, to adapting to new challenging situations while not minding changing jobs frequently in search for professional progress and for a higher economic compensation.

1.4 EDISON´s potential role in the DSP labour market

Working on both sides of each of the gaps presented in the previous section increases the speed/probability of closing them. Therefore, the three groups of market players defined in the market model: a) Employers, b) Education, training and/or certification suppliers and c) DSP candidates, are potential targets for one or more services to be offered on the basis of EDISON´s results. In this regard, EDISON outputs can support the activities of "win-win" matchmaking mechanisms, helping the Supply side of the DSP labour market to address the two gaps simultaneously, by supporting the three main types of market players involved in the DSP labour market. (as shown in Figure 2).

Regarding the supply side, EDISON can collaborate with educational and training organisations, and contribute to set certification standards, to help harmonise the best possible DSP workforce offering in a wide variety of cases. EDISON can also address the DSP´s needs, helping them structure the study programme that best fits each person in order that they find the best job position, and can also help DSP improve in their profession, as it is acknowledged that learning is inherent in a good Data Science Professional's career. Regarding the demand side, EDISON outputs could provide tangible value for DSP employers from research, industry and public administration sectors contributing with the definition of their particular needs for each occasion and helping them find the DSP that are fit for their purposes.

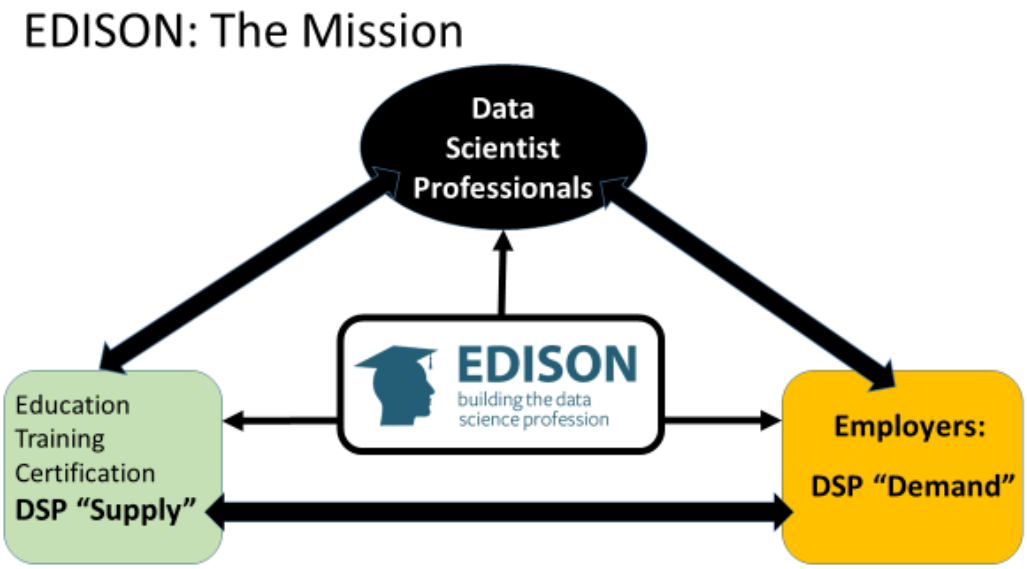

Figure 2 – EDISON´s role in the DSP labour market

2 Overview of the Data Science Professional labour market

2.1 A changing landscape fostering the demand of DSP

The analysis of the context for the future development of the DSP labour market is framed by a combination of trends, reflecting ongoing deep changes in society as a whole. This section addresses the analysis of the so-called PEST factors: *Political, Economic, Socio-cultural, and Technological factors* that act as **market shaping forces**. The value of this analysis rests on the notion that future evolution of these trends will help to set up the landscape for the likely evolution of the digital driven labour markets, among them the DSP labour market.

In this context, two "macro" trends have been widely recognised as driving pervasive change: a) Globalisation has already affected the way organisations operate to deliver their services, produce and market their products, which includes the way they create and manage human workforces; and b) The accelerated technological changes posing systematic adaptation, continuous development and structural improvements. **The phenomena acting as drivers for Data Science and DSP development not only will persist, but will also grow deeper**. **The explosive acceleration of the so-called digital transformation of all aspects of the economy and society as a whole will foster continuous growth in the demand of digitally skilled professionals capable of working with data driven technologies, where DSP play a key and consolidating role.**

These trends have in turn triggered a wide combination of change/adaptation processes in the Political, Economic, Social and Technological factors influencing the DSP labour market.

2.1.1 Political Factors

Transformations in the **Political conditions** and governmental policy not only influence macroeconomic conditions, regulations and the political realm, but also impact on public and private sector organizations affecting the ease or difficulty in being successful and profitable. The changes in the political environment affect the availability of capital, cost and demand of acquiring technology, material and human resources to remain competitive and profitable. National governmental actions and transnational policies, (e.g. EU) foster the use of technology, looking beyond the recent economic crisis to the need for preventing and overcoming similar conditions in the future through improved regulation and visibility: the Digital Agenda (DA) and the Digital Single Market (DSM) or the policies to foster technological actions in order to stay globally competitive force industry to make the best of the data being produced daily. **The 830.000 unfilled vacancies for IT professionals that are estimated for 2020 are leading to the need for digital skills for nearly all jobs where IT complements existing tasks**[5].

- Regulations that affect the education systems and security and big data, and legal frameworks that guide the way public officers are hired, as different from the private sector, also affect the data science labour market. Regarding education, the Bologna process contributes to harmonising and levelling higher education studies across Europe, and recommendations on how universities develop and implement courses affect the way universities can design and develop degree studies.
- European Open Science Cloud[6] policy and European Data Infrastructure: Europe is the largest producer of scientific data in the world, but insufficient and fragmented infrastructure means this 'big data' is not being exploited to its full potential. By bolstering and interconnecting existing research infrastructures, the new European Open Science cloud will offer, according to the EC, Europe's **1.7 million researchers and 70 million science and technology professionals working in industry and the public administration** a virtual environment to store, share and re-use their data across disciplines and borders.
- The Grand Coalition for Digital Jobs promoted by the EC is another political reaction to the need for digital skills for nearly all jobs where digital technology complements existing tasks**. In the near**

[5] https://ec.europa.eu/digital-single-market/en/skills-jobs
[6] http://europa.eu/rapid/press-release_IP-16-1408_en.htm

future, 90% of jobs - in careers such as engineering, accountancy, nursing, medicine, art, architecture, and many more - will require some level of digital skills. There are new patterns of knowledge acquisition and this is coupled with rising demand from new industries and from the transformation of traditional services and manufacturing sectors into the digital economy. Furthermore, facts show that this **process is accelerating, challenging organisations and institutions to adapt in less and less time, making conventional education, training and re-skilling practices to become obsolete equally quickly, thus forcing demand to replace them by other new ones** (MOOC, Youtube, gamification and Serious Games are some examples).

2.1.2 Economic conditions

The dynamics of Economic factors are key drivers for new markets such as the DSP labour market. **Built on top of globalization and digital transformation, the *Data driven economy* will lead to more focused and effective business opportunities, productivity growth, and increased competitiveness in data across the whole economy.** This will require **speed in the acquisition of new skills and adeptness in changing behaviours and developing culture across the productive system, thus pushing forward the needs for novel DSP profiles**. One particular issue to be taken into account when analysing the DSP labour market is that nowadays there are inevitably serious difficulties in measuring the Digital Economy (DE) because it is hard to define precisely in itself, it is moving at a very high speed and the traditional statistical measuring instruments have become plainly obsolete. Some DE roles are completely new; others are essentially re-skilled versions of long established professions. These constraints obviously apply to the interlinked markets such as labour.

2.1.3 Social factors

Socio-cultural dynamics and cultural models are affected by the Digital Transformation of society: the need to have more and more highly skilled workers that can play a role in high technology environments is unstoppable. Social networks affect the lives of increasing amounts of people (people depend on them for finding a job, communicating, buying and receiving payments, etc.). Culture is becoming increasingly global with the aid of the world wide web. Open access to information and to growing amounts of data entail security issues, which is a serious worry in Europe, while at the same time make information and knowledge more open and easy to access, use and re-use. A growingly heterogeneous population, with massive arrival of people from other cultures have to be harmonised while preserving the European cultural roots.

In this context, socio-cultural, behavioral transformation has been influenced by the shift from pre-digital to a digital world, and the development of the new social dynamics and cultural models of the Information Society. *Social Disruption and widespread changes in societal cultural behaviours such as the* Cultural Globalisation and the "born digital" generations are trends supported and fostered by the web, as the internet is changing everyday life, with their socio-economic standing influencing how they use the web. Some of the new personality traits, values, attitudes, interests, or lifestyles include:

- ***The trend to create active communities***: The dynamic growth of digital content includes user-created content generated by the rapid growth of Web 2.0 social networks, which have become commonplace, and the related phenomenon of collaborative creation (co-creation) of content.
- ***Social media as a primary method of communication and creative expression*** *("self-casting", pervasive user created content,…)*: Users are becoming active participants as well as information providers. These phenomena are changing the way organisations must compete in delivering services, thereby triggering requirements (engagement, dissemination, training and re-skilling) where Applied or Serious Games[7] can provide a pathway toward the solution.

[7] i.e. games designed for a purpose other than pure entertainment. In Applied Games – In search of a New Definition. R. Schmidt, K. Emmerlich and B. Schmidt. (Entertainment Computing ICEC 2015. Dec. 2015).

- ***Globalisation and Openness of knowledge for tackling social challenges*:** Globalisation also has impact on science, where the sharing of scientific outputs is growing rapidly, deconstructing boundaries and disciplines. Additionally, increasing conditions for complying with open access policies to make content and data easily discoverable and accessible to communities is focused towards solving major societal challenges such as climate change, resource scarcity, and demographic shifts.
- There is an **overwhelming need to transform all avenues of life and enable digital citizens' deeper fulfillment via digital knowledge and understanding as new workers, better students, etc**. These challenges are not currently being successfully addressed via the traditional approaches in education and training, nor by the formal educational and vocational training systems, working with the conventional (i.e. outdated) educational, training, engagement and methodologies, technologies, tools and processes. EDISON is to address these market challenges by means of the foreseen Community portal services.

2.1.4 Technological factors

In multiple reports issued since 2007 IDC, the leading ICT market analyst, identifies four trends currently affecting ICT and driving their current growth: mobile communications, cloud computing, social, and Big Data and Analytics technologies. General technological trends include decreasing prices and the improved performance of ICT products and services, the improved ability to store at lower costs and share information, the vastly increased number of people online, and businesses routinely using the web in increasingly knowledge-driven, customer-centric organisations. Under this umbrella we should include other vast areas such as the Internet of Things, Cybersecurity, and the corresponding continuous professional development and re-skilling. **All in all, this pervasiveness of changes has a high impact on the need to upskill workforces**. Within this space, the impact multiplier potential of Data Science teams becomes a great opportunity for new education, training and certification services offerings. In this context EDISON should become a unique source of competitive advantage for DSP Supply side players.

2.2 The DSP labour market today

The DSP labour market is incipient, and therefore has all the distortions that are typical of an emerging market. The following observations arising from desk research analysis and widely confirmed by the interviews carried out with DSP labour market players, summarise the current infancy of the DSP labour market:

- Up to now, there is no widely accepted – at market level – definition of what a DSP is, and there are different pragmatic definitions around value creation and transfer demanded of these professionals, which lead to blurred boundaries and misinterpretation on both the demand and the supply sides of the market. There are clear signs of a) ignorance, as e.g. employers do not know the value added DSP could offer, and therefore cannot specify what kind of employee they want to hire; and b) confusion on where and how DSP can be recruited or how they should be educated and/or trained. Ambiguities in the definition of DSP creates problems all along the value chain. Recently Dataversity illustrated this lack of a unified definition: "*After a lengthy discussion, I realized we were using the same term to mean different things. As a result, there was no agreement on what we had just agreed to*"[8].
- Development of both sides of the market are still being driven by early adopters, mostly located in the Silicon Valley and/or linked to Internet industry and global retailers and financial services; with a supply side coming from an unstructured space ranging from the internal corporate efforts of innovative companies to the leading Universities in the Bay area and Massachusetts. From a

[8] Selling the Value of Data Science, Governance or Analytics http://www.dataversity.net/back-basics-selling-value-data-data-science-governance-analytics/)

geostrategic point of view, the DSP labour market shows its highest degree of progress in the USA, linked directly to their worldwide lead in the digital transformation and globalisation processes. But even within this leading segment of the market, definitions, rules, *de facto* standards (in the widest sense), procedures, workflows, best practices, etc. are still in their infancy or yet have to be created. Transferring this huge gap into our European lagging behind context, the overall preliminary conclusion translates into an interesting space of diverse opportunities for future products and services enabled by the EDISON Data Science Framework (EDSF).

- DSP jobs are being filled (when it happens) after longer than standard recruiting periods; and jobs are occupied by a variety of professional profiles which at best should be considered as "proxies". Vacancies are announced for different positions named by old/existing profile names (because companies do not have DSP well defined yet). Jobs are offered to candidates that have best match of required competences and skills even if not defined as DSP. Many are self-educated and usually come from different areas that have experience of using data related technologies in their original professional areas.
- The defining lines among the different market players are blurred, as their roles in the market (who are the DSP, who helps them to become fit-for-purpose, who recruits them, who selects them, etc.) are not clearly established, contracting channels and prices (remuneration packages) are not yet settled. There are, indeed, forecasts on future demand [9] and as to what constitutes the DSP labour market, but still quite imprecise to date. Furthermore, as part of the Digital Economy novelty and the lack of clear definitions, it is almost impossible today, even within the USA space, to define the size of the DSP labour market. The EC has addressed this total lack of quantitative information by launching a "European Data Market" monitoring tool[10]. Its Initial reports published in late 2015 have started to provide basic estimates, referred to the Data Economy, and including the so-called data companies and data workers, which provide at least orders of magnitude that help to make some inference with respect to the DSP labour market:

    - "In 2014, the European data industry (data producers) comprised approximately 243,000 companies with a share of 14% of the 1.7 million enterprises populating the ICT and professional services sectors. The population of companies with a strong reliance on data (data users) was of 642,000 enterprises in 2014, corresponding to 6.3% of the 10.3 million potential user companies (excluding the government sector) - a relatively low penetration".
    - "6.1 million data workers in 2014, representing more than 3% of the total employment in the EU. The number of data workers therefore increased by almost 6% year-on-year in 2014, which is well over the 2.2% growth rate for employment in the EU registered in 2014 (European Commission, spring 2015). All the industries are potential users of data and therefore employ data workers, although there are industries which are much more data intensive than others, such as professional services, wholesale and retail, information and communication, finance, and manufacturing, which together count for more than 70% of total data workers".
    - "Data Scientists in 2014 represented less than 2% of the estimated population of data workers in the EU, counting about 119,000 workers".
    - "In 2014 in the EU there was a gap between total demand and supply of data workers of 509,000 jobs, corresponding to 7.5% of total demand".

2.3 Variety of DSP job functions and roles

An important sign of the early stage of development of the DSP labour market relates to the functions and roles played by DSP, or at least expected by their employers. Desk research and interviewees have provided diverting insights about this dimension of the market. In this regard, the situation can be described as follows: The profiles and roles that professionals referred to as DSP play in their working

[9] http://www.forbes.com/sites/gilpress/2015/04/30/the-supply-and-demand-of-data-scientists-what-the-surveys-say/#4178857c205e
[10] SMART 2013/0063 European Commission, DG CONNECT, performed by IDC

environment are substantially different according to the sector they work in: the difference between DSP working in scientific research organisations, e.g. in Research Infrastructures, reveal patterns which widely differ from those working in the corporate world. DSP do not do the same job, do not have the same needs, do not have the same knowledge, and do not get the same salary.

- In the business world, a Data Science Professional is seen as specializing in deriving business intelligence and analytic insights from structured and unstructured data sources. Data science and analytics should be perceived and managed by companies as a professional function with its own clear career path and well-defined roles. From Tom Davenport's article on "Light quants" and analytical translators[11], "light quant" is someone who knows something about analytical and data management methods, and who also knows a lot about specific business problems. The value of the role comes, of course, from connecting the two. An "analytical translator" may also have some light quant skills, but this person is also extremely skilled at communicating the results of quantitative analyses, both light and heavy. It is by taking advantage of these soft skills that the value extracted from data is transferred to business decision makers and into competitive advantage, justifying their high compensation packages.
- In Research organisations, DSP fill support positions. According to the experts interviewed in Research organisations, DSP do not lead research activities. They play roles such as research assistants, big data analysts, research librarians, and operators, working in big data curation, management and administration. Furthermore, DSP act as a "bridge" between the research teams and the IT services/specialists.
- A major difference between DSP working for research or for industry is their time-to-results scope. Research DSP usually do not work under time-bound pressures, in terms of the urgency to deliver results. Their mission is to contribute to moving the boundary of knowledge forward. In the opposite situation, the goal of DSP working for the Internet industry, massive e-retailers (e.g. Amazon) or in financial services, is to "move" products forward at minimal cost and in almost real time conditions. So the ability to take on deadlines, constrained resources – even company's political climate – and push a product out in a reasonable time is a really important skill [12].

In industry, Demand players define DSP by the problems they can solve: **"**What companies are looking for is ultimately the capability to make predictions based on that data. Companies use those predictions to help drive everything from marketing strategy to resource allocation, personnel levels and staffing, or to predict retail sales."[13]**.** In the USA corporate market, advertisements asking for DSP usually describe what they have to be capable of doing, probably their previous experience, and some of the tools required. There is an implicit agreement in that proficient DSP have an advanced degree, usually a master's or a PhD, most regularly in quantitative disciplines, and are proficient users of tools such as Hadoop and languages such as R, python, and they use a variety of methods to derive useful information from data. Some of the advertisements also emphasise soft skills if they are needed. "The most successful – and consequently sought after – DSP possess that rare combination of analytical skills, technical prowess and business acumen needed to effectively analyse massive data sets while thinking critically and shifting assumptions on the go, ultimately transforming raw intelligence into concise and actionable insights."[14] Ultimately, a key requirement is that the DSP candidate must be familiar with the research methods (or business process management) to be able to discover new opportunities for companies to improve their current business or propose new data driven services. Table 1 summarises differences between industry and research oriented DSP characteristics. Details behind this synthesis are presented in Appendix 1.

[11] http://dupress.com/articles/new-big-data-analytics-skills/
[12] What makes a true Data Scientist, Thomson Nguyen
[13] Will Cukierski, Data Scientist, Kaggle
[14] http://www.itcareerfinder.com/it-careers/big-data-scientist.html

Table 1: Main differences between industry and research DSP jobs

Issues/ traits	Business world	Research world
Ultimate driving force for employers	Decision making	Improve excellence in research results
DSP Functions	Support decisions; improve operations	Support research process; optimise data handling
DSP Mobility	High between companies	Career via stipulated scale
Professional career risks	Inclined to	Avoidance of
Job satisfaction	Financially driven. Challenging environment. Dynamic	Science oriented / recognition. Stability. Self-development
Recruitment processes	Flexible	Rigid, constrained by public sector regulatory framework
Training to become fit-for-purpose	Fast. MOOC's, bootcamps	Solid. university + on-the-job training
Background	Master, PhD, MBA.[15] Industry knowledge	Master, PhD, Post-Doc. Beneficially: scientific education and subject domain specific

These differences are relevant when it comes to the design of educational and /or professional training and/or certification products and services aiming at building up a DSP skilled workforce. The analysis also needs to take into consideration the overall size of the sectors where DSP could get a job. In this regard, it is key to understand that while the **Research Infrastructures in Europe as a whole are in the order of magnitude of some 1.000 potential employer institutions (plus up to 4.000 Higher Education institutions as many of them will employ DSP for their own research activities), the data industry and the data user companies total some 1.000.000 potential employers, i.e. a labour market segment potential 200 times bigger** at least, given that the number of DSP a leading company could hire should be higher than those hired by in research teams. And the corporate world offer, in general terms, much higher remuneration packages to their highly regarded employees.

2.4 Forecasting trends in the DSP labour market

There is a strong coincidence across the desk research sources forecasting a clearly growing demand of data workers, especially in the big data domain. This growth in the needs to increase hiring data workers has been confirmed throughout our interviews on both sides of the Atlantic:

- According to the European Data Market monitoring tool[16]: The annual demand for Data workers will grow from 6.1 million (2014) to a range of between 7.6; 8.5 or even 11.8 million in 2020, depending of the chosen growth scenario. In parallel, the gap between demand and supply of data workers will grow from 509,000 jobs in 2014 up to 536,000 unfilled positions in 2020.
- The data science market will be active within the framework of the Digital Agenda for Europe. Public expenditure to develop ICTs will be doubled, and digital literacy and skills will be enhanced.

[15] Salaries of Data Scientists. The Burtch Works Study, Burtch Works Executive Recruiting. Burtch Works LLC. April 2014

[16] European Data Market SMART 2013/0063. D6, First Interim Report. Gabriella Cattaneo et al. IDC Open Evidence. Oct. 2015.

- Managers of conventional industries will have to have more analytics knowledge if they want to stay competitive. “Within 10 years, if you're not a data geek, you can forget about being in the C-suite.” [17]
- According to Forbes[18] The market for big data technology and services (will) grow at a 23.1% compound annual growth rate, reaching $48.6 billion in 2019. The larger market for business analytics software and business intelligence solutions which now includes the new disciplines of data science and cognitive computing is at least 5 times bigger.
- McKinsey[19] estimated that the US needs to retrain 1.5 million managers with the know-how to use the analysis of big data to make effective decisions. Given the comparable size of the EU labour market, we can extrapolate a very rough figure of the opportunities for DSP related re-skilling demand.
- The Open Data policy across Europe will call for professionals who can make good use of it. Awareness needs to spread. Efforts are oriented in this direction.
- With regard to the Data Scientists (as IDC renamed Big data workers), the European Data Market monitoring tool[20] estimates indicate that:
    - While “In 2014 Data scientists represented less than 2% of the estimated population of data workers in the EU counting about 119,000 workers, the estimated figure for 2020 is of 267,000 big data analyst jobs demanded”.
    - The potential supply-demand gap of 66,000 unfilled positions corresponds to approximately 17% of demand. This forecast gap at 17% is almost three times as high as the data workers' gap estimated for the Baseline scenario in 2020 (approximately 6%), even if the absolute number for DSP is tiny compared to the millions of data workers.
    - This appears to confirm the concern among Big Data stakeholders about the potential lack of Big Data skills and the need for Europe to catch up in the training and education for these skills.

[17] http://mashable.com/2014/12/25/data-scientist/#Nq7qAk0zrZqp

[18] http://www.forbes.com/sites/gilpress/2015/12/15/6-predictions-for-big-data-analytics-and-cognitive-computing-in-2016/#2a9c609a409e

[19] http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation

[20] European Data Market SMART 2013/0063. D6, First Interim Report. Gabriella Cattaneo et al. IDC Open Evidence. Oct. 2015.

3 The Employers: Demand of DSP

The explosion in data to be managed is driving employers in the different sectors to resort to DSP to adapt to the digital transformation of their businesses:

- Research organisations need to cope with data coming out of "in silico" research and make good use of it, both for their own research mission and to give back to society, for sustainability purposes. These needs are becoming mainstream but they differ from one research domain to another.
- Industry market demand requires businesses to push forward, and every minute counts to stay competitive. Employers need fast and useful answers from good, adaptable DSP who can integrate heterogeneous data and quickly yield results for the organisation.
- In public administration institutions, there is a growing weight of the so-called "evidence based policy formulation" and the corresponding KPIs, as proofs of improving services for the general benefit of society; but as it happens with the research environment, there is less sense of urgency compared back-to-back with industry and commercial sectors of the economy.

3.1 Who is hiring DSP today

Industry and commercial activities are at present the players most actively seeking skilled DSP, especially in data intensive sectors. This means for them a change in skills, leadership and the need for the organisational structure to make room for them. Companies are facing the data challenge by giving preference to people with analytical skills when hiring and promoting their professional careers, and also by developing analytical skills through in-house training, and by integrating new talent with more traditional data workers.

Companies understand that it is critical to have data analytics capabilities in order to drive digital transformation successfully. "Looking five years ahead, expectations of the level of digital integration are broadly comparable in all regions: Europe Middle East and Africa, the Americas and Asia Pacific. A large majority of companies expect to have advanced digitisation and integration capabilities..."[21]

If we take the three level classification of analytics-driven companies (analytically challenged, practitioners and innovators) more analytics driven companies attract-hire-retain the greatest number of DSP. Whereas less analytically driven companies outsource their data analysis functions either because they lack the knowledge or do not have (yet) suitable employees. In these cases, they hire consultants, or an advisory panel. Outsourcing DSP functions is a way of introducing data analysis into the company and filling knowledge gaps.[22]

In any case, it is the C-level[23] managers or senior teams who need to "solve a strategic problem" and resort to DSP. In companies where there is already a DSP in the company, there is better understanding of their roles and consequently these companies focus on recruiting correctly and retaining DSP. Larger companies where there is a Chief Data Officer (CDO) usually have a department or a small structure to look for DSP in universities, the social media or other means of scouting and eventually recruiting.

Some very important sectors find it difficult to cover positions. For example, big European energy distribution companies predict that data-science-related IT positions will be very difficult to fill, as they will not only be competing among themselves, but will have to "face off against companies from other

[21] https://www.pwc.com/gx/en/industries/industries-4.0/landing-page/industry-4.0-building-your-digital-enterprise-april-2016.pdf
[22] http://sloanreview.mit.edu/projects/analytics-talent-dividend/)
[23] C-level, also called the C-suite, is an adjective used to describe high-ranking executive titles within an organization. C, in this context, stands for chief…As compared to the skills associated with other organizational positions, which are often more functional and technical in nature, C-level executives must demonstrate leadership skills and business expertise, as well as team-building abilities, rather than functional and technical know-how. http://searchcio.techtarget.com/definition/C-level

data-intensive or more digitally mature industries that are more accustomed to attracting this type of talent"[24].

As in other industry sectors, banks and other financial organisations – despite being traditional institutions – have understood that digitisation will touch every aspect of their business operations, not only their products for the sake of benefit, but will reach human management and even the culture. The more innovative banks are already hiring heavily and foresee more.

Regarding the size of companies, managers are much better informed in large organisations than in SMEs.Technological firms and start-ups, as well as other more traditional ones like insurance companies, manufacturing, big retailers, banking and finance, in sum, organizations in every industry sector are building up their data science teams and are increasingly hiring DSP [25].

In r**esearch institutions**, DSP work together with other domain scientists and are distributed in teams in areas such as data management, data analysis, storage management, operations, software integration, federated service management, and user support. They report to mid-level management, or to the Principal investigator in the cases of grant projects. Research organizations will keep on hiring DSP. The DSP role is defined locally by organisations and may not match a formal or standard definition. These are defined by The European Normalisation Committee (CEN). Others aim at having a 20 to 30% of their staff working as DSP. All have declared that they will keep on hiring DSP, both for growth and for replacement of professionals who retire.

Diverse **public administration** bodies employ DSP in the USA: government agencies, networking hubs, and the U.S. military are hiring DSP and the Health care sector is building data sets on their electronic patient records. The function (data officer, chief data officer) exists in numerous governmental organisations such as city or state governments, or national entities**. The White House appointed the first U.S. Chief Data Scientist in 2015**, to "harness the power of technology and innovation to help government better serve the American people"[26], an example of which is the launching of the Police Data initiative, which accelerates progress around data transparency and analysis[27]. Freshly graduated DSP are hired in the USA administration: Local Government and Policing are both a really good training grounds… "some stay, most quickly move on towards much higher pay. The reason is that we have very few of us and shed loads of data! You get a chance to really flex your skills and get a massive amount of real world experience from analysing domestic violence risks to helping predict the impact of the next flu pandemic, as well as maintaining the systems and their links. While we don't pay the best, we do of course have great benefits like flexible working..."[28]. The scarce information found regarding the public administration sector in Europe leads us to consider this a territory which is yet to be developed, but where numerous professionals will be needed.

3.2 Recruitment conditions

Given the early stage in demand maturity, the companies still resort to known patterns as a first choice to recruit DSP. According to Burtch works [29] a large majority (92%) of DSP have a graduate degree, with 48% holding a PhD. and another 44% a Master's degree. At almost every level, DSP with a PhD. earn more than their counterparts with a Bachelor's or Master's degree. The most frequent degrees held by DSP are in mathematics or statistics, one fifth of them holding a degree in computer science. Significantly more DSP hold a PhD. compared to other predictive analytics professionals (48% in data science vs. 17% in predictive analytics).

[24] The Transforming Utility Industry Context — What Utilities Envision for the Future, Gaia Gallotti, Dec. 2015
[25] http://www.burtchworks.com/2015/01/12/predictions-2015-analytics-data-science-hiring-market/
[26] https://www.whitehouse.gov/blog/2015/02/18/white-house-names-dr-dj-patil-first-us-chief-data-scientist
[27] https://www.whitehouse.gov/blog/2015/05/18/launching-police-data-initiative
[28] https://www.quora.com/What-are-the-best-companies-to-hire-fresh-data-scientist
[29] The Burtch works Study on Salaries of Data Scientists (April 2015)

Some companies will accept DSP with undergraduate diplomas in an analytical concentration, such as Computer Science, Math and Statistics, Management Information Systems, Economics, Engineering and Hard Sciences. A solid base for a DSP career is provided by career-focused courses, degrees and certificates in analytical disciplines like database management, predictive analytics, business intelligence, big data analysis and data mining.

A combination of hard skills and soft skills is required, and communication and other soft skills are widely accepted as essential. Computer expertise is a must, the differential aspects, the ones that discriminate among people with similar technical proficiency levels, are the business (in the wide sense) and management capabilities and their communication skills. "A key skill involves explaining big data outcomes to executives—in visual displays or verbal narrative... "tell a story with data," and relate well to decision-makers".[30]

Competitive DSP must show strong technical background, but it has to be value delivery driven. Creativity is also demanded. In the end, non-technical Skills are also highly demanded. "Soft" skills include team work, strong communication, presentation and interpersonal skills; and these skills are sought by nearly 40 % of employers. Other key responsibilities include the capability of providing business recommendations (35%), and of "translating" business needs to data requirements for Business Intelligence. "*The number one complaint we hear from companies about Data Scientists is that they lack business knowledge and skills."*[31]

Although it is understood and accepted that soft skills are essential to the profession, the current problem is that many organisations and institutions do not test them thoroughly, because they lack the legal, managerial culture and technical tools to assess and really screen people, or to detect possible failures. Some organisations are simply not ready to assess them at all, especially in Europe. For many organisations -especially in the public administration and the research sectors- hard skills are also easier to demonstrate or prove. In public bodies recruitment procedures (which are rigid and tend to leave them unconsidered) are already set, and it is difficult to change them in the short run. However, other organisations assess them, and do not mind accepting professionals that lack these skills, on grounds such as "they do not need to communicate with the public", "there must be room for everybody", etc.

For research institutions in Europe, a Master's degree plus some (variable) experience is required. Many have a PhD. degree, covering a wide range of data related topics. They are also required to have high domain knowledge in the respective discipline. Being a new field, there are no DSP trained as such with 5+ years' experience, although many professionals graduated in computer science, mathematics and statistics, engineering and physics consider themselves DSP. At one of the interviewed organisations, they all consider themselves DSP, call themselves "engineers", and only one has a degree as such. DSP come from several disciplines, and organisations are ready to complement their education with training as long as they are scientists.

In some research organizations experience is vital, whereas other institutions prefer freshly graduated professionals with experience in their research projects. More experienced professionals demand higher remuneration or may not be willing to adapt to their particular conditions. Some institutions require somebody able to know the ontology of the domain; the skill to "translate" and so say the same things in two different languages, whether they are talking with the IT or with the researchers.

[30] Big Data in Big companies, T. Davenport and Jill Dyché, May 2013 international institute for analytics https://www.sas.com/resources/asset/Big-Data-in-Big-Companies.pdf

[31] 6 Key Career Strategy Tips for Data Scientists, August 10th, 2015. *http://www.burtchworks.com/2015/08/10/6-key-career-strategy-tips-for-data-scientists/)*

3.3 DSP recruitment processes

As stated before, the infancy of the DSP labour market is also perceived by the lack of well-established recruitment mechanisms. There is a wide variety of recruitment practices, from the strictly regulated and formal procedures that rule procurement in public institutions, to informal appointment through contacts or by word of mouth.

Insights provided by desk research -notably up-to-date articles on how recruitment processes are carried out in the USA- on the one hand, and the interviews with employers (in global companies and Research infrastructures mainly) on the other hand, show that recruitment criteria are different when working for private companies or for research and public administration:

- Regarding requirements, commercial companies ask for decision makers, able to provide fast answers and take risks, and so DSP may be remunerated accordingly. There is high mobility and candidates must be willing to change working environment, location and working conditions; they must assume greater workload and tougher working conditions. Working for the public administration, on the other side, offers security about the job and deadlines are less stringent. Remuneration, however, may not be as high, and promotion schemes are more rigid, and slower moving.
- In the industry sector, there is strong competition among companies that wish to hire DSP professionals, and given the urgency to recruit and the difficulty to verify a DSP's added value, it is a common practice to hire away ("*poach ", as mentioned in interviews)* from other companies.
- In data driven companies, remuneration policies, together with the profile required, are clearer than in challenged or lagging types of companies. Candidates' assessment risks being more superficial in companies where there is no previous experience on data science.
- 0ne of the functions of a DSP is to look for and hire other DSP, for building a team and as advisors on the best candidates to choose for a company's specific Data Science Professional. They usually make use of their own networks.
- Recruitment processes are still finding a way to ensure that the candidates approached are fit for the position[32].
- The social networks, notably LinkedIn, are used for identification of candidates. Another usual channel is through the networks of links between companies and universities. A third channel of identification are bootcamps and other training courses, professional and student associations, other companies, communities, and job fairs.
- The channels also vary according to the position level. For managerial, top salary (the C-suite) positions, where salary is higher than that of the person who does the recruitment process, a head hunter or professional contacts are preferred.
- The selection process typically contains three steps. Strong quantitative and programming skills are non-negotiable and are tested in the first place. Secondly, other skills such as problem solving and communication, which are more subjective, are checked. The final step is to assess how the candidate works in a team and fits into the culture of the organisation.
- Though not yet settled, recruitment mechanisms are being refined. Recommendations now include in the final stages of the hiring process an assignment and spending a working day at the company that is to hire the candidate. The candidate shares the environment and problems faced, and is assessed on his/her capability to face them, suggest possible solutions, and the way he/she might evolve within the company[33].
- Numerous consultancy firms and companies that develop tools and offer training services and certification also provide job services, as do DSP associations. A few consultancy companies are

[32] Flash survey Big Data professionals contacted more than ever before: http://www.burtchworks.com/2015/02/16/flash-survey-big-data-professionals-contacted-more-than-ever/

[33] http://firstround.com/review/how-to-consistently-hire-remarkable-data-scientists/

starting to become specialized in services for data driven professionals, and match job hunters with potential employers[34].

- In public institutions hiring mechanisms and processes are fixed and differ from the private sector, irrespective of the position or qualifications sought. The public administration hires according to general conditions for public officers. Candidates undergo the same process as any other candidate. Positions are advertised through fixed channels.
- Research institutions which are publicly owned also hire on fixed mechanisms, and there are cases when candidates' qualifications are made to "fit" into the official processes: CV review, exams/assignments, interviews. Problems can arise if a candidate with demonstrated higher qualifications invokes to their right to appeal if they do not agree with the selection process.
- Research infrastructures and similar organisations typically post advertisements in their own specialised pages or have contacts with universities, use the social media, or channels of individuals. The basic requirements are the competences needed for each case or research project, domain knowledge (or quick adaptability to it). Soft skills are sought in most cases. At big infrastructures, such as CERN, a lot of the recruitment is done semi-internally, as they have many programmes related to summer students, technical students, and fellows.
- Universities are associated with research institutions, and are a source of candidates either through the teaching staff who detect possible candidates, by the students' research projects at the point of developing masters or PhD degrees, or when offering places for working experience (bachelor level). In some research institutions, part of the researcher's responsibility is to keep their networks updated, as they may need to hire a highly specialised and specific profile not advertised by the normal channels. Many universities have fairly formal recruitment processes, which are coordinated through their institutional websites. Academic job portals are also used, and some research institutions have an agency for jobs.
- In research organisations, the personal interview is considered a very important step in the recruitment process, and is usually carried out by the coordinator or the principal investigator, with the presence of Human Resources staff. This interview allows the recruiter to appraise the candidate's capabilities for the specific tasks, and their communication and team working skills. In public agencies, this stance may have less weight in the process, as set procedures are difficult to change to allow for specific cases.

Insights from interviews and desk research also indicate that DSP cannot be hired following the same procedures as other, already established professions. DSP hiring has its own specific and in many cases has ad hoc nature. However, companies still use their established procedure that is in many cases defined by law or sector regulations. Criteria for recruiting varies according to the companies, the key requirements being experience working on problems with real world data, and how these problems can be overcome, ability to work in a team, and to explain technical matters and the complexity of data analysis in a way that non-technical decision makers will understand.

Assessment of recruiters about the recruitment experience highlight that on top of the lengthy search and hiring-selecting processes, it takes 3 to 6 months for a recently hired professional to integrate into the organisation. Although the conditions vary depending on the type of company, position, responsibility, the hired professional is expected to start producing in a short period. Employers need DSP to start being productive from the beginning. However, some organisations feel they are not fully integrated before a year. Even when a DSP is hired out of a competing company, where the way he/she works and what can be expected of him/her is known, there is still a need to adapt, so extra training is brought in. It is common for successful DSP to keep updated for continuous professional development, self-learning and adopting new technologies and tools.

[34] where data scientists and employers find the best fit (http://correlation-one.com/)

Research institutions' criteria vary according to the grant given to each research project. Research done in the area is preferred to experience in the case of young professionals. Enthusiasm about the work is also sought. For more experienced or senior positions, the demand is for a scientist who is settled in his or her discipline, with a strong IT affiliation that makes him or her quickly adaptable to the new conditions. Needs have to be described correctly and attractively, especially in research institutions where working conditions differ from a "9 to 5" job. Description has to clearly state what makes it different from working elsewhere in order to make the position appealing to possible candidates.

4 The Supply Side: Education and training of DSP

4.1 Characteristics of the Supply side of the DSP labour market

The education and training of DSP is being carried out by universities, other training centres and institutions, and training units inside companies and institutions. There are also "self-made" DSP that have not received formal education as such, but perform data science tasks and actually become a source for covering job demands.

There are some 4.000 Universities in Europe, but only a short minority offer some sort of educational programmes directly contributing to build up a DSP workforce. As can be seen in EDISON Deliverable D2.2[35], there are over 110 universities and higher educational institutions offering programme studies in Europe, and over 110 in the USA. Many of these universities offer more than one programme. Interviews with Higher Education institutions experts coupled with desk research indicate that universities are continuing to open new study programmes every term, and that programmes are being extended and diversified. Data science is more and more extended to other sciences, medicine, arts, and humanities. Universities that offer Data Science programmes are mostly those that were already offering computer science and mathematics, and also schools of engineering. They are mostly located in the areas that require this type of graduates, notably industrialised ones. Most data science courses are built on the basis of existing strengths and aim at developing wide sets of skills and competences.

Universities offer masters programmes, building on Science, Technology, Engineering and Mathematics (STEM) disciplines. Scientifically oriented master degrees do not require 1-2 years' work experience in contrast to business masters courses. As one of the interviewees said, masters courses "do not ensure you the worker you need because master courses are very generic". However, most of the interviewed institutions state that among their students there is from 90 to 100% employability. Universities set students up on their professional career paths. According to our interviews nearly 100% of data science graduates enter the job market, sometimes even before they finish studies. These are preferred for research but do not offer much domain experience for the business world. Universities can offer distance, online and face-to-face education, and combinations of all of these. Some courses are eligible for certification, and some of the certifications can count towards a degree.

Europe's higher education institutions operate within the legal and administrative frameworks of their national or regional higher education systems where considerable diversity remains. Educational systems and institutions retain their own characteristics, although harmonization is being achieved through the Bologna process and the EHEA (European Higher Education Area) to ensure comparability in standards and quality.

The most frequent Data Science studies are at master's level. Good STEM knowledge and skills are required for applying for most of them. There are other, more varied and flexible master courses aimed at a wider range of educational backgrounds such as human centred multimedia, applied games, marketing science, still requiring strong affiliation to computer science disciplines.

Another growing trait is interdisciplinarity. Universities generate data science programmes building on their strengths, and involving several departments in their development. In this way, other than the traditional Computer Science, mathematics or statistics departments originating the core of these studies, departments such as Economics, Biology, Chemistry, Graphic Design also participate.

Universities and other training organisations are offering new Data Science courses in a flexible way to suit different publics. Existing/developed courses combine online and face-to-face/residential

[35] Edison D2.2 Existing Educational and Training Resources Inventory and Analysis.

education. Some universities offer accelerated formats and the possibility of taking full or part time studies. Flexibility is achieved by creating modular courses that can be combined in a full Data Science programme, and courses are offered as single courses or included into other degree studies. The offered courses may have several starting dates throughout the year and flexible online schemes for working professionals. Another form of flexibility is to offer education with different entry requirement levels.

Training centres (physical and virtual). Given the urgent need for DSP, it is accepted that not everyone has to go through university studies to become a DSP. There are training organisations which offer a wide variety of courses and training combinations oriented to fulfilling the different sectors of demand: some of these are associated to software developers and consultancy companies and tend to offer training on more technical skills. Some organisations offer courses with different combinations of individual modules lasting an average of one hour each. Others offer courses followed by comprehensive exams. They offer a wide variety of short and long courses and training usually geared to graduates (engineers, mathematicians, statisticians…) who wish to specialize and fast track their Data Science Professional career. Outside of universities, providers of specific tools such as Hadoop, Spark, NoSQL, SQL, R, Python, Matlab, SAS, and Tableau are other focus of learning. Cloudera, IBM, Microsoft, SAP, and other well-established institutions offer professional courses which are highly sought and well-respected.

There is also a wide range of online courses available, ranging from MOOCS to commercial offerings. A landscape of the offer can be seen in EDISON Deliverable D2.2. Many of these organisations provide complementary education to support “self-made” DSP, who want to obtain the set of competences and skills to pursue a Data Science career themselves.

Training centres and in-house training are clearly geared to industry and constantly adapt their curricula to demand, aiming at on-the-spot solution of business issues. Companies’ Human Resource departments/services also organize their own courses, whether hired from training centres or tailored to the company’s needs. Private and professional organizations offer other training with Data Science related certifications. The leading analytics software providers offer certificated courses, mostly associated to their products. Consulting companies and private enterprises that already offer analytics and data management services offer additional Data Science training as a complement to their services.

There are solutions such as the immersive self-styled bootcamps which offer intensive training during 10 to 12 weeks, with demanding admission conditions and also high fees. The offered training is mostly practically oriented and ensures that participants have the necessary knowledge and skills to perform DSP’ tasks and theoretically easily find a job.

The leading training centres that offer these courses are usually in close contact and relationship with the industry sector and facilitate links between students and industry and offer hands-on training. Examples of these are Bit Bootcamp, Data incubator, Data Science Dojo, Metis, Zipfian, Data Science Europe. They guarantee success, offer job assistance via associated hiring firms or hiring partners, and some even offer to refund amounts paid if the student does not get a job. Their success is based on the fact that they ensure a job position relatively quickly (within 2 months or so), the size of their classes is small, they have expert mentors who can teach, they use tested material and usually include in their services putting students in contact with companies seeking Data Science professionals. At the same time, fees are usually high. A certificate of attendance is given, not always a degree. Training periods are variable (2 weeks to 2 months). Training centres and in-house training are clearly geared by industry demand and constantly adapt their curricula to adjust to changing demand and specific solutions for business. Companies’ Human Resource departments/services also organize their own courses, whether hired from training centres or tailored to the company’s needs.

In-house training and self-made DSP: Numerous computer science professionals or coming from STEM disciplines have opted for building themselves a curriculum to cover the skills required to perform data science functions. Many DSP (the information we have comes mainly from the USA) build up their careers from different backgrounds, starting with a master degree, most often in STEM disciplines, and use continuous further professional (self-)education by using available online classes, books, and regularly dedicating time to studying the available content. The self-made Data Scientists, who come from diverse sectors/domains also need to be addressed by providing targeted training, professional improvement and also certification. Self-made DSP, who have strong skills in mathematics, data handling, computer science, programming languages and statistics are the best candidates to get Data Science positions in industry. Certification mechanisms will have to recognise them in order to keep them in the market. Re-skilling and professional development solutions have to be found for these market players.

As mentioned by interviewed corporate employers, these professionals, in some cases supported by their companies, resort to universities (under bilateral agreements between academia and corporations) and other training centres to keep their skills updated, complement them and remain competitive with more certified DSP that start graduating from universities. Up to the present moment many DSP positions are covered by self-made DSP, who are well placed in the market. Certification mechanisms will have to recognise them in order to keep them in the market. Re-training or updating solutions have to be found for these market players. In the USA a high percentage of companies (63%) are providing formal or on-the-job training in-house.[36] We do not have the corresponding information for Europe.

The advantage of training in-house is that staff already know the business. Efforts are made to integrate DSP with the traditional data workers dealing with databases and data management, but developing insights into business process management and continuous improvement remains among the main goal of implementing Big Data and Data Science technologies by companies [37].

4.2 Current market gaps and future evolution

The most significant gap is the lack of capacity of the whole educational training and certification system to ensure the timely delivery of competent DSP both in terms of quality and quantity The solution for this strategic failure can´t be solved quickly, and furthermore, below the surface there are diverse other issues to be addressed:

a. Universities seem to be offering the same training to all students irrespective of the market sector they are targeting when looking for a professional career. Some universities mention in their promotions that graduates can go to the public sector, or deepen into research by progressing on to PhD programmes. Although universities are actively working to design study programmes and launch new courses, they are the stakeholders that currently play the least within this market. Universities focus more on their mission, which does not necessarily include looking at the future job market for their graduates. These existing gaps cannot be solved in the short term through the conventional higher education and training channels. Active participation from other specialised training - both in-house and specialised training companies - is required.
b. Universities need to stay active in this market, and increase education in quantity and variety, they need to adapt quickly.
c. Some of the challenges that universities face in relation to curriculum design include an incomplete identification of the real needs of the demand side and students' expectations, lack of flexibility in curriculum design and lack of revision of the curriculum to match changing needs; not sufficient

[36] The Talent Dividend Sam Ransbotham, David Kiron, Pamela Kirk Prentice. Findings from the 2014 Data & Analytics global executive study and research project. MIT Sloan Review, in collaboration with SAS. Spring 2015

[37] http://www.forbes.com/sites/gilpress/2015/04/30/the-supply-and-demand-of-data-scientists-what-the-surveys-say/#6d95f828205e

connection between new Data Science curriculum developers and the previous education; coordination and best practices exchange in curriculum development.

d. There is not enough collaboration between HEI and employers to ensure that all students can have hands-on practice. The number and range of agreements with industry and companies, where students could do their work experience and develop their practical projects working on real problems should be widely enlarged. This would increase the graduates' employability.
e. Many universities offer PhD programmes in a framework of cooperative projects with research infrastructures and other research organisations. This constitutes a good source of future DSP and eventually future industry leaders. PhD degrees (in Data Science) will create wider opportunities for students to continue in research and to move to industry (for higher paid jobs). There have to be positions for PhD students, and universities have to guide them adequately. As one of the interviewees stated, the institution has to make clear to the student that "preparing for a well-paid job is a perfect motivation, (but) that is typically not a good motivation for a research track."
f. The need for an interdisciplinary approach to the integration of technology in areas such as finance, banking, healthcare, medicine, energy or defence motivate requirements for vertical education in Data Science. "To be effective, a Data Scientist curriculum has to be closely linked to commercial and societal needs"[38].
g. The wide variety of courses and programmes reflects the existing differences in the understanding of what a Data Science Professional is and has to be, in spite of the National Institute of Standards and Technology[39] attempts to define Data Science and Big Data. This lack of a unified definition has consequences in the market and it is difficult for universities to fit their offer to what employers are looking for. The analysis of the current Data Science programmes (both European and non-European) revealed that the first implemented programmes have imbalances between different components that constitute the major Data Science competence areas (as specified in the EDISON Data Science Framework). Following best practices in competences/outcome based learning and related educational models will help combine strong subject oriented academic education and the practical training sought by employers.
h. Data Science (or associated) education and training are at present primarily concentrated in the industrially advanced and data driven countries. It is explained by both local demand from companies and strong requirements to technological base availability.
i. Universities need approval of programmes by higher academic bodies in order to get resources to run the programmes (teaching staff, administrative and support staff, logistics and lab capacities, etc.) and to adjust to fast changes in technologies.
j. The introduction of Data Science related courses at undergraduate level is complicated because of a stronger need for course harmonisation among departments and ultimate need for accreditation by the corresponding national accreditation body. The introduction of changes in undergraduate programmes is complex and takes longer (from 6 months to 2 years) to be approved.
k. When opening new Data Science related programmes, universities face problems on how to balance between general or core courses (related to the technology foundations) and specialised courses that are oriented to the companies' particular needs. This also includes a need to differentiate from other studies, e.g. in the case of several masters at one university. The curriculum/programmes should be organised in such a way that they combine both the foundational Data Science courses and the domain specific science courses. Some universities are already offering different tracks: at least research and business oriented. Students are advised on which track to follow, whether to start a career in business with notably higher salaries, or follow a research track. Strong demand for and shortage of expert DSP also affect universities by creating difficulties in hiring (or growing professionally in-house) teaching staff who can do research and also teach the growing number of degree courses that universities are opening.

[38] Big Data and Data Scientists, it's an issue of degrees. Tom Groenfelt http://www.forbes.com/sites/tomgroenfeldt/2012/01/21/big-data-and-data-scientists-its-an-issue-of-degrees/#34e75d697fb8

[39] NIST Request Comments on Its Big Data Framework Development. http://www.nist.gov/itl/bigdata/20150406_big_data_framework.cfm

l. Strong requirement for future DSP to have practical skills and domain specific knowledge for working with Big Data technologies and domain specific data impose departments that plan to open Data Science programmes and/or tracks to provide sufficient facilities and teaching mentoring staff to assist the students in acquiring practical knowledge and skills. This can be achieved either by building capacities at universities or involving project advisers and mentors from cooperating external organisations and companies that can also provide place for staging practice or internship. The above factors define the size of new programmes that university can offer maintaining the quality of education, while extending them in the future will require solving scalability problems.

Ultimately, even though universities primarily aim at educating for the long term, rather than present day problems, they make a strong effort to adapt to real world needs and problems. When generating or designing courses, companies are present in the advisory boards to contribute with indications that help alignment with organisations' needs. The industry sector also contributes by offering places for work experience or projects, and asks for certain content or tools. They do not usually decide on concepts or content. Changes in curriculum are easier to make (and have the approval by the accreditation body) at Masters level than at undergraduate level. Decision makers (masters' programme director, manager) solve curriculum design by describing content in a general way, to allow for changes; or by defining the course as a specialisation course.

In spite of these shortcomings, data science education and training offerings will need to expand and improve its fit-for-purpose traits:

1. Companies need to find faster ways to bring the DSP they need to their companies, so online solutions and MOOCs will continue to be used, together with in-house training. Education levels in data science will be different, so mechanisms will be needed to assess skills. Until supply is able to provide more professionals, salaries will keep increasing and market will regulate deficit. Acting DSP and practitioners will need to keep the professional knowledge and skills up-to-date in this quickly evolving field. "Both certificate and degree programmes are needed. The process of incorporating Big Data into the operations of business, government, and education will require hundreds of thousands of new, specially trained knowledge workers."[40]
2. The education and training offering will keep growing to keep professionals updated regarding the skills to work with unstructured data, and as more graduates pursue their careers, we will start having information on the effectiveness of these quick study programmes. The number of education programmes will increase and there will be a variety of training options to cater for the different sectors in the economy, both in formal education and informal options such as MOOCs, bootcamps, etc. Part time studies are seen as an option for people who are already working and want to become DSP, or for companies who need to re-train their staff. Still, universities will continue to provide strong quantitative foundation.
3. At the current technology development pace and changing market demand, companies treat university studies as a "too slow" solution. However, they are interested in continuous flow of future DSP on junior and stage positions. And universities also need to adjust and balance study programmes, establish more clearly which skills employers want, and improve work experience opportunities.
4. Many students seem not to be aware about career opportunities lying in Data Science, as it is relatively new. Data Science study calls for efforts that require strong motivation from students. Although some efforts have been detected, universities will have to focus on marketing their new programme offerings to reach more students. This includes among others, flexible accreditation formulas for recognition of different study levels.

[40] Unleashing the potential of Big Data, white paper: http://www.essi.upc.edu/~aabello/publications/13.Whitepaper.pdf

5 Conclusions: opportunities for exploitation of EDISON outputs

The general conclusion of the analysis we have performed so far is that **the DSP labour market is incipient, in its infancy stage and therefore shows all the distortions which are typical of an emerging market.** The positive aspect of this situation is that the market gaps depicted in the previous chapters offer diverse opportunities for exploiting EDISON's outputs while collaborating in building up the Data Science profession.

5.1 Main DSP labour market gaps

The lack of a proper functioning of this market can be summarised as follows:

1. The growing quantitative gap between the number of DSP needed and the professionals that the education, training and certification suppliers are able to offer. At the same time, **there is a low usage of DSP in many key industrial sectors**, in the research ecosystem and especially in the public administrations, even in the most advanced economies and particularly in Europe.
2. The lack of practical common understanding among labour market players on what a Data Science Professional is, or worse, what value DSP can deliver for those who hire them. Within an organisation, there need to be data workers that can interpret data at all levels, from the clerical level to the executive level. All positions have to be catered for. And in order to achieve this, data literacy has to be instilled and oriented from an early age. **There is the opportunity of raising awareness and acting towards data literacy**.
3. The existing differences in the definition of what DSP have to be or how they benefit those who hire them has negative effects in the labour market:
    a. Employers do not know how to specify DSP jobs, have not developed efficient recruiting processes and need support to integrate them into their organisation when they have finally hired them.
    b. The EU offering of education, training and certification is insufficient to cope with the likely growth in DSP demand.
    c. The definition of the educational offer not necessarily coincides with the demand's needs, and the mechanisms to understand them and build fit-for-purpose offerings seem quite insufficient.
    d. The flows of potential candidates to become DSP through EU formal education channels seem to be insufficient.
4. Education, training and certification players need to ensure their courses appeal to students as well as to employers. There is an opportunity to seriously improve the attractiveness of the profession, as companies progressively understand the value of data and will be prepared to reward those able to extract this value. There is a need for generating, collecting and disseminating success stories in DSP hiring.

The **urgency to mitigate the identified deficits and support strategic changes can only be addressed by aligning interests and efforts from the demand and the supply sides of the DSP labour market**. To date, there is scarcely visible progress in this convergence, largely due to the lack of relevant dissemination and DSP community mobilisation processes. **An ecosystem providing these mechanisms is needed as soon as possible; and the exploitation of EDISON results can provide its building blocks.**

5.2 Market-driven guidelines for the sustainable exploitation of EDISON outputs

Based upon the conclusion of our analysis, the following subsections comprise preliminary suggestions for exploitation aimed at helping the market to better operate, by delivering products and services based on the EDISON project results.

5.2.1 Exploitation guidelines targeted at supporting DSP Demand side market players

1. Help companies with no previous experience on data science to improve the definitions of their needs regarding deployment of Data Science teams.
2. Provide advice on recruitment techniques to appraise DSP candidates' fit-for-purpose potential
3. Offer engagement mechanisms between Employers and DSP to make identification, recruitment and selection more productive and cost-efficient
4. Provide prospective DSP with information on DSP potential as a professional career, and about educational, training and certification offerings.
5. Provide links with Supply side providers to place students for getting DSP work experience.
6. Shape EDISON results to the specific needs of the different employers´ segments.
7. Facilitate Employers-Supply institutions agreements on practical traits of DSP career paths.
8. Offer advice and tools for companies to carry out in-house training.
9. Tailoring training and certification services to sectors/companies' specific characteristics.
10. Offer training and certification services to "self-made" DSP, irrespective of their background.

5.2.2 Exploitation guidelines targeted at supporting DSP Supply side market players

1. Provide advice and support to help universities to understand employers´ needs and adapt their DSP oriented educational offering to these unmet needs.
2. Help European universities set up Data Science degree courses using the Competences Framework (CF-DSP) and the Body of Knowledge (BoK) as a source.
3. Provide advice on improving universities marketing: Collaborate with universities' market positioning in the DSP space.
4. Approval of content is important for many universities. Help them work on the process.
5. Give universities market intelligence on recruiters and students (i.e. give information about their customers' employers); and about other educational offering to compete or collaborate with.
6. Services to help universities´ staff prepare the business cases for their courses.
7. Dramatically improve linkage of universities with possible employers.
8. Work on Universities' strengths to ensure quality education and contribute to the body of research that supports them as institutions as well as a founding ground for future teaching staff.
9. Contribute to adapt universities' DSP admission criteria as the starting point of the whole Data Science value chain.
10. Make EDISON offering compatible with EU diversity of cultures, languages and legal frameworks, thus reducing current market fragmentation, at least in the DSP space.

References

1. GMAC (Graduate Management Admission Council) in partnership with MBA Career Services & Employer Alliance**, Corporate Recruiters Survey, 2014 Survey Report**
2. GMAC (Graduate Management Admission Council) in partnership with MBA Career Services & Employer Alliance, **Corporate Recruiters Survey, 2015 Survey Report**
3. **Salaries of Data Scientists**. The Burtch Works Study April 2015, Burtchworks Executive Recruiting, Linda Burtch, Managing Director.
4. **Getting Value from your Data Scientists** by Jeanne G. Harris and Vijay Mehrota, I MIT Sloan Management Review, fall 2104 vol.56, No 1.
5. **The Graduate Market in 2015.**Annual review of graduate vacancies and starting salaries at Britain's leading employers. Produced by High fliers. High Fliers Research Ltd. http://www.highfliers.co.uk/download/2015/graduate_market/GMReport15.pdf
6. **How to Hire Data-Driven Leaders**. Michael Fitzgerald MIT Sloan Management Review
7. **What Makes a Data Scientist Great?** MIT Sloan Management Review, Special Collection. Fall 2015. http://www.sas.com/content/dam/SAS/en_us/doc/research2/mit-what-makes-data-scientist-great-107969.pdf
8. **Survey: What Employers Are Looking For in a Business Intelligence Analyst** June 27, 2014 by **Abe Selig.** In: Plotting Success. Insights on how Business intelligence can create intelligent businesseshttp://www.plottingsuccess.com/what-employers-look-for-in-bi-analyst-0614/
9. **The Supply And Demand Of Data Scientists: What The Surveys Say** by Gil Press, Contributor Forbes, Apr30, 2015 http://www.forbes.com/sites/gilpress/2015/04/30/the-supply-and-demand-of-data-scientists-what-the-surveys-say/
10. **Building Strategic Advantages with Machine Learning,** by Ray Villeneuve, Skytree, the Machine learning company.http://www.gartner.com/technology/media-products/newsletters/skytree/1-2BI2G9P/index.html
11. **Staffing Data Science Teams**, Sky Tree White Paper.http://www.skytree.net/resources/
12. **The Data Scientist of the Future,** Robert Dutcher**.** April 16, 2015. Skytreeblog http://blog.skytree.net/2015/04/16/the-data-scientist-of-the-future/
13. **IT career Finder - Data scientist** Comprises skills & responsibilities, salaries, educational requirements and trainingprogrames**http://www.itcareerfinder.com/it-careers/big-data-scientist.html**
14. **How do you recruit Data scientists? John Salvatore et al** https://www.quora.com/How-do-you-recruit-data-scientists
15. **The struggle for strategic planning** in European higher education: the case of Portugal. Maria de Lourdes Machado (Center for Research on Higher Education Policy (CIPES)) and James S. Taylor (University of Aveiro, Center for Research on Higher Education Policy (CIPES).In: Research in Higher Education Journal
16. **Data scientists vs data analysts. Why the distinction matters**. Thomson Nguyen, October 2, 2015http://blog.import.io/post/data-scientists-vs-data-analysts-why-the-distinction-matters
17. **International Trends in Higher Education 2015.**Univeristy of Oxford. University of Oxford International Strategy Office, 2015.
18. The complete list of data science boot camps and fellowships http://www.skilledup.com/articles/list-data-science-bootcamps
19. **The Talent Dividend** Sam Ransbotham, David Kiron, Pamela Kirk Prentice. Findings from the 2014 data & analytics global executive study and research project. MIT Sloan Review in collaboration with SAS. Spring 2015
20. **Status of the Chief Data Officer**, an update on the CDO role in organisations today. By Tony Shaw, John Ladley and Charles Roe. White paper sponsored by Paxataand CA Technologies, Dataversity, 5 November 2014
21. **Higher education in the 21st century: Meeting real-world demand.** Written by Lauren Barack and edited by Steven Mintz and Christine Emba**.** The Economist Intelligence Unit research report, sponsored by Academic Partnerships**.** March 2014
22. **Business model generation**, Alexander Osterwalder & Yves Pigneur. Design Alan Smith. Editor TimClark.2009 Self published
23. **Curriculum Design Requirements and Challenges of the Learning Society Approach** Sedighe Karimi, Ahmad-Reza Nasr, Mostafa Sharif Department of Curriculum Planning, Faculty of Education, University of Isfahan, Isfahan, Iran. Online Published: October 15, 2012http://www.ccsenet.org/journal/index.php/jel/article/view/19032/14239
24. **Graduate Programs in Big Data Analytics and Data Science**. Last updated: July 29, 2015 http://whatsthebigdata.com/2012/08/09/graduate-programs-in-big-data-and-data-science/
25. **Local and regional government in Europe Structures and competences CEMR** http://www.ccre.org/docs/Local_and_Regional_Government_in_Europe.EN.pdf
26. **6000 Companies Hiring Data Scientists,** Vincent Granville. December, 2013 **http://www.datasciencecentral.com/profiles/blogs/6000-companies-hiring-data-scientists**
27. **Discussions on Digital: the new war for talent.** McKinsey Digital, October 2015
28. **Higher Education Market Leadership.** A look at how higher Education institutions used research across 2013 to make high impact decisions. Hanover Research. Hannoverresearch.com
29. **The art of hiring data scientists** Sara Vera, October 8, 2015http://blog.import.io/post/the-art-of-hiring-data-scientists
30. **Thriving in an increasingly digital Ecosystem**. P. Weill, S.Woerner. MIT Sloan Management Review. Summer 2015.

31. **How to hire data scientists and get hired as one**. Derrick Harris, Apr 13, 2013. https://gigaom.com/2013/04/16/how-to-hire-data-scientists-and-get-hired-as-one/
32. **DESI Digital Economy and Society Inde**x. EU Fact Sheet http://europa.eu/rapid/press-release_MEMO-16-385_en.htm
33. **Rutgers Discovery Informatics Institute: RDI2**. http://www.dataversity.net/data-science-education-rutgers-business-oriented-approach/ http://rdi2.rutgers.edu/
34. Why your Company needs a Data Hero (and how to find one) http://www.dataversity.net/why-your-company-needs-a-data-hero-and-how-to-find-one/
35. University statistics UK http://unistats.direct.gov.uk/Subjects/Overview/10007152FT-BSDSCADF
36. Tracking the Data Science talent Gap, Alex Woodie; March 25, 2016 http://www.datanami.com/2016/03/25/tracking-data-science-talent-gap/
37. New emphasis on data science valuable to students in all fields. *UC Berkeley Chancellor Nicholas Dirks.* http://www.dailycal.org/2016/03/01/346188/
38. **So you wanna be a data scientist? A guide to 2015's hottest profession** by Todd Wasserman, December 2014 http://mashable.com/2014/12/25/data-scientist/#Nq7qAk0zrZqp
39. **European Data Market SMART 2013/0063. D6 – First Interim Report**. Gabriella Cattaneo et al. IDC, Open Evidence, 16 October 2015. http://www.open-evidence.com/wp-content/uploaDSP/2014/11/European-Data-Market-D3.2_12.07.2014_Revisedv.2.pdf
40. **Selling the Value of Data Science, Governance or Analytics. Kimberly Nevala** , April 20, 2016 http://www.dataversity.net/back-basics-selling-value-data-data-science-governance-analytics/
41. **How to Be a Data Scientist: Data Science Skill Development.** Paramita Ghosh / April 7, 2016 http://www.dataversity.net/data-scientist-data-science-skill-development/
42. **Data Scientis**t. Data Scientists analyze business data for actionable intelligence. **http://www.itcareerfinder.com/it-careers/big-data-scientist.html**
43. **From Bricks to clicks. The potential of data and analytics in higher education**. Higher Education commission, UK. January 2016. http://www.policyconnect.org.uk/hec/sites/site_hec/files/report/419/fieldreportdownload/frombrickstoclicks-hecreportforweb.pdf
44. **Big Data: Getting a better read on performance,** Jacques Bughin. McKinsey Quarterly, Feb 2016. http://www.mckinsey.com/industries/high-tech/our-insights/big-data-getting-a-better-read-on-performance
45. **Industry 4.0: Building the digital Enterprise** 2016 Global Industry 4.0 Survey PwC https://www.pwc.com/gx/en/industries/industries-4.0/landing-page/industry-4.0-building-your-digital-enterprise-april-2016.pdf
46. **Not Just a Title: How to Identify a Data Scientist** May 11th, 2015 http://www.burtchworks.com/2015/05/11/how-to-identify-a-data-scientist/
47. **Big Data Predictions – Linda Burtch 2015 Predictions for Analytics & Data Science Market you tube video: https://www.youtube.com/watch?v=-raUCus9FY0&feature=youtu.be**
48. **Guide to Data Science Bootcamps** http://www.mastersindatascience.org/schools/data-science-bootcamps/
49. **You're hiring the wrong data scientists** http://www.computerworld.com/article/2929766/data-analytics/youre-hiring-the-wrong-data-scientists.html
50. **Predictions for the 2016 Analytics & Data Science Hiring Market February 8th, 2016** http://www.burtchworks.com/2016/02/08/predictions-for-the-2016-analytics-data-science-hiring-market/
51. **Five Data Science Predictions for 2016 Ian Swanson, Editorial Team, December, 2015 by, http://insidebigdata.com/2015/12/21/five-data-science-predictions-for-2016/**
52. **Big Data Analytics: Assessment of Demand for Labour and Skills 2013-2020**. SAS /The Tech Partnership, October 2014**. https://www.thetechpartnership.com/globalassets/pdfs/research-2014/bigdata_report_nov14.pdf**
53. **UNLEASHING THE POTENTIAL OF BIG DATA** A white paper based on the 2013 World Summit on Big Data and Organization Design. Interdisciplinary Center for Organizational Architecture Aarhus University. IBM Paris-Sorbonne University, Interdisciplinary Center For Organizational Architecture **http://www.essi.upc.edu/~aabello/publications/13.Whitepaper.pdf**
54. **LERU Roadmap for research data LERU research data working group. Advice paper no.14 December 2013 league of European Research Universities http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf**
55. **Dirección de Mercadotecnia. Análisis, planeación ay control. Philip Kotler, Diana, México, 1985, 4ª edición (Marketing Management, 1980)**
56. **Lifecycle Strategy: Product, Market, Execution Fit**. Lex Sisney, October 18, 2011. http://organizationalphysics.com/2011/10/18/lifecycle-strategy/
57. **Gartner Reveals Top Predictions for IT Organizations and Users for 2016 and Beyond**. http://www.gartner.com/newsroom/id/3143718
58. **For Data Scientists, What's in a Name Really Matters.** Alex Woodie. March 29, 2016. http://www.datanami.com/2016/03/29/data-scientists-whats-name-really-matters/
59. **5 IT industry predictions for 2016 from Forrester and IDC**, Bernard Golden CIO Nov 20, 2015 http://www.cio.com/article/3006976/it-industry/5-it-industry-predictions-for-2016-from-forrester-and-idc.html

60. **What's the Big Data? 7 Predictions for 2016 from IDC**. Dec 6,2015 https://whatsthebigdata.com/2015/12/06/7-predictions-for-2016-from-idc/
61. **Data Scientists Still Hot, Salaries Cool Off**, May 28, 2016 https://whatsthebigdata.com/2016/05/28/data-scientists-still-hot-salaries-cool-off/
62. **Digitising European Industry Reaping the full benefits of a Digital Single Market, Communication from The Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions**. European Commission, Brussels, 19.4.2016 COM(2016) 180 final https://ec.europa.eu/digital-single-market/en/news/communication-digitising-european-industry-reaping-full-benefits-digital-single-market
63. **Big Data, the management revolution**. Andrew McAfee and Erik Brynjolfsson. HBR. Org October 2012 http://www.rosebt.com/uploaDSP/8/1/8/1/8181762/big_data_the_management_revolution.pdf
64. **Wait – Is this Person even a Data Scientist?** Paul Von Bünau. May 13, 2016 http://dataconomy.com/wait-person-even-data-scientist/?mc_cid=d94edb9162&mc_eid=466f9bb071
65. **3 Reasons why People, not Robots, are Key to Data Science,** Ron Bodkin · February 11, 2016. http://dataconomy.com/3-reasons-why-people-not-robots-are-key-to-data-science/
66. **17 Predictions About the Future of Big Data Everyone Should Read**. Bernard Marr Mar 15, 2016 http://www.forbes.com/sites/bernardmarr/2016/03/15/17-predictions-about-the-future-of-big-data-everyone-should-read/#1928e55c157c

APPENDICES

Appendix 1: Differences between industry and research DSP jobs

The following table illustrate/exemplifies some basic differences between DSP working for industry versus those working for public sector research organisations. These have been extracted from trade articles and from EDISON interviewees. These differences have no proper statistical value, but provide mainly direct opinions coming from real world DSP labour market players.

issue	Industry	Research Institutions
Job description	• Perform data-mining, modelling and hypothesis generation in support of high-level business goals. Successful DSP are required to have strong aptitudes for understanding and even providing business insights. • Need strong oral & written communication skills to present data as a concise story for diverse audiences. DSP often work in heterogeneous team settings, with non-technical managers, graphic designers, and experts in the company's products or service.	• They work side by side with researchers, help with experiments to get more of the data, also design tools for the lab, scripts and so. • 2 roles: a) serve scientists to maintain operations b) develop infrastructures to provide the services that scientists need. A DSP is a "bridge" between the scientist and the IT staff. • DSP are hired on project basis in many cases. • DSP work is interdisciplinary but among scientists. Contact with non-scientific and/or technical staffs not usual or even minimal, making communication skills become less demanding.
Profile	• Major issue: Inconsistent use of the "Data Scientist" title • Education Requirements (Burtch: "Approximately 40% of DSP positions require an advanced degree, such as a Master's, MBA or PhD". Some employers accept undergraduate diplomas in analytical fields.	• Masters; PhD; post-docs. • Computer scientists need to have a good amount of knowledge in their specific discipline in their bachelor or master's degree). • Given the interdisciplinary science environments, "translation" is one of the main skills required.
Soft skills: empathy, communication and leadership	Critical factor in heterogeneous settings e.g. when reporting to not technical management or interaction with, e.g. marketing or financial co-workers.	Limited relevance due to work on project basis and the rather homogeneous peer-to-peer work environment
Mobility across employers	Highly dynamic	• Low, but uncertainty in keeping jobs given the project-by-project contracts • Problems in retaining professionals (salaries)
Salary	• Varies among industries • Higher in USA than EU	• Well defined salary scale, given public sector rules • Low compared to industry and to researchers
DSP expecta-tions	• job satisfaction included challenging environment	Stability/Expect paced growth in hierarchy/Regularity. Recognition of their contribution to science.
Recruiting channels	• Social networks for screening • Personal contacts • Head hunters for top jobs	• Set by Research Project leaders • Personal contacts versus Public sector hiring rules
Report to:	Top or mid-level management	Principal Investigator /Research team manager
limitations to hire	• Lack of candidates • Lack of experience recruiting • Mobility/lack of loyalty	Positions and salaries scarcely attractive
training	Bootcamps and MOOCs	Seniors training juniors

Appendix 2: Methodological details

As stated in Chapter 1, this appendix provides operational details of the tasks performed for gathering basic information for the market analysis performed by Inmark. It includes the list of subjects (Interview guidelines) that guided the in-depth interviews for both the demand and the supply sides. They comprise a list of issues commonly confronted, and were not meant to be answered fully and in great detail, but rather to move interviewees to express themselves on the issues that most affect them and which consequently have to be taken into consideration for describing the market.

We also include the list of experts actually interviewed, including their names and the institutions they work for. They belong to institutions that are already working with data science education and/or are already employing DSP, and in the supply side have experience in the design of novelty courses or data science courses. Interviews lasted between 30 and 90 minutes and were carried out either face-to-face or remotely, between March and May 2016. Finally, we also include relevant documents reviewed in the desk research tasks.

Interview guidelines

a) Demand side: Current/future DSP employers

1. Why do you need DSP? What drove you to hire the first DSP? When was this? What difference does it make to hire DSP in your organization? What functions do you expect DSP to fulfil in the workflows of your organisation?
2. How do they work? In teams? Distributed throughout the organisation? Which areas do they work in? What level positions are DSP in? Who do they usually report to?
3. Do the different areas in the organisation require DSP with different specialisations? For which areas, tasks, roles are DSP recruited? What positions do DSP fill? How many years' experience do your DSP have in average?
4. How many DSP currently work in your organization? Out of how many employees total? Do you have all those you need? (unmet need) Main problems encountered when you try to hire a Data Science Professional?
5. Does your organization have any specific regulations to hire DSP? Would it be a problem if hiring a Data Science Professional implied paying him/her a salary well above the rest of the personnel?
6. How many DSP do you plan to hire in 2016-2018? How many DSP per 100 employees this number represents?
7. How will the situation of DSP in your organization change in 5 years' time? According to our organisation's plans will you keep hiring DSP in the future? When will this trend consolidate/stabilize or even slow?
8. How will tasks/functions performed by DSP evolve in the next years?
9. When you look for a Data Science Professional, do you demand any specific profile? What skills do you require of DSP to work in your organisation? Hard skills vs. Soft skills. Who is in charge of designing this profile? What qualifications/ certification do you require? What study level (Bachelor, Master, PhD) do you hire to work as DSP? Is certification a must? How many years' experience do your DSP have? Is DSP experience important for your organisation? How long does it take to integrate a Data Science Professional in your organisation?
10. Are starting salaries and salary growth expectations different from that of other employees at the same level? How much does a Data Science Professional earn? Are the specific benefits or career development you offer DSP different from those you offer other employees?
11. How are DSP recruited? What is the recruitment process like? Is it different from that of other employees? Who (post/profile) is in charge of defining needed profiles? Who interviews the DSP candidates?
12. Do you resort to external agencies/ consultants to recruit DSP? Where do you look for DSP? Universities? Other training centres? Websites? Do you consider hiring DSP from outside the EU?
13. How long does it take to hire a Data Science Professional? When you hire a Data Science Professional, do you get what you need? If you have to do extra training, what do DSP lack?
14. How do education institutions get to know your needs regarding hiring DSP? Are you currently in contact with universities or other HEI or training centres to inform them about you organization's needs in terms of DSP skills? Does your organization deploy special actions in order to recruit DSP (visit universities, job fairs, etc)?

15. Do you get from Universities the DSP you need (quantity) and with the required skills (Quality)? Do you train existing employees to become DSP? How do you train them? Externally? Have you installed any in-house training programme for DSP? Is there a specific employee profile in the organization that you train as a Data Science Professional? Who is in charge of DSP training within your organization?

b) Supply Side: Universities Interview Guideline

Main issues arising when designing a DSP oriented educational programme/course

- Content? The project develops the Data Science Competence Framework (CF-DSP), Data Science Body of Knowledge (DSP-BoK). Would these help?
- Student recruitment: Number of students? How to attract them? Process?
- Getting resources to create the course?
- Competition from / collaboration with other universities?
- Which is the business case / cost-benefit of the programme?
- Compliance with Requisites to have the new course approved?
- Identifying teaching staff?
- Any other issues not mentioned?

About the process of creating <u>and getting the approval</u> to implement innovative programmes/courses:

- How to respond to the needs of :
    - The Public administration?
    - Industry?
    - Not-for-profit organisations?
    - Large European Research infrastructures or other research organisations?
- In what way do you take into consideration what other universities /institutions are doing in this domain?
- What factors e.g. number of graduates that will find a job quickly, new technologies being used in organisations, legislation, HEI rankings, ultimately influence the design of your courses?
- If your Univ. has standard procedures to design courses, how do you adapt specific requirements to them?

Students

- What makes students want to study in your university, as compared to others? How to attract them?
- What rules the way you fix the number of vacancies (a maximum, a minimum?)? By your agreements with companies? By the number of potential students? By your available resources?
- What are the mechanisms by which you will recruit students for the course? Is there a standard procedure, or is it tailored to the potential students' interests, vocations and expectations?
- Do you consider varying your course offering according to what potential students ask for, according to the organisations you have agreements with or according to what other universities are currently doing?
- How do you take into account students expectations in the design of the courses?

Employability

- How long does it take graduates to find a job? What kind of organisations hire them? What percentage of graduates are hired? Do you have tracking mechanisms?
- Do you consider different positions /tasks/roles to cover, and therefore, design different courses of study?
- Certification/ recognition/ standards: how are these important? For graduates placement?

Wrap-up

- Bearing the issues previously discussed, where do you feel pain points are centred? How can we help YOU, as one of the administrative role responsible in the university, to create a Data Science course that produces the best DSP that the market needs?

List of interviewees

Univ. Complutense de Madrid	Baltasar Fernández Manjon
Univ. of Trento	Yannis Velegrakis
Univ. of Southampton	Jeremy Frey
Univ. of Southampton	Jon Forester
Univ. of Southampton	Ruben Sanchez Garcia -
Univ. of Southampton	Bill Brocklesby
Univ. of BedforDSPhire	Ingo Frommholz
Univ. of Stavanger	Tomasz Wiktorski
ESIC	Pablo Lopez Tenorio
Univ. of Utrecht	Alexandre Bonvin
Univ. of Amsterdam	Karen Maex
Univ. of Amsterdam	Andrea Haker
Univ. of Amsterdam	Maarten de Rijke
Vrij Universiteit of Amsterdam	Henri Bas
Univ. of Amsterdam	Marc Salomon
AWI	Angela Schaefer
CERN	Jamie Shiers
SIB Swiss Institute of Bioinformatics	Martial Sankar
EPOS	Daniele Bailo
LIFEWATCH Unicantrabria	Jesus Marco de Lucas
Univ. of Utrecht	Bas Van Breukelen
Netherlands eScience Centre	Oscar Martínez Rubí
Universidad de San Andrés	Ernesto Gore
IBM	Cortnie Abercrombie
Talenom	Sakari Jorma
Smart Data SEO	Eric Axel Franzon
Verve	Brian Crook
Dataversity	Shannon Kempe
SD Supercomputer Center	Ilkay Altintas
British Army	Lt Col Mike Servaes RA
Reeve Consulting	April Reeve
Machine intelligence	Steve Ardire
Data Kitchen	Christopher Berg
Silicon Valley Data Science	John Akred