EDISON Data Science Framework: A Foundation for Building Data Science Profession For Research and Industry

Yuri Demchenko, Adam Belloum, Wouter Los
University of Amsterdam, The Netherlands
{y.demchenko, A.S.Z.Belloum, W.Los}@uva.nl

Andrea Manieri
Engineering Ingegneria Informatica S.p.A., Italy
Andrea.Manieri@eng.it

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

Tomasz Wiktorski
University of Stavanger, Norway
tomasz.wiktorski@uis.no

Holger Brocks, Jana Becker, Dominic Heutelbeck, Matthias Hemmje
FTK – Research Institute for Telecommunication and Cooperation, Germany
{hbrocks, jbecker, dheutelbeck, mhemmje}@ftk.de

Abstract— Data Science is an emerging field of science, which requires a multi-disciplinary approach and should be built with a strong link to emerging Big Data and data driven technologies, and consequently needs re-thinking and re-design of both traditional educational models and existing courses. The education and training of Data Scientists currently lacks a commonly accepted, harmonized instructional model that reflects by design the whole lifecycle of data handling in modern, data driven research and the digital economy. This paper presents the EDISON Data Science Framework (EDSF) that is intended to create a foundation for the Data Science profession definition. The EDSF includes the following core components: Data Science Competence Framework (CF-DS), Data Science Body of Knowledge (DS-BoK), Data Science Model Curriculum (MC-DS), and Data Science Professional profiles (DSP profiles). The MC-DS is built based on CF-DS and DS-BoK, where Learning Outcomes are defined based on CF-DS competences and Learning Units are mapped to Knowledge Units in DS-BoK. In its own turn, Learning Units are defined based on the ACM Classification of Computer Science (CCS2012) and reflect typical courses naming used by universities in their current programmes. The paper provides example how the proposed EDSF can be used for designing effective Data Science curricula and reports the experience of implementing EDSF by the Champion Universities that cooperate with the EDISON project.

Keywords—Data Science, Data Scientist Professional, Big Data, EDISON Data Science Framework (EDSF), Data Science Competences Framework (CF-DS), Data Science Body of Knowledge (DS-BoK), Data Science Model Curriculum (MC-DS), Data Science Professional profiles.

I. INTRODUCTION

Data Science is an emerging field of science, which requires a multi-disciplinary approach and has a strong link to Big Data and data driven technologies that created transformational effect to all research and industry domains. Their sustainable development requires re-thinking and re-design of both traditional educational models and existing courses. However, at present time most of the existing university curricula and training programs are built based on available courses and cover limited set of competences and knowledge areas that are related to multiple Data Science and general data management professional profiles as defined in the project. This potentially may create gaps in knowledge and competences of the future Data Scientist graduates for their smooth integration in the real working environment (both in industry and academia).

In recent years, Europe created advanced Research e-Infrastructures (eRI) supporting numerous European research communities. The complexity of eRIs is continuously growing and their normal operation requires more and more qualified engineers, specialists and researchers. The tasks of these experts expanded from equipment maintenance and operation in the past to solving complex tasks with data management and assisting researchers with new scientific and data analytics tools. With the growth of data driven research, the IT and data specialists are getting directly involved into the research process, and their ability to provide deep insight into generated and collected data becomes an important component of modern research. Subject domain and data domain specialists need to work together to benefit from available technologies and to obtain reliable scientific results for practical products and implementation.

978-1-5090-1445-3/16/$31.00 ©2016 IEEE (DTW 2016)

This paper presents a research and coordination activity done in the framework of the EU funded EDISON project to establish the new profession of Data Scientist [1, 2]. The paper provides information about the proposed EDISON Data Science Framework (EDSF) and its components that include the Data Science Competence Framework (CF-DS), Data Science Body of Knowledge (DS-BoK), Data Science Model Curriculum (MC-DS), and Data Science Professional profiles (DSP). The EDSF is intended to provide a basis for building effective Data Science curricula and enable the whole Data Science supply-demand-community ecosystem.

The paper is organized as follows. Section II introduces into the problem area of establishing a new profession of Data Scientist for new emerging technology and science domain. Section III describes the proposed EDSF and sections IV-VI provide details about CF-DS, DS-BoK and MC-DS components correspondingly. Section VII discusses how the EDSF can be used for designing and assessing Data Science curricula, and section VII reports on the experience of working with champion universities adopting the EDSF.

II. The Emergence of the Data Scientist Profession

The revolutionary value of data in modern computer powered e-Science is recognized in earl works by technology visionaries. It is first described in the book by Tony Hey and others "The Fourth Paradigm" [3] and confirmed in the HLEG report "Riding the wave: How Europe can gain from the rising tide of scientific data" [4], that computational (and statistical) methods and data mining on large sets of scientific and experimental data will 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 in problem solving for managing enterprise operations, staying innovative and competitive, and being able to provide advanced customer-centric service delivery. Modern agile data driven companies are transforming their organizational to reflect the important role of data in optimizing business and operational processes. These changes have increased the demand for new types of specialists with strong technical background and deep knowledge of the data intensive technologies. This has been defined as a new profession of the Data Scientist.

A. Data Science Professional definition and skills

There is no well established definition of the Data Scientist due to a diverse number of competences and skills expected from these specialists. We will take as a basis the definition provided in the NIST SP1500-1 document [5]: "*A **Data Scientist** is a practitioner who has sufficient knowledge in the overlapping regimes of expertise in business needs, domain knowledge, analytical skills, and programming and systems engineering expertise to manage the end-to-end scientific method process through each stage in the **big data lifecycle**."* The document defines the following groups of skills required from the Data Scientists: domain experience, statistics and data mining, and engineering skills [5].

Other definitions [6, 7] admit such desirable features as ability to solve variety of business problems, optimize performance and suggest new services for the organisation employing Data Scientist. Many practitioners admit a need for a successful Data Scientist to develop a special mindset, to be statistically minded, understand raw data and "appreciate data as a first class product" [8].

The qualified Data Scientist should be capable of working in different roles in different projects and organisations such as Data Engineer, Data Analyst or Data 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.

B. Data Science Education: New approach required

Educating and training Data Science specialists requires a new model, reflecting in its design the whole lifecycle of data in research and industry domains and requirement to have a broad range of skills to use data to deliver insight into organisational processes and their improvements. Such a model must be built on a thorough analysis of the requirements of modern Data Science to define the competence profile, required skills and other professional intelligence characteristics. There is also conceptual challenge to connect different terminology, operational models in interconnected sectors of science and technology, research, business, and education that all together create an ecosystem for a new emerging profession.

Currently there is no widely accepted Data Science professional education curricula, neither generic Big Data technologies training programs. Further, there is no common approach to effectively build professional level Data Science curricula. Universities both in Europe and in the USA (and beyond) do not offer sufficient possibilities for educating the large number of this new type of specialist. Universities, Industry and Research Infrastructures (RIs) must cooperate to establish a common Data Science competence profile and a common component-based curriculum for education and training to realise this profile.

III. EDISON Data Science framework components

The EDISON vision for building the Data Science profession will be enabled through the creation of a comprehensive framework for Data Science education and training that includes such components as Data Science Competence Framework (CF-DS), Data Science Body of Knowledge (DS-BoK) and Data Science Model Curriculum (MC-DS).

Figure 1 below illustrates the main components of the EDISON Data Science Framework (EDSF) and their inter-

relations that provides conceptual basis for the development of the Data Science profession (references to published discussion documents are provided):

- CF-DS – Data Science Competence Framework [9]
- DS-BoK – Data Science Body of Knowledge [10]
- MC-DS – Data Science Model Curriculum [11]
- DSP - Data Science Professional profiles and occupations taxonomy [12]
- Data Science Taxonomy and Scientific Disciplines Classification (including Vocabulary)

The proposed framework provides a basis for other components of the Data Science professional ecosystem:

- EDISON Online Education Environment (EOEE)
- Education and Training Marketplace and Directory
- Data Science Community Portal (CP) that also includes tools for individual competences benchmarking and personalized educational path building
- Certification Framework for core Data Science competences and professional profiles

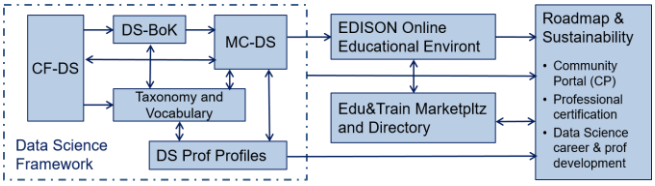

Figure 1 EDISON Data Science Framework components.

The CF-DS includes common competences required for successful work of Data Scientists in different work environments in industry and in research and through the whole career path. The future CF-DS development will include coverage of the domain specific competences and skills and will involve domain and subject matter experts.

The DS-BoK defines the Knowledge Areas (KA) for building Data Science curricula that are required to support required Data Science competences. DS-BoK is organised by Knowledge Area Groups (KAG) that correspond to the CF-DS competence groups. DS-BoK incorporates best practices in Computer Science and domain specific BoK's and includes KAs defined based on the Classification Computer Science (CCS2012) [13], components taken from other BoKs and proposed new KAs to incorporate new technologies used in Data Science and their recent developments.

The MC-DS is built based on CF-DS and DS-BoK where Learning Outcomes are defined based on CF-DS competences and Learning Units are mapped to Knowledge Units in DS-BoK. Three mastery (or proficiency) levels are defined for each Learning Outcome to allow for flexible curricula development and profiling for different Data Science professional profiles.

The DSP profiles and Data Science occupations taxonomy are defined based on and as an extension to the European Skills, Competences, Qualifications and Occupations (ESCO) [14]. DSP profiles definition will create an important instrument to define effective organisational structures and corresponding roles. DSP can also be used for building individual career path and corresponding competences and skills transferability between organisations and economy sectors.

The Data Science Taxonomy and Scientific Disciplines Classification will serve to maintain consistency between four core components of EDSF. To ensure easy navigation and mapping between the EDSF components, all attributes and properties are enumerated: competences in CF-DS, KAGs and KAs in DS-BoK, LOs and LUs in MC-DS, professional profiles in DSP.

IV. Data Science Competence Framework

The CF-DS includes the common hard and soft skills (i.e., technical and collaborative skills, also called social or professional intelligence) required to have Data Scientists engaged in a team and to act in the modern agile data-driven enterprise, as well as the subject-specific knowledge and skills allowing to work in different scientific and technical domains. The EDISON CF-DS development follows the European e-Competences Framework (e-CF3.0) guiding principles [15].

The EDISON study on Data Science competences revealed that two new groups of competences should be included that have not been explicitly identified in previous studies and frameworks. The figure presents the following competences.

3 competence groups identified in the NIST document and confirmed by analysis of collected data:

- Data Analytics including statistical methods, Machine Learning and Business Analytics
- Engineering: software and infrastructure
- Subject/Scientific Domain competences and knowledge

2 identified competence groups that are highly demanded and are specific to Data Science

- *Data Management, Curation, Preservation (new)*
- *Scientific or Research Methods (new)*

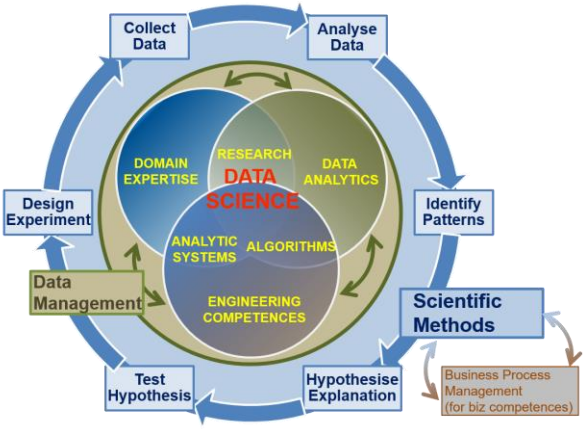

Figure 2. Data Science competence groups

Data management, curation and preservation are already included in the existing (research) data related professions such as data steward, data archivist, data manager, digital librarian, data curator, and others. Research data management is an important component of European Research Area policy. Companies also recognize needs for data management skills when they start using data driven technologies.

Knowledge of the scientific research methods and techniques makes the Data Scientist profession different from all previous professions. For business related professions, similar role belongs to business process management that need to be adopted to new data driven agile business model, in particular, to adopt continuous data driven business processes improvement [16].

The identified demand for general competences and knowledge on Data Management and Research Methods needs to be implemented in the future Data Science education and training programs, as well as to be included into re-skilling training programmes. It is important to mention that knowledge of Research Methods does not mean that all Data Scientists must be talented scientists; however, they need to know general research methods such as formulating hypothesis, applying research methods, producing artefacts, and evaluating hypothesis (so called 4 steps model). Research Methods training are already included into master programs and graduate students.

The identified competence areas provide a better basis for defining education and training programmes for Data Science related jobs, re-skilling and professional certification.

V. Data Science Body of Knowledge

The CF-DS provides a basis for the definition of the Data Science Body of Knowledge (DS-BoK), the knowledge needed by the Data Science practitioners to perform all the data related processes of his/her profession. The BoK defines the content of a curriculum and needs to be mapped to desired Learning Outcomes, which in its own turn are defined by required competences for target professions.

The DS-BoK should contain the following Knowledge Area groups (KAG) that are defined after CF-DS competence groups:

- KAG1-DSDA: Data Analytics group including Machine Learning, statistical methods, and Business Analytics
- KAG2-DSENG: Data Science Engineering group including Software and infrastructure engineering
- KAG3-DSDM: *Data Management group including data curation, preservation and data infrastructure*
- KAG4-DSRM: *Scientific or Research Methods group*
- KAG5-DSBP: Business process management group
- KAG6-DSDK: Data Science Domain Knowledge group includes domain specific knowledge

DS-BoK includes both KAGs from existing BoKs (currently used BABOK, ACM-BOK, DAMA-BOK, PM-BOK) and those defined based on the ACM CCS2012. The subject domain related knowledge group (scientific or business) KAG6-DSDK is recognized as essential for practical work of Data Scientist what in fact means not professional work in a specific subject domain but understanding the domain related concepts, models and organisation and corresponding data analysis methods and models. These knowledge areas will be a subject for future development in tight cooperation with subject domain specialists.

VI. Data Science Model Curriculum

Model Curriculum can be regarded as a blueprint that can be used by educators and trainers to develop curricula at various educational institutions and for different target groups. Definition of MC-DS should incorporate best practices and be grounded in education theory to achieve required learning outcome. The following learning and instructional models are considered: Bloom's Taxonomy, Constructive Alignment and Problem Based Learning, Competence Based Learning, that have being partly evaluated in early authors' works [17, 18, 19].

From the practical perspective, the Model Curriculum represents a tool for

- supporting the development of new Data Science programmes (including selection of appropriate learning units) tailored according to proficiency levels required to address competences required for identified DSP profiles
- assessing the compliance of existing Data Science programmes, facilitating the elicitation of potential gaps related to specific competence groups and knowledge areas implied by target professional profiles.

Hence, the Model Curriculum helps matching the supply-side and demand-side requirements for Data Science education. The formal MC-DS definition will create a basis for Data Science educational and training programmes compatibility and consequently Data Science related competences and skills transferability.

A. Best practices in effective curricula design

The proposed MC-DS will reuse the best practices in curriculum design and new educational model to facilitate the students learning as well as existing staff professional training and re-skilling for data related technologies.

There are several concepts that can guide the development of an effective curriculum: Alignment and Coherence, Scope, Sequence, Continuity, and Integration [20]. These 5 basic concepts help develop a logically consistent curriculum which components (courses, and learning units) complement each other and are ordered in such a way that it form a continuous, logical, and progressive learning path. There are several common frameworks used to develop model curricula, some are subject or discipline centric while others are organized around concept and skills that are revised as we progress across the curriculum. In practice, model curricula should

define either the time students have to spend learning given topics (usually using credit units) or the outcome assessing whether students have mastered the given competences (knowledge, abilities and skills). The latter approach is known as Competence Based Education (CBE) or Outcomes Based Learning.

B. *Competence Based Education*

CBE uses a different from the traditional education approach. Instead of focusing on how much time students spend learning a particular topic or concept (Carnegie unit credit hour, so called "sit time"), the CBE assesses whether students have mastered the given competencies, namely the knowledge, skills, and abilities [21, 22]. For this, the well-defined learning outcomes are specified for all academic activities or classes, and students' progress is assessed against those learning outcomes. The learner (student or trainee) is evaluated on the specified (group of) competences, and only after mastering them they can move on to others.

The CBE is also associated with more flexible study model for already working learners or those who undergo professional re-skilling or want to train for a new profession based on their existing experience, competences and skills. In this case, they can skip learning modules entirely if they can demonstrate require competences through the assessment system or formal testing.

The CBL can also allow the students to learn in their own pace, practicing necessary skills as much as they need to achieve necessary mastery level. It works naturally with both individual self-study and with teacher or instructor supervised/facilitated study, so well suited for online and remote education, and in particular for post-graduate education. CBL is also associated with such educational technologies and models as MOOCs, flipped classrooms, learning analytics, and others targeting growing needs of life-long learning and self-re-skilling dictated by current fast technologies development.

C. *MC-DS Design approach*

Model Curriculum defines a set of 6 Learning Outcomes for each competence group listed in Section IV. Each LO is defined of three mastery levels following levels of Bloom's Taxonomy [23] and also aligned with European Qualification Framework [24], European e-Competence Framework, and ACM/IEEE Computer Science Curriculum Definition [25]. For each Data Science professional profile MC-DS assigns ECTS points to particular LOs on relevant mastery levels.

The Model Curriculum is organized as core and elective topics. Core topics are required to every Data Science program while Elective topics aim to cover in depth the knowledge on a specific area of data science.

The EDISON approach to defining the Data Science Model Curriculum follows a CBE model and can be summarized in the following steps:

1. For each enumerated competence from CF-DS, define LO according to knowledge or mastery level (defined as Familiarity, Usage, Assessment)
2. Each KAG from DS-BoK is mapped to existing academic subject classification groups that is primarily based on ACM CS2012 and complemented with the domain or technology specific classifications to be defined by subject matter experts.
3. For each KA or Knowledge Unit, specify related Learning Units (LU) defined according to academic subject classification or following current practices by universities
4. For each Learning Unit, assign its category as core/mandatory (Tier1 or Tier 2), elective or prerequisite
5. For both Core or Elective LU, define a list of Learning Outcomes

VII. Examples of Practical use of EDSF

A. *New Data Science courses development*

MC-DS can be used for design a new Data Science curricula for target group of students or learners. In practice when designing a new programme it is necessary to decide on the set of courses with a specific number of credits. The standard in Europe is to use European Credit Transfer System, which defines bachelor programs to have 180 points and Master programs 120 points. This gives usually 30 points per semester. At American institutions credit hours systems are used and they are not fully uniform between institutions.

Required proficiency in each competence group for each professional profile is illustrated in Figure 3. It creates a basis for division of points between LOs and related LUs. In addition, each Learning Outcome can be achieved on three different knowledge or mastery levels (familiarity, usage, assessment). Typically, Bachelor programs focus on two lower levels and Master programs on two higher levels.

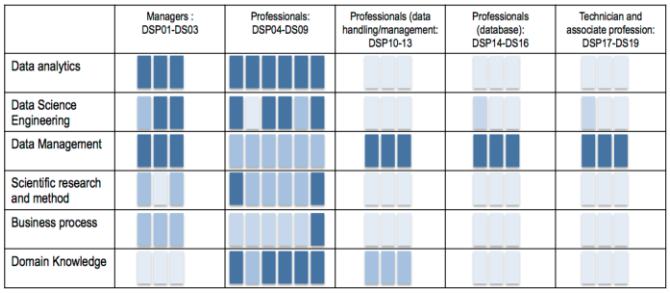

Figure 3. Proficiency/mastery level needed by different Data Science Profile for each of Data Science competence groups
Legend: (1) Bars represent individual DSP profiles [12], 2) mastery levels: familiarity –light blue; usage - blue; assessment – dark blue.

B. *Assessment, certification and accreditation*

The proposed EDSF and MC-DS provide a strong foundation for existing Data Science programmes assessment and accreditation based on MC-DS and

according to target DSP profiles. CF-DS and DSP profiles together with DS-BoK provide a basis for individual professional certification. Such use of EDSF will help to close the gap between the offered Data Science education and demand from the job market.

VIII. Validation – Piloting Use Cases

Two types of *Piloting Use Cases* are defined to support the validation and (iterative) refinement of the proposed EDISON Data Science Framework, but also to further exemplify its adoption in a number of practical settings with diverse stakeholder groups:

- **Use Case A University Champions** – Universities and Scholars will use the EDSF to align/structure their curricula as well as their corresponding educational offering.
- **Use Case B Lifelong Learning/Training** – Industries will use the EDSF to define employment policies and of course by trainers and education institutions to define and execute lifelong learning activities.

Hence, UC A targets the application of EDISON outcomes for supporting academic education to educate young students to understand and develop their unique capabilities and careers. UC B addresses Continuous Professional Education/Lifelong Learning scenarios re-skilling or planning the career path.

The implementation of both above mentioned use cases requires the active participation of "EDISON Champions", which represent different types of educational profiles as well as address a diverse range of target sectors (such as science, public institutions, industry).

A. EDISON Champions

An "EDISON Champion" is an institution that wants to align its Data Science programme (either running or planned) with EDISON outcomes. The project and its champions collaborate to develop and share relevant information and other resources needed, and to share results of such endeavor which contribute to better fulfil their respective institutional mandates. Champions are pioneers (or early-adopters) participating in piloting use cases to assess, align, and adopt EDISON outcomes before further exposure to larger communities.

B. Implementation Strategy

The Piloting Use Cases are conceived as agile, i.e. their implementation will be iterative, incremental, and evolutionary (i.e. inherent input/feedback loops). The Champions are the Primary Actors (their mentors are supporting Secondary Actors), whereas the EDISON Data Science Framework characterizes the System under discussion. The scenarios are then defined by engaging the Primary Actors in the application of the components of the EDISON Data Science Framework.

Each use case instance is initialized by a reflection (desk-based, assessing scientific and technical soundness, scope and relevance, readability, appearance and structure, usability, etc.) of the particular EDISON Framework constituent, relying on the documents provided by the project. Then it is followed by hands-on application w.r.t. the Primary Actors (Champions) own offerings, supported by the respective Secondary Actors (EDISON Mentors). Experiences, issues, and insights will be captured systematically and reported back to the respective constituent owners for consideration and management/implementation of changes, if need be.

IX. Conclusion and Further Developments

The presented EDSF includes components to be implemented by the main stakeholder of the supply and demand side: universities, professional training organisations, standardisation bodies, accreditation and certification bodies, companies and organisations and their Human Resources department to successfully manage competences and career development of the data related jobs. All working EDSF documents are published at the project web site under the Creative Common Attribution License (CC BY). The proposed EDSF has been widely discussed at numerous workshops and community forums. It is already used by few institutions associated with the EDISON project. The future development will include collecting feedback from experts and communities of practice to include extensions needed for different scientific and technology domains.

To ensure successful acceptance of the proposed EDSF and its core components, essential role belongs to standardisation in the related technology and educational domains. This work is being done in the project. Necessary contacts with the European and international standardisation bodies and professional organisations are being established.

It is anticipated that real life implementation and adoption of the EDISON Data Science framework will includes both approaches top-down and bottom-up that will allow universities and professional training institutions to benefit from EDISON recommendations and adopt them to available expertise, resources and demand of the Data Science competences and skills.

Acknowledgment

This paper presents the ongoing results of the EDISON project. The EDISON project is supported under H20202 Grant Agreement n. 675419 by the European Commission.

References

[1] EDISON Project: Building Data Science Profession [online] http://www.edison-project.eu/

[2] Andrea Manieri, et al, Data Science Professional uncovered: How the EDISON Project will contribute to a widely accepted profile for Data Scientists, Proc.The 7th IEEE International Conference and Workshops on Cloud Computing Technology and Science (CloudCom2015), 30 November - 3 December 2015, Vancouver, Canada

[3] 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/

[4] 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

[5] NIST SP 1500-1 NIST Big Data interoperability Framework (NBDIF): Volume 1: Definitions, Sept 2015 [online] http://nvlpubs.nist.gov/ nistpubs/SpecialPublications/NIST.SP.1500-1.pdf

[6] What is a data scientist? 14 definitions of a data scientist! [online] http://bigdata-madesimple.com/what-is-a-data-scientist-14-definitions-of-a-data-scientist/

[7] Cortnie Abercrombie, What CEOs want from CDOs and how to deliver on it [online] http://www.slideshare.net/IBMBDA/what-ceos-want-from-cdos-and-how-to-deliver-on-it

[8] LinkedIn's Daniel Tunkelang On "What Is a Data Scientist?" [online] http://www.forbes.com/sites/danwoods/2011/10/24/linkedins-daniel-tunkelang-on-what-is-a-data-scientist/

[9] Data Science Competence Framework [online] http://edison-project.eu/data-science-competence-framework-cf-ds

[10] Data Science Body of Knowledge [online] http://edison-project.eu/data-science-body-knowledge-ds-bok

[11] Data Science Model Curriculum [online] http://edison-project.eu/data-science-model-curriculum-mc-ds

[12] Data Science Professional Profiles [online] http://edison-project.eu/data-science-professional-profiles

[13] The 2012 ACM Computing Classification System [online] http://www.acm.org/about/class/class/2012

[14] ESCO (European Skills, Competences, Qualifications and Occupations) framework [online] https://ec.europa.eu/esco/portal/#modal-one

[15] European e-Competence Framework 3.0. A common European Framework for ICT Professionals in all industry sectors. CWA 16234:2014 Part 1 [online] http://ecompetences.eu/wp-content/uploads/2014/02/European-e-Competence-Framework-3.0_CEN_CWA_16234-1_2014.pdf

[16] Five Steps to a Data Driven Organization, Blog article by Shelley Sweetv[online] http://www.bpminstitute.org/resources/articles/five-steps-data-driven-organization

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

[18] Demchenko, Yuri, David Bernstein, Adam Belloum, Ana Oprescu, Tomasz W. Wlodarczyk, Cees de Laat, New Instructional Models for Building Effective Curricula on Cloud Computing Technologies and Engineering. Proc. Workshop "Requirements Engineering for Cloud Computing (RECC)", in conjunction with The 5th IEEE International Conference and Workshops on Cloud Computing Technology and Science (CloudCom2013), 2-5 December 2013, Bristol, UK.

[19] Wlodarczyk, Tomasz Wiktor, and Thomas J. Hacker. "Problem-Based Learning Approach to a Course in Data Intensive Systems." Cloud Computing Technology and Science (CloudCom), 2014 IEEE 6th International Conference on. IEEE, 2014.

[20] Alignung and Building Curriculum: General Concepts and Design Principles [online] http://gototheexchange.ca/index.php/curriculum-overview/curriculum-models-and-design-principles

[21] Sally M. Johnstone, Louis Soares, Principles for Developing Competency-Based Education Programs [online] http://www.changemag.org/Archives/Back%20Issues/2014/March-April%202014/Principles_full.html

[22] Information Technology Competency Model of Core Learning Outcomes and Assessment for Associate-Degree Curriculum (2014) http://www.capspace.org/uploads/ACMITCompetencyModel14October2014.pdf

[23] Bloom, B. S.; Engelhart, M. D.; Furst, E. J.; Hill, W. H.; Krathwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. New York: David McKay Company.

[24] European Qualifications Framework (EQF) [online] https://ec.europa.eu/ploteus/content/descriptors-page

[25] Computer Science 2013: Curriculum Guidelines for Undergraduate Programs in Computer Science http://www.acm.org/education/CS2013-final-report.pdf