Data Science Professional uncovered

How the EDISON Project will contribute to a widely accepted profile for Data Scientists

Andrea Manieri
R&D Lab
Engineering Ing. Inf. SpA
Rome, Italy
Andrea.manieri@eng.it

Yuri Demchenko
System and Network Engineering
University of Amsterdam,
Amsterdam; The Netherlands
y.demchenko@uva.nl

Tomasz Wiktorski
Department of Electrical
Engineering and Computer Science
University of Stavanger
Stavanger, Norway
tomasz.wiktorski@uis.no

Steve Brewer
Chemistry,
University of Southampton,
Southampton, UK
S.Brewer@soton.ac.uk

Matthias Hemmje
Faculty of Mathematics and
Computer Science
University of Hagen
Hagen, Germany
mhemmje@ftk.de

Tiziana Ferrari
European Grid Initiative
Amsterdam, Nederland
tiziana.ferrari@egi.eu

Ruben Riestra
Innovation Management Unit
INMARK sa.
Madrid, Spain
ruben.riestra@grupoinmark.com

Jeremy Frey
Chemistry
University of Southampton
Southampton, UK
j.g.frey@soton.ac.uk

Abstract— The digital revolution made available vast amounts of data both in industry and in the research landscape. The ability to manipulate and extract knowledge and value from this data represents a new profession called the Data Scientist: expected to be the most visible job in future years.

The EDISON project has been established in order to support Universities, Research Centers, Industry and Research Infrastructure organisations to cope with the potential shortfall of Data Scientists, to define the framework of competences as well as the body of knowledge for this profession.

In this paper the EDISON team describes how it intends to nurture the profession of Data Scientist to cope with the expected increase in demand. The strategy proposed is based on both the analysis of the demand side (Industries, Research Centers and Research Infrastructure organisations) and the supply side (Universities and training centers) bridging between the providers and employers by cooperating on the establishment of a Competence Framework and a Body of Knowledge for the Data Scientist Professional. The project will exploit piloting initiatives in cooperation with pioneer universities and also involve external experts as evangelists.

Keywords—Data, Big Data, Data Intensive Science, Big Data technologies, Data Science Curricula, Data Scientist Professional, Data Science Competences Framework, Data Science Body of Knowledge, EDISON Project

I. INTRODUCTION

The emergence of Data Science technologies (also referred to as Data Intensive Science or Big Data technologies) is having an impact, at a fundamental level, on nearly every aspect of how research is conducted, how scientists think, and how research data are used and shared. Similarly, the availability of an extensive amount digital data about business processes has created several opportunities for Data Scientists to be involved in new roles in industries and business organisations.

The effective use of Data Science technologies requires new skills and demands for new professions, usually referred to as the Data Scientist: an expert who is capable both to extract meaningful value from the data collected and also manage the whole lifecycle of data, including supporting Scientific Data e-Infrastructures. The future Data Scientists must possess knowledge (and obtain competencies and skills) in data mining and analytics, information visualisation and communication, as well as in statistics, engineering and computer science, and acquire experiences in the specific research or industry domain of their future work and specialization. We call this the Data Science Professional (DSP).

II. THE DATA SCIENCE LANDSCAPE

A. The Rise of the Data Science Profession

As described in the visionary book by Tony Hey and others "The Fourth Paradigm" [1] and confirmed in the HLSG report "Riding the Wave" [2] computational (and statistical) methods and data mining on large sets of scientific and experimental data play a key role in discovering hidden and obscure relationships between processes and events that are necessary in order to make new scientific discoveries and support innovation in industry and the modern digital economy. Industry also recognises the benefits of Big Data technologies and the use of scientific methods in business/operational data analysis and problem solving for managing enterprise

operations, staying innovative and competitive, and being able to provide advanced customer-centric service delivery. These changes have increased the demand for new types of specialists with strong technical background and deep knowledge of the Data Intensive Technologies that have been identified as the new profession of the Data Scientist.

The U.S. National Institute for Science and Technology (NIST) defined the following groups of skills required/expected from Data Scientists: domain experience, statistics and data mining, and engineering skills. The qualified Data Scientist should be capable of working in different roles in different projects and organisations such as Data Engineer, Data Analyst or Architect, Data Steward, etc., and possess the necessary skills to effectively operate components of the complex data infrastructure and processing applications through all stages of the Data lifecycle till the delivery of expected scientific and business values to science and/or industry.

Currently there is no widely accepted Data Science professional education or generic Big Data technologies training programs and there is no common approach as to how to effectively build professional level Big Data curricula. Universities both in Europe and in the USA (and beyond) do not offer sufficient possibilities for educating the number of this new type of specialist needed. There are some Data Science programs offered in Europe and the US, mainly as Certificates. Some of them are based on slight modification or simple re-branding of existing curricula on Business Analytics, Data Analytics or Machine Learning. Although traditional scientific disciplines can create a strong basis for a new Data Science curriculum, there should be a number of new courses to reflect the whole scientific data lifecycle management from data production, processing, analysis, visualization, management, archival, preservation, access, adaptation, dissemination and re-use.

The European Union (EU) and Member States, anticipating the e-Science paradigm, invested significant resources in building advanced Research Infrastructures (RIs) in Europe. The continuously growing complexity of scientific experiments, and consequently the supporting e-Infrastructure[1], requires a significant number of specialists to support such digital infrastructure as well as the whole scientific discovery processes. Currently this need is partially satisfied with "self-made" Data Scientists who have grown from the scientific community or Engineers that have approached a scientific discipline in a consultant role: both without any explicit recognition of their roles, competences and profession.

The new multiannual funding program from the European Commission, HORIZON 2020, has urged us to recognize "self-made" Data Scientists as professionals and to provide them with additional and more systematic training on general Data Science knowledge and competencies. Similarly, the education and training of new Data Scientists for the future research careers and projects as well as for the future labour market is recognized as a key factor needed to ensure that new generations are fully prepared to have a leading role in tomorrow's digital economy.

[1] The notion of e-Infrastructure is similar to the one of CyberInfrastructure in US and includes any electronic device and instruments plus algorithms and services needed to perform scientific processes and experiment.

B. Recognised Demand for Data Science

To lead and benefit from emerging Data Science technologies we need Data Scientists who are familiar with new data-driven technologies and data-centric infrastructures and can work both in research and in industry. Assuming the current and future researchers have their base in university/academic education in their native/original professional/scientific domain, they will require additional education and training to obtain the necessary knowledge and skills to work with existing and future data intensive infrastructures and tools.

The current number of university-trained Data Scientists or other formally trained practitioners across Europe each year is probably a few hundred graduates. This must grow by an order of magnitude in the short term.

The HLSG report "Riding the Wave" and the follow on report, "The Data Harvest" [3], set the scene by urging the European Commission to promote, and the member-states to adopt, new policies to foster the development of advanced-degree university programs for the emerging field of Data Science.

A first estimation of the need and shortage of trained Data Scientists in the short-term has been delivered by McKinsey [4]. The report predicts that in the US alone between 140-190 thousands workers are needed with 'deep data analytical' experience, and also that 1.5 million managers will need to become data literate.

The Report on the Consultation Workshop "Skills and Human Resources for e-Infrastructures within Horizon 2020" (May 2012) [5] identified nine areas as having potentially the most significant skills gaps. The areas range from data mining, data management and stewardship, as well as data preservation. For these nine areas 25-50% more experts will be required in the next 2-5 years. A whitepaper "Unleashing the potential of Big Data" published after the World Summit on Big Data and Organization Design (2013) [6] has stated that "*the process of incorporating Big Data into the operation of business, governance and education will require hundreds of thousands of new, specially trained knowledge workers*".

Another report, Big Data Analytics (assessment of demand for Labour and Skills 2013-2020), published in October 2014 [7], indicates that the demand for big data refiners is already outstripping supply, and the people needed to mine big data are becoming as precious and scarce as oil itself. This report highlights the urgency of this need. The results of their survey in the UK reveals a tenfold increase in demand for big data staff in the past five years, with vacancies rising from 1,800 in 2008 to 21,400 in 2013 – an average annual increase of 212 per cent. Over the past year, there has been a 41% increase in the number of big data jobs posted. This would imply for the UK an increase of demand for big data specialists by 160 per cent between 2013 and 2020, adding 346,000 big data jobs. Extrapolating this number to the EU scale (relative to GDP figures) would imply adding almost 2.5 million jobs. On 13 October 2014, The European Commission and data industry launched a €2.5 billion partnership to master big data [8]. A

public-private partnership (PPP) named Big Data Value Association is aiming to strengthen the data sector and put Europe at the forefront of the global data race [9].

The predictions for the growth of "data jobs" in these reports are different, but they demonstrate that a targeted effort is required to train many more specialists. The numbers above are not comparable as it is not clear to which level of trained persons they are referring. This varies from data literate managers, to data-related jobs, up to academic trained data scientists. Even when assuming from the job numbers above that only 20% require an academic training, this implies tens of thousands of graduate data scientists must graduate from universities in the next ten years. Many of these will be needed to work in RIs or comparable challenging work environments in both the public and private sectors.

III. Re-Engineering of Data Science Education

Educating and training Data Science specialists requires a new model, which reflects by design the whole lifecycle of data, and is aligned by construction with the target research and industry domains context and technologies. Such a model must be built based on a thorough analysis of the requirements posed by modern Data Science, and will define the Body of Knowledge (BoK) and Skills Profile (SP) of a Data Scientist. We envision this model will be adopted by academic programs in agreement with public and private employers, and must be supported by both teaching materials and experimental infrastructure, where the students and trainees can obtain the necessary theoretical and practical knowledge. Furthermore, for the "self-made" Data Scientists, the model should provide alternatives for obtaining structured and consistent professional knowledge on general Data Science technologies, and guidelines for formal certification from universities or other professional training organizations.

Universities, Industry and Research Infrastructures (RIs) must cooperate to establish a common Data Science competencies profile and a common component-based curriculum for education and training to achieve this profile. For achieving these goals, we foresee a need for re-engineering current education programs.

Apparently the set of mandatory skills desirable for a Data Scientist are already quite frequently targeted by modern high-level education. However they are usually addressed in a scattered way, i.e. in different programs and directions of study, often perceived as orthogonal in traditional, process-centric education. As Data Science requires a complete paradigm shift towards a data-centric view of the world, we must make two re-engineering steps towards a sustainable model for Data Science education:

(1) identify and "mix and match" all the required building blocks of existing education programs into a coherent Data Science program, and

(2) revise and/or redesign every such building block to reflect the data-centric paradigm. Finally, we must fill in the gaps that current programs rarely address, i.e. specific big data concerns directly related to the size and complexity of data that are not homogeneously reflected in traditional education programs.

The main challenges we foresee in this context are:

(1) the analysis of requirements, which must be addressed together with the Data Science industry and research partners,

(2) the paradigm-shift for an alternative, data-centric view of traditional education programs, which must be addressed together with the educators,

(3) the integration of modern, data-specific challenges in this education program, which must be tackled by both educators and practitioners,

(4) the domain specific Data Science education and availability of datasets and virtual learning environment.

IV. Supply vs. Demand side

The McKinsey report [4] identifies three key types of talent and expects that the supply of these talents will be a significant constraint on the ability of organizations around the world to capture value from Big Data. These key talent categories are: deep analytical talent, data-savvy managers and analysts, and supporting technology personnel. The report associates these talents with technical skills in statistics, machine learning, and development, implementation, and maintenance of hardware and software tools such as databases and analytic programs. Traditional university curricula do not cover such a wide range of competences within one single educational program. As a consequence, people working in the area of big data analytics in both industry and academia even now do not have all competencies needed for the work because they have been following traditional education curricula covering only part of the required skills.

In the O'Reilly book Analysing the Analysers [10], provides one of the first analysis of the Data Scientist skills and roles based on a survey over more than 200 Data Scientists done in 2009. The book identifies four organisational roles for Data Scientists (Data Business people, Data Creatives, Data Developers and Data Researchers) that describe the work done by data scientists which has in turn been mapped to five groups of the specific skills required for those roles: Business, Machine Learning and Data Structures, Mathematics and Operations research, Programming, Statistics.

A number of initiatives throughout Europe have started to address the need for a new profession. The EC has lead a number of multi-stakeholder partnership initiatives to raise awareness on e-skills and the demand for digital jobs by organizing several events in recent years such as the e-Skills Week organized in March 2012 and the Grand Coalition for Digital Jobs in March 2013. The e-Skills week has demonstrated a strong mobilisation of stakeholders (2.235 events organized in 37 European countries and 1.8 million participants). A second campaign, e-Skills for Jobs, was launched in 2014, and third one is planned in 2015-2016 [11].

Other initiatives lead by stakeholders from the European ICT sector took the challenge to define a list of competences, skills, and capability levels. One relevant to EDISON is the European e-Competences Framework version 3.0 (e-CF) [12] is the result of 8 years continuous effort and commitment by multi-stakeholders from the European ICT sector. It provides a reference of 40 competences as required and applied at the Information and Communication Technology workplace, using a common language for competences, skills and capability levels that can be understood across Europe. The e-CF framework is developed under the umbrella of the CEN

Workshop on ICT Skills [13] operational since 2003 and has since then contributed e-Skills agenda of the European Commission (EC). The CEN Workshop on ICT Skills constitutes an exchange platform for experts representing the ICT industry, academic, training organizations, ICT professional associations, and research institutions.

Within the Research Data Alliance (RDA) which aims to build the social and technical bridges needed to enable the open sharing of data, an Interest Group on Education and Training on handling of Research Data (IG-ETRD) [14] has been established and is already actively working on elaborating reference models for the setting of quality standards for appropriate education and training programmes aimed at researchers and other professionals. In this way it is already preparing the ground for practical applications of these standards in educational environments. The IG-ETRD is aiming to create taxonomies of the skills required by different groups of data management specialists/professionals and is coordinating development of framework curriculum for different domains.

A few non-aligned educational programs and individual courses have been created recently to fill the gap. One of them is the EIT ICT Labs Data Science Master program [15] which offers an academic programme whereby students can study data science, innovation, and entrepreneurship at leading European universities. In this programme, students learn about scalable data collection techniques, data analysis methods, and a suite of tools and technologies that address data capture, processing, storage, transfer, analysis, and visualization, and related concepts (e.g., data access, data pricing, and data privacy).

V. Competence Framework

The first pillar of the EDISON project is to define the Data Science Professions in a formal way in terms of defining required Competences Framework and Skills Profile (CF-DS) and mapping them to the available knowledge domains and academic disciplines currently taught at universities whilst also defining what new disciplines and training subjects/topics still need to be created. The CF-DS will include the common hard and soft skills (i.e., technical and collaborative skills) required for successful Data Scientists to work as a part of a team and be an actor in the modern agile data-driven enterprise, as well as subject-specific knowledge and skills related to different scientific and technical domains across the labour market.

The EDISON CF-DS development will follow the European e-Competences Framework (e-CF) guiding principles:

- CF-DS will adopt a holistic e-CF definition: “Competence is a demonstrated ability to apply knowledge, skills and attributes for achieving desirable results” in organisational or role context.
- CF-DS should work as an enabler for multiple applications that can be used by different types of users from individual to organisational; it should support common understanding and not mandate specific implementation.
- A competence can be a part of a job definition but cannot be used to substitute similar named job definition; one single competence can be assigned to multiple job definitions.

CF-DS will include not only specific Data Science competences but also other more generic competences, like Problem Management, Business or Risk Management, etc. It is intended that CF-DS will be integrated with the e-CFv3.0 and similarly to e-CF will require maintenance and regular updates to continue working as an organisational tool and a basis for Data Science curriculum update.

VI. Body of Knowledge

The CF-DS is complemented by the creation of the Data Science Body of Knowledge (DS-BoK), the knowledge needed by the practitioner to perform all the data related processes of his/her profession.

We will perform an analysis of existing academic and professional courses and curricula, including available books and other training resources. We will identify common conceptual elements and gaps amongst the present offering and the needed one (by discipline and market sectors).

This work will be strongly grounded in education theory, including application of Bloom’s Taxonomy, Constructive Alignment and Problem-based Learning [16]. Altogether, this will lead to a creation of modular taxonomy for Data Science, based on the CF-DS and reflecting differences of the related professions like Data Engineer, Data Steward, Scientific Data or e-Infrastructure manager, etc. This taxonomy will serve as a basis for the formalization of a unified BoK and later implementation of an adaptive curriculum of worldwide relevance for European application.

It is intended that CF-DS will be integrated with the e-CFv3.0 and comply with the European Qualifications Framework (EQF) [17] which defines eight reference levels describing what a learner knows, understands and is able to do – 'learning outcomes' [18]. EQF can act as a translation tool to make national qualifications more readable across Europe, providing a basis for European level certification, and promoting workers' and learners' mobility between countries and facilitating their lifelong learning.

Using pilot programs and champion institutions, both the model and the curriculum can be fine-tuned and further exposed for adoption to the large community of universities in search of a well-constructed Data Science education program.

VII. Engaging with Experts

There are many important stakeholders with vested interests in the successful establishment of the Data Scientist profession, including Research Infrastructure (RI) managers, industry managers and commercial laboratory chief executives, academic leaders, policy makers and of course the data scientists themselves. As we have described, the success of this professional role is predicated on the broad agreement of a process to harmonise and correlate curricula, terminology and key practices addressing this multi-stakeholders audience. This can best be achieved through collaboration and efficient interactions between these sectors in ways that would either not occur otherwise or would not happen soon enough.

EDISON will foster such interactions through a number of mechanisms, chief among them being the Expert Liaison Groups (ELG). Three such groups bring together prominent and influential individuals in order to share knowledge and

ideas, review and steer emergent material from EDISON and also to champion the ideas and outputs from the project. The three groups are: employer sectors, data experts and academic scholars and officers. These three groups will enable the project to facilitate a structured dialogue between the EDISON team and these key stakeholder communities. Other channels will also be established whether directly at events and workshops, through dissemination material, as well as through social media channels but the ELGs will provide a permanent, robust and effective route for input and impact.

VIII. The EDISON Project

In order to cooperate in the implementation of previously initiated actions, authors and related institutions joined at the 3rd Research Data Alliance (RDA) plenary of the RDA IG-ETRD activities and promoted EDISON with the aim to coordinate the establishment of the Data Science profession and development of the homogenous curricula in Europe in coordination with other similar initiatives. The EDISON project – recently funded by the European Commission - is pursuing the establishment of the Data Scientist as a new profession in support of the e-Infrastructure needs and beyond.

A. Aims and Objectives

Establishing Data Scientist as Profession is a long reaching perspective; EDISON is conceived to pave the way for such long-term challenge. In particular, EDISON will work to achieve the following interim objectives:

- To facilitate the offering of the Data Science curricula by increasing the number of universities and professional training organizations, including producing a definition of the Data Science Professional competencies and skills profiles and a corresponding Model Curricula (MC).
- To propose a framework and an ICT environment for re-skilling and (eventually) certifying Data Scientist's expertise for graduates, practitioners and researchers throughout their careers that would create alternative paths, including the "self-made" Data Scientists, to satisfy the rapidly increasing demand for skilled Data Scientists (e.g. Data Analysts, Data Engineers, Data Archivists, Data Stewards, etc.) in addition to academic education and training.
- To develop a sustainable education model for Data Science and Data Intensive technologies that is supported by a multi-stakeholder business model that incorporates a sustainability roadmap for European academia, research and industry that would be mutually beneficial for multiple stakeholders from the e-Infrastructure, academic and industrial players, the professional organizations as well as citizens at large.

B. The Project Background

The EDISON project is inspired by the report "Riding the Wave" (2010) and its follow-up "The Data Harvest" (2014) that predicted the growth of data driven technologies that require new data-centric infrastructure and indicated the importance of education and training on Data Science to empower European research and industry to benefit from the new technologies. In particular, the project will directly address four of the seven recommendations in the "The Data Harvest" report including: 2) promote data literacy across society, from researcher to citizen, 4) developing tools to build trust between data scientists, 5) supporting international collaboration and 7) don't stop what has begun well – EDISON is predicated on the idea of capturing the best of achievements to date.

The current proposal is the result of activities started by consortium members as far back as 2013 when (the initiative group from) UvA initiated a first "Birds of a Feather" (BoF) meeting at The first RDA Plenary Meeting in Gothenburg on March 18-20, 2013. The first BoF confirmed a common interest and the importance of addressing education and training needs to the European research community. The project partners has already proven experience in developing pioneering educational courses on Big Data [19], Data Intensive technologies [20] and Cloud Computing [21, 22] for Computer Science programs. The project will extend this experience to other science and industry domains by cooperating with other stakeholders.

C. Open Issues

Realization of the expected results and impacts will require an intensive combination of effort and creativity as there are multiple factors, some of them of a structural nature, that hamper the quick solution to the qualitative and quantitative gaps in the Data Science education, training and labour market. Among these factors, the following are to be addressed by EDISON partners:

- Insufficient understanding of the urgency and/or capacity among even leading European Higher Education Institutions (HEI) to establish Data Science oriented curricula and professional training programmes, and the adoption of innovation driven approaches in designing such programmes to enable the graduates to follow the technology development and constantly improve their knowledge and skills,
- Mathematical and technological complexity within the Big Data and Data Intensive Science domain that requires a wider range of knowledge to be taught/provided in the Data Scientists educational programmes,
- Need for access to real life scientific, experimental or production data and well as to modern Scientific or Big Data infrastructure which may not be available to the majority of universities and those seeking training;
- Heterogeneity among HEI due to national differences and/or curriculum capabilities,
- Heterogeneity among RI, among industries, among regions and a lack of common requirements for a new type of Data Scientist specialist making it difficult to quickly develop a common curriculum and a standard educational programme,
- Low levels of awareness of – or worse, misconceptions about – what a DS is/should be and the value they provide, in the perception of the various stakeholders groups,
- Unwillingness of both the prospective students and employers to wait for the 2 years Master education cycle in conditions of urgent Data Scientists needs and trends among companies to provide internal training for their own staff,
- Lower attractiveness of the scientific in general and e-Infrastructures in particular, or public sector, as professional career paths among the incoming Generation

Y/Millennial students compared to the commercial and business sector,

- A widening gap and vital need for cooperation between the science/research sector and industry in using emerging Big Data and Data Intensive Technologies; at the same time overhyped expectation from companies in fast return of investment in new technologies.

IX. Summary and Future Actions

The EDISON project is supported by the EC under Grant Agreement n. 675419 and will officially start its activities on the 1st Sept 2015. The following are the first steps that will be addressed by the project team:

- Defining Education and Training needs and creating a model of the Data Scientist Professional: In this phase EDISON will gather information about the educational needs for Data Scientist profession from different research and industry domains, collect information about existing education and training resources focused on and related to Data Science, create a taxonomy of existing academic and technical disciplines, and based on this create competences and skills framework for Data Science (CF-DS) together with a definition of the Data Science Body of Knowledge (DS-BoK)
- Developing the EDISON Education and Training Model that includes the Model Curriculum in Data Science (MC-DS) and its profiles for a selected number of scientific and industry domains. EDISON MC-DS will be tested with pilot implementations for the three basic use cases defined for the project. Supporting the development of Data Science education and training courses, and developing accreditation scheme for education and training courses implementing MC-DS.
- Creating components of the EDISON Online Education and Training Infrastructure that will provide one-stop shop for educational and training resources to support both residential (university) and online education (including curricula, courses, textbooks, online resources, educational and computational platforms, educational datasets).
- Developing a sustainable business model for Data Science related education and training that should be supported by the Certification scheme for Data Scientist Professionals that should open possibilities both for future university graduates in Data Science, for practitioners who want to obtain a new profession and built a career in Data Science, including self-made Data Scientists. This activity will also produce the Roadmap for sustainability of Data Science education and training in Europe.

Acknowledgment

The EDISON project is supported under Grant Agreement n. 675419 by the European Commission. The positions expressed by the EDISON consortium during the execution of the Grant Agreement are not necessarily those of the European Commission nor other European Institutions.

References

[1] The Fourth Paradigm: Data-Intensive Scientific Discovery. Edited by Tony Hey, Stewart Tansley, and Kristin Tolle. Microsoft Corporation, October 2009. ISBN 978-0-9825442-0-4 [Online]. Available: http://research.microsoft.com/en-us/collaboration/fourthparadigm/

[2] Riding the wave: How Europe can gain from the rising tide of scientific data. *Final report of the High Level Expert Group on Scientific Data. October 2010*. [Online]. Available at http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

[3] The Data Harvest: How sharing research data can yield knowledge, jobs and growth. An RDA Europe Report. December 2014

[4] Big data: The next frontier for innovation, competition and productivity, Report, McKinsey Global Institute, 2011 [online]

[5] Skills and Human Resources for e-Infrastructures within Horizon 2020, The Report on the Consultation Workshop, May 2012. [online] http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/report_human_skills.pdf

[6] Unleashing the potential of Big Data. A Whitepaper "World Summit on Big Data and Organization Design, 2013

[7] Big Data Analytics: Assessment of demand for Labour and Skills 2013-2020. Tech Partnership publication, SAS UK & Ireland, November 2014 [online] https://www.e-skills.com/Documents/Research/General/BigData_report_Nov14.pdf

[8] European Commission and data industry launch €2.5 billion partnership to master big data, Press release, Brussels, 13 October 2014 [online] http://europa.eu/rapid/press-release_IP-14-1129_en.htm

[9] Big Data Value Association [online] http://www.bdva.eu/

[10] H.Harris, S. Murphy, M. Vaisman, "Analysing the Analysers: An Introspettive Survey of Data Scientists and Their work", O'Reilly Media, Inc, 2013

[11] e-Skills for the 21st Century Fostering Competitiveness, Growth and Jobs http://ec.europa.eu/enterprise/sectors/ict/e-skills/index_en.htm

[12] European eCompetences Framework http://www.ecompetences.eu/

[13] CEN ICT skills Workshop http://www.ecompetences.eu/cen-ict-skills-workshop/

[14] Research Data Alliance (RDA) IG Education and Training on handling of research data (IG-ETRD) [online] https://www.rd-alliance.org/ig-education-and-training-handling-research-data.html

[15] EIT ICT Labs Master Program on Data Science (DSC) [online] http://www.masterschool.eitictlabs.eu/programmes/data-science/Bloom's taxonomy: the 21st century version. [online] http://www.educatorstechnology.com/2011/09/blooms-taxonomy-21st-century-version.html

[16] John Biggs, Constructive alignment in university teaching, HERDSA Review of Higher Education, Vol. 1 [online] http://www.herdsa.org.au/wp-content/uploads/HERDSARHE2014v01p05.pdf

[17] Descriptors defining levels in the European Qualifications Framework (EQF) [online] http://ec.europa.eu/ploteus/en/content/descriptors-page

[18] Demchenko, Y., E.Gruengard, S.Klous, Instructional Model for Building effective Big Data Curricula for Online and Campus Education. In Proc. 6th IEEE International Conference and Workshops on Cloud Computing Technology and Science (CloudCom2014), 15-18 December 2014, Singapore

[19] Wiktorski-Wlodarczyk, T. Hacker, T.J. "Problem-based Learning Approach to a Course in Data Intensive Systems", In Cloud Computing Technology and Science (CloudCom), 2014 IEEE 6th International Conference on (pp. 942-948). IEEE.

[20] Demchenko, Yuri, David Bernstein, A.S.Z. Belloum, Ana Oprescu, Tomasz W. Wlodarczyk, and Cees de Laat, *New Instructional Models for Building Effective Curricula on Cloud, Compting Technologies and Engineering,* IEEE International Conference on Cloud Computing Technology and Science, Bristol, Dec 2013.

[21] Demchenko, Y., Bernstein, D., Belloum, A., Oprescu, A., Wiktorski-Wlodarczyk, T. & Laat, C. D.. "New Instructional Models for Building Effective Curricula on Cloud Computing Technologies and Engineering". In Cloud Computing Technology and Science (CloudCom), 2013 IEEE 5th International Conference on (Vol. 2, pp. 112-119). IEEE.

[22] DRAFT NIST Special Publication 1500-1: NIST Big Data Interoperability Framework, Volume 1-7. [Online]. Available: http://bigdatawg.nist.gov/V1_output_docs.php