EDISON Data Science Framework (EDSF) Extension to Address Transversal Skills required by Emerging Industry 4.0 Transformation

Yuri Demchenko
University of Amsterdam, The Netherlands
{y.demchenko, w.los}@uva.nl

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

Cuadrado Gallego Juan José
University of Alcala
jjcg@uah.es

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

Abstract—The emerging data-driven economy (also defined as Industry 4.0 or simply 4IR), encompassing industry, research and business, requires new types of specialists that are able to support all stages of the data lifecycle from data production and input, to data processing and actionable results delivery, visualisation and reporting, which can be collectively defined as the Data Science family of professions. Data Science as a research and academic discipline provides a basis for Data Analytics and ML/AI applications. The education and training of the data related professions must reflect all multi-disciplinary knowledge and competences that are required from the Data Science and handling practitioners in modern, data-driven research and the digital economy. In the modern era, with ever faster technology changes, matched by strong skills demand, the Data Science education and training programme should be customizable and deliverable in multiple forms, tailored for different categories of professional roles and profiles. Referring to other publications by the authors on building customizable and interoperable Data Science curricula for different types of learners and target application domains, this paper is focused on defining a set of transversal competences and skills that are required from modern and future Data Science professions. These include workplace and professional skills that cover critical thinking, problem solving, and creativity required to work in highly automated and dynamic environment. The proposed approach is based on the EDISON Data Science Framework (EDSF) initially developed within the EU funded Project EDISON and currently being further developed in the EU funded MATES project and also the FAIRsFAIR projects.

Keywords—Data Science, Data Scientist Professional, Big Data, Industry 4.0, EDISON Data Science Framework (EDSF), Data Science Competences Framework, Data Science professional skills, FAIR principles, Data Management and Stewardship.

I. Introduction

The emerging data-driven economy, as a part of a more general Fourth Industrial Revolution (also referred to as Industry 4.0 or simply 4IR) is powered by the convergence of previously disconnected fields such as Mathematics, Cloud Computing, Big Data, Data Science and Analytics (DSA), Artificial Intelligence (AI), robotics, mobile technologies, 3D printing, internet of things (IoT), nanotechnology and biotechnologies, that are all based on automation and the digitalisation of organisational, industrial and business processes. Industry 4.0 is characterized by fast development, high levels of technology convergence, and an increased role for knowledge, skills and human factors to enable continuous and sustainable science and technology development. Such a type of economy requires new types of data-driven approaches, and Data Science and Data Analytics enabled competences and related workplace skills.

In conditions of continuous technological development and shortened technology change cycles, Data Science education requires an effective combination of theoretical, practical and workplace skills. The importance of effective application of existing data analytics and data management platforms and tools, and corresponding hands on experience is growing, and their elements need be generically incorporated into modern curriculum design.

The EDISON Data Science Framework (EDSF) [1, 2-5], which is the products of the EDSION Project, provides a basis for building such effective education and training curricula. This paper discusses ongoing developments to extend EDSF with the set of transversal competences and skills that are required for modern and future Data Science professions. These include sought after workplace and professional skills such as critical thinking, problem solving and creativity, and the ability to thrive in a highly automated and dynamic environment.

This paper refers to previous work by the authors that proposed new approaches to building effective curricula in Cloud Computing, Big Data and Data Science [6, 7. 8, 9] and applies them to define a set of foundational courses that are broadly applicable, or the general Data Science curricula, and these can enable digital transformation of organisations.

The paper is organized as follows: section II provides references to recent studies on the skills demand for Industry 4.0 and the data-driven economy and describes the challenges in professional education and training for data scientists and other data workers. Sections III and IV briefly describe the EDISON Data Science Framework and its components. Section V describes example courses to support commonly required competences in Data Science curricula that are important to facilitate the digital transformation of organisations. Section VI describes the proposed Data Management and Data Stewardship course content. Section VII provides summary and suggestions for future work, including vision for adopting FAIR principles in Open Education.

II. Demand for Data Science and Data Skills

Growing demand for Data Science and Analytics enabled and general data-driven professions is confirmed by multiple European and global market studies. The IDG report 2017 [10] provides deep analysis of the European data market and growing demand for data workers. The EU data market volume is estimated as EUR 60 Bln with growth to EUR 106 Bln in 2020. With the total number of data workers to grow 6.1 mln (2016) 10.4 million in 2020 the data worker skills gap is estimated as 769,000 or 9.8% (2020). Addressing this demand and gap is becoming critical for the European economy and a challenge for universities. The report stresses that unsatisfied demand for data workers will lead to an under-performing economy, industry, research and a loss of competitiveness.

The European Commission report 2018 on Skills for Smart Industrial Specialisation and Digital Transformation [11] provides further insight into the development of the mainstream technologies, high-tech skills demand and proposes a strategy for addressing skills sustainability, and also stresses the importance of T-shaped skills profiles, combining deep specialisms with common transversal skills.

Demand for data-related professions will grow even more with the further growth of Industry 4.0 [12] which will deliver tremendous changes to both business models and the labour market with the strong change in the skills set needed to thrive in the new economic landscape. The key Industry 4.0 elements that both empower new data-economy and will be facilitated by the new business and consumer models are listed in Table 1.

Table 1. Top Transversal Skills for Industry 4.0 [9, 10]

Key Industry 4.0 Elements	Top 10 Skills for Industry 4.0
• Cyber-physical systems • Internet of things • Internet of services • Smart factory • Mobile technologies • Cloud computing • Big data	1. Complex problem solving 2. Critical thinking 3. Creativity 4. People management 5. Coordinating with others 6. Emotional Intelligence 7. Judgement and decision making 8. Services orientation 9. Negotiation 10. Cognitive flexibility

The World Economic Forum (WEF) published report “The Future of Jobs” (2016) [13] focused on the employment, skills and workforce strategy for the future economy. The report summarised a vision for the leading high-tech companies and their future skills demand. The following top 10 skills are identified as critical for 2020 (reflecting a shift from currently required skills in the direction of independent critical thinking, creativity, and cognitive flexibility) [14, 15].

III. EDISON Data Science Framework (EDSF)

The EDISON Data Science Framework (EDSF) has been developed within the work packages of the EDISON Project and provides a basis for the Data Science profession definition and links together the main components related to the Data Science education, training, organisational roles definitions and skills management, as well as professional certification and career transferability.

Figure 1 below illustrates the main EDSF components and their inter-relations:

- CF-DS – Data Science Competence Framework [2]
- DS-BoK – Data Science Body of Knowledge [3]
- MC-DS – Data Science Model Curriculum [4]
- DSPP - Data Science Professional profiles and occupations taxonomy [5]
- Data Science Taxonomy and Scientific Disciplines Classification.

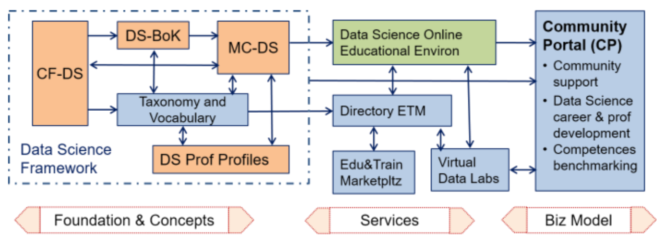

Figure 1. EDISON Data Science Framework components and Data Science Educational environment.

The framework provides the basis for the definition and design of other components of the Data Science professional environment such as

- Data Science Education Environment (DSEE) intended to be cloud based, customizable and aligned with the new workplace practices and skills
- Education and Training Directory connected to Marketplace and Virtual Data Labs
- Data Science Community Portal (CP) that provides information and community support services. It also provides gateway to DSEE, Marketplace and Virtual Data Labs. CP is intended to include tools for individual competences benchmarking and personalized educational path building

A. *Data Science Competence Framework (CF-DS)*

The CF-DS provides the overall basis for the whole framework. The CF-DS includes the core competences required for the successful work of a Data Scientist in different work environments in industry and in research and through the whole career path. The CF-DS is defined using the same approach as e-CFv3.0 [16] (competences defined as abilities supported by knowledge and skills with applied proficiency levels) but have competence structured according to the major identified functional groups (as explained below).

The CF-DS is structured along four dimensions (similar to the European e-Competence Framework e-CFv3.0 [16]) that includes (1) competence groups, (2) individual competences definition, (3) proficiency levels, and (4) corresponding knowledge and skills. In this context, each individual competence includes a set of required knowledge topics and a set of skills type A and skills type B. Such CF-DS structure allows for competence based curriculum design where competences can be defined based on the professional profile (see DSPP [5] for mapping between professional profiles and competences) or target leaners group when designing a full curriculum, or based on competence benchmarking for tailored training to address identified competences and knowledge gaps.

The following core CF-DS competence and skills groups have been identified (refer to CF-DS specification [2] for details):

- Data Science Analytics (including Statistical Analysis, Machine Learning, Data Mining, Business Analytics, others) (DSDA)
- Data Science Engineering (including Software and Applications Engineering, Data Warehousing, Big Data Infrastructure and Tools) (DSENG)
- Data Management and Governance (including data stewardship, curation, and preservation) (DSDM)
- Research Methods and Project Methods (DSRMP)
- Domain Knowledge and Expertise (Subject/Scientific domain related)

Data Science competences must be supported by knowledge that is defined primarily by education and training and skills that are defined by work experience correspondingly. The CF-DS defines two types of skills:

- Skills Type A which are related to the professional experience and major competences, and
- Skills Type B that are related to wide range of practical computational skills including using programming languages, development environment and cloud based platforms (refer to CF-DS [2] for full definition of the identified knowledge and skills groups).

B. *Workplace skills*

Workplace skills, also referred to as “soft” skills or professional attitude skills, are becoming increasing important in the data-driven Industry 4.0 economy.

The CF-DS defined two groups of skills that are demanded by employers and required for the Data Scientist to work efficiently in modern data-driven agile companies:

- Data Science Professional and Attitude skills (thinking and acting like Data Scientist) that defines a special mindset that can be developed by a practicing Data Scientist along their career progression
- 21st Century skills that comprise a set of workplace skills that include critical thinking, communication, collaboration, organizational awareness, ethics, and others.

Universities should pay attention to developing such skills and include them in curricula or extra-curricula activity. Refer to CF-DS for detailed skills definition.

C. *Data Science Body of Knowledge and Model Curriculum*

The DS-BoK defines the Knowledge Areas (KA) and corresponding Knowledge Units (KU) for building Data Science curricula that are required to support identified Data Science competences. DS-BoK is organised by Knowledge Area Groups (KAG) that correspond to the CF-DS competence groups. DS-BoK is based on ACM/IEEE Classification Computer Science (CCS2012) [17], incorporates best practices in defining domain specific BoK’s and provides reference to existing related BoK’s. It also includes proposed new KA to incorporate new technologies and scientific subjects required for consistent Data Science education and training.

The MC-DS [4] is built on DS-BoK and linked to CF-DS where Learning Outcome definitions are based on CF-DS competences (specifically skills type A), 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. Practical curriculum elements should be supported by corresponding educational environment for hands-on labs and educational projects development.

The formal DS-BoK and MC-DS definition creates a basis for Data Science educational and training programmes compatibility, and consequently Data Science related competences and skills transferability.

IV. Data Science Professional Profiles (DSPP) and Competences Mapping

DSPP provides an effective tool for organisational skills management and capacity building: professional profiles are linked to organisational roles and processes; mapping profiles to competences and skills allows for competences and career management.

A. *DSPP profiles definition*

The proposed DSPP definition is based on the analysis of the research and industry demand in data-related professions. The identified professional profiles are classified using ESCO taxonomy [18], and necessary extensions are proposed to support the following hierarchy of data handling related occupations:

- Managers: Chief Data Officer (CDO), Data Science (group/department) manager, Data Science infrastructure manager, Research Infrastructure manager
- Professionals: Data Scientist, Data Science Researcher, Data Science Architect, Data Science (applications) programmer, Data Analyst, Business Analyst, etc.
- Professional (database): Large scale (cloud) database designers and administrators, scientific database designers and administrators
- Professional (data handling/management): Data Stewards, Digital Data Curator, Digital Librarians, Data Archivists

- Technicians and associate professionals: Big Data facilities operators, scientific database/infrastructure operators
- Support and clerical workers: Support and data entry workers.

The individual profiles are defined in accordance with the CWA 16458 (2012): European ICT Professional Profiles [19] standard (and its revision 2018)

The DSPP document defines competences relevance or mastery levels and corresponding learning outcome for the targeted education or training mapped to each profile. This provides direct linking to mastery levels defined in MC-DS and allows easy curricula and courses design [8, 9].

B. *Digital Competences and Data Literacy*

In the context of digital transformation and growing AI-based automation, all professional profiles should possess sufficient levels of digital competences and skills to operate successfully in Industry 4.0. The EC study and report on "Digital Competences for Citizen" (DigComp) published in 2018 [20] provides a good advice for addressing both digital skills and data skills in professional workplace training and vocational education. Important part of digital competences is understanding the role of data and processes related to data handling in modern applications, social media, industrial processes, research, and specifically how data are used in AI based decision making and control. Special training and skills development must be focused on data processing and management issues.

V. Courses to Facilitate Data Science Professional Skills and enable Data Driven Transformation

This section provides examples of supportive courses to facilitate the Data Science professional competences and skills that can be also used as a basis for general digital competences and data literacy training. All described courses have been taught and currently are in the development phase answering demand for research and industry to enable an effective Agile Data Driven Enterprise model (ADDE).

A. *Professional Issues in Data Science*

The goal of this course is to equip students and practitioners with the knowledge and skills for further focused study of more specific Data Science and Analytics areas and courses.
Data Science Competences and Skills management and capacity building, EDISON Data Science Framework.
Data Science professional skills ("Act and think as Data Scientist") and 21st Century Skills.
Data Science and Analytics methods and technologies overview.
Research Methods; Business processes management.
Project management; Agile development process and best practices, PMI process groups.
Data Management in research, industry and personal: standards and best practices.
FAIR (Findable, Accessible, Interoperable, and Re-usable) principles in Open Data and enterprise data management.
Privacy enabling technologies.
Ethical and legal principles and regulations.

It is also beneficial to supply this course with the guided/tutored groups and/or individual training on essential professional skills such as complex problem solving, critical thinking, creativity, etc. defined as critical for Industry 4.0 workforce.

B. *Data Science and Analytics Foundation (DSAF)*

The goal of this course is to introduce students to the whole spectrum of Data Science and Analytics technologies and at the same time provide strong statistical background for future mastering the core data analytics methods and Machine Learning technologies. This defines the main emphasis in the DSAF course on statistical methods, probability theory, hypothesis testing, data preparations, methods of qualitative and quantitative analytics. The primary analytics platform for this course is recommended to have a low programming threshold to enable fast learning. The RapidMiner visual data analytics environment (https://rapidminer.com/) was identified as a good alternative (in same cases preferable choice) against typically used R or python tools, to introduce the trainees into the key data analytics methods and enable active experimentation. Exercises in the course should be also available for Open-Source Python-based data analytic stack.

The following topics are suggested for the DSAF curriculum:
Introduction and course overview: Data Science and Big Data technologies, Data Science competence and skills, Research Methods in Data Science, Machine Learning and Data Mining overview.
Statistical methods and Probability theory
Data description and Statistical Data Analysis
Data preparation: data loading, data cleaning, data pre-processing, parsing, transforming, merging, and storing data
Qualitative and Quantitative data analysis
Classification: methods and algorithms
Cluster analysis basics and algorithms
Performance of data analytics algorithms and tools
Building engaging visualizations of data analysis
Organizing data analysis following CRISP-DM and Data Science Process
Open Data repositories, test datasets, developer communities
Data Management, FAIR Data Principles

It is important to provide such a course at the beginning of the Bachelor or Master Data Science programme and repeat key components in the more specialised courses such as Machine Learning or Data Mining to allow gradual knowledge building that benefits from repetition. DSAF course can be also recommended for other applied Data Science programmes, Elements of the DSAF course can be included into the Data Literacy training. To support students with limited programming background we suggest taking a preparatory course that refreshes key programming techniques, elements of Python language, and basic database operations.

C. Research Methods and Process Management

Research Methods and Project Management is one of the Data Science Competence group and DS-BoK Knowledge Area Group [3]. The Data Science uses research methods in its foundation to drive experiment based hypothesis validation, taking insight into or extracting value from the business or industrial data.

Besides the critical importance of understanding and applying different research methods and experiment design, a Data Scientist must have a comprehensive knowledge and ability to effectively use existing process models that represent best practices in solving practical Data Analytics tasks and provide a basis for the project organisation and management. A well-defined process model should answer the following questions:

- What is the whole process to obtain information from the data? What are stages?
- How are datasets transformed at each stage? How are datasets organised, accessed and stored at each stage?
- What is done in each stage? What roles and activities are involved?
- Which techniques must be applied?
- Which technology and tools must be employed?
- How are the results evaluated and quality ensured?

The following presents the reference and a short description of the popular data analysis process models.

1) CRISP-DM, CRoss-Industry Standard Process for Data Mining [21]

The CRISP-DM was the first model to formalise the data mining process and its relevance to the business processes and intended actionable outcome. The following 6 sequential phases are defined which can be also organised into iterative/continuous development and improvement cycle:

Business understanding: Understanding business processes, problem definition, required outcome.

Data understanding: Obtaining available data, data inspection, data preparation, data observation and initial hypothesis formulation.

Data preparation: Prepare datasets for analysis by extracting necessary data from the raw data.

Modelling: Create one or several data models, select corresponding techniques, possibly test and experiment with several to assess performance.

Evaluation: Ensure that the solution fulfils the business objectives established in the first phase, and to obtain this certainty a in deep evaluation of the model must be performed during this phase. Evaluated if the knowledge obtained with the model has the desired value for the customer.

Deployment: Solution deployment in a production environment for customer, providing the results in easy understandable form to customer.

2) ASUM, Analytics Solutions Unified Method (IBM)[22]

ASUM is an hybrid of agile and traditional process model. It has five (5) sequential phases and a set of processes to manage and monitor the progress and maintenance of the project:

Analyze. Requirements specified and agreed; contract or services agreement is signed.

Design. Define all components of the solution and their relationships and dependencies, identify necessary resources.

Configure and Build. The solution is developed all components are integrated and configured.

Deploy. Create a plan to run and maintain the developed solution, including configuration management and migration plan if necessary.

Operate and Optimize. The solution is operational is monitoring data are collected and maintained.

3) TDSP, Team Data Science Process (Microsoft) [23]

TDSP is an agile and iterative process model. It has five sequential phases: Business Understanding, Data acquisition and understanding, Modeling, Deployment, Customer acceptance. TDSP framework document provides valuable information for Data Science applications developers for planning and managing the whole process from development to deployment and operation.

4) KNIME Model Factory (KMF) [24]

KMF defines the process model specifically adopted to Data Mining and Machine Learning processes that includes similar stages such as Init, Load, Transform, Learn, Score, Deploy.

VI. DATA MANAGEMENT AND DATA STEWARDSHIP

Establishing effective Data Management and Data Governance (DMG) in organisation is considered as a first step in the digital transformation. Best practices in DMG are well defined by the DAMA (Data Management Association) and published as Data Management Body of Knowledge (DMBOK) [25] that defines a set of knowledge, competences and responsibilities of the main organisational roles and actors in the Data Management and Governance. The DNV GL Data Quality Assessment Framework [26] provides a set of industry best practices recommendations on how to achieve best use of company's data resource, converting them into assets and bringing competitive benefits. Widely adopted in research community, the FAIR data principles [27, 28] and Data Stewardship competence framework [29] provide a good contribution to building practically oriented DMG curricula. Below are outlines of the two distinctive courses that are related to DMG: Enterprise Data Management and Governance and Research Data Management and Stewardship.

A. Data Management and Governance (enterprise scope)

The DMG course uses DMBOK as general framework covering majority of topics, extending them with the Data Science and Big Data Analytics platforms and enriching with the FAIR and industry best practices. The following are the main topics that can be included in the course:

- Introduction. Big Data Infrastructure and Data Management and Governance.
- Data Management concepts. Data management frameworks: DAMA Data Management framework, the

Amsterdam Information Model. Extensions for Big Data and Data Science.

- Enterprise Data Architecture. Data Lifecycle Management and Service Delivery Model. Data management and data governance activities and roles. Data Science Professional profiles and organisational roles, Skills management and capacity building.
- Data Architecture, Data Modelling and Design. Data types and data models. Data modeling. Metadata. SQL and NoSQL databases overview. Distributed systems: CAP theorem, ACID and BASE properties.
- Enterprise Big Data infrastructure and integration with enterprise IT infrastructure. Data Warehouses. Distributed file systems and data storage.
- Big Data storage and platforms. Cloud based data storage services: data object storage, data blob storage, Data Lakes (services by AWS, Azure, GCP). Trusted storage, blockchain enabled data provenance.
- FAIR data principle and Data Stewardship, Data Quality assessment and maturity model. Data repositories, Open Data services, public services.
- Big Data Security and Compliance. Data security and data protection. Security of outsourced data storage. Cloud security and compliance standards and cloud provider services assessment.

The outlined topics above can be included in the practical courses for different target groups and at the different competence levels from Data literacy courses to professional training and academic curricula.

B. Research Data Management and Stewardship (RDMS)

The research data management has numerous implementations and well supported with training materials but in most cases, this is focused on the specific scientific domain. The courses also include growing popular FAIR principles and Data Stewardship related topics.

The following RDMS course example is structured along practical aspects of the research data management.

A. Use cases for data management and stewardship

- Preserving the Scientific Record

B. Data Management elements (organisational and individual)

- Goals and motivation for managing your data
- Data formats, Metadata, related standards
- Creating documentation and metadata, metadata for discovery
- Using data portals and metadata registries
- Tracking Data Usage, data provenance, linked data
- Handling sensitive data
- Backing up data, backup tools and services
- Data Management Plan (DMP)

C. Responsible Data Use (Citation, Copyright, Data Restrictions)

- Data privacy and GDPR compliance

D. FAIR principles in Research Data Management, supporting tools, maturity model and compliance

E. Data Stewardship and organisational data management

- Responsibilities and competences
- DMP management and data quality assurance

F. Open Science and Open Data (Definition, Standards, Open Data use and reuse, open government data)

- Research data and open access
- Repository and self- archiving services
- RDA products and recommendations: PID, data types, data type registries, others
- ORCID identifier for data and authors
- Stakeholders and roles: engineer, librarian, researcher
- Open Data services: ORCID.org, Altmetric Doughnut, Zenodo

G. Hands on practice includes the following topics:

a) Data Management Plan design

b) Metadata and tools

c) Selection of licenses for open data and contents (e.g. Creative Common, and Open Database)

VII. Conclusion and Further Developments

EDSF provides a common semantic basis for interoperability of all forms of the Data Science curriculum definition and education or training delivery, as well as knowledge assessment based on fully enumerated definitions for EDSF components and individual units. Besides defining academic components of the effective and consistent curriculum, EDSF also provides advice on the required Data Science Education Environment to facilitate fast practical knowledge and skills acquisition by students and learners.

An effective professional education needs to provide a foundation for future continuous professional self-development and the mastering of new emerging technologies, that can provide a basis for the life-long learning model adoption. Wide use of available online resources and platforms for demanded Data Science and other digital and data skills will facilitate the adoption of the FAIR principles Findable, Accessible, Interoperable, and Re-usable, to enable Open Education to follow the vision originally proposed for Open Data [26]. Universities can contribute to building a FAIR life-long educational space that can serve both organisational and individual needs of students and learners, including support for widely apprised citizen scientists.

The EDSF, and its proposed further integration with the Data Science Education Environment, offers an expanded meta-framework that will enable efficient configuration of education and training solutions for highly demanded Data Science and Analytics competences and skills.

Acknowledgment

The research leading to these results has received funding from the Horizon2020 projects MATES (grant n. 591889), FAIRsFAIR (grant n. 831558) and EDISON (grant n. 675419).

References

[1] EDISON Data Science Framework (EDSF). Available at http://edison-project.eu/edison/edison-data-science-framework-edsf

[2] Data Science Competence Framework. Available at http://edison-project.eu/data-science-competence-framework-cf-ds
[3] Data Science Body of Knowledge. Available at http://edison-project.eu/data-science-body-knowledge-ds-bok
[4] Data Science Model Curriculum. Available at http://edison-project.eu/data-science-model-curriculum-mc-ds
[5] Data Science Professional Profiles. Available at http://edison-project.eu/data-science-professional-profiles
[6] Demchenko, Yuri, Emanuel Gruengard, Sander Klous, Instructional Model for Building effective Big Data Curricula for Online and Campus Education. 1st IEEE STC CC and RDA Workshop on Curricula and Teaching Methods in Cloud Computing, Big Data, and Data Science, in Proc. The 6th IEEE International Conference and Workshops on Cloud Computing Technology and Science (CloudCom2014), 15-18 December 2014, Singapore.
[7] Manieri, Andrea 2015, 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
[8] Yuri Demchenko, Adam Belloum, Wouter Los, Tomasz Wiktorski, Andrea Manieri, Steve Brewer, Holger Brocks, Jana Becker, Dominic Heutelbeck, Matthias Hemmje, EDISON Data Science Framework: A Foundation for Building Data Science Profession For Research and Industry, 3rd IEEE STC CC and RDA Workshop on Curricula and Teaching Methods in Cloud Computing, Big Data, and Data Science (DTW2016), in Proc. The 8th IEEE International Conference and Workshops on Cloud Computing Technology and Science (CloudCom2016), 12-15 December 2016, Luxembourg.
[9] Yuri Demchenko, Luca Comminiello, Gianluca Reali, Designing Customisable Data Science Curriculum using Ontology for Science and Body of Knowledge, 2019 International Conference on Big Data and Education (ICBDE2019), March 30 - April 1, 2019, London, United Kingdom, ISBN978-1-4503-6186-6/19/03.
[10] Final results of the European Data Market study measuring the size and trends of the EU data economy, EC-IDC, March 2017 [online] https://ec.europa.eu/digital-single-market/en/news/final-results-european-data-market-study-measuring-size-and-trends-eu-data-economy
[11] Skills for Smart Industrial Specialisation and Digital Transformation Interim Report, European Commission, Written by PwC, November 2018
[12] The Future of Jobs, World Economic Forum Report, 18 January 2016 [online] http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf
[13] The 10 skills you need to thrive in the Fourth Industrial Revolution [online] https://www.weforum.org/agenda/2016/01/the-10-skills-you-need-to-thrive-in-the-fourth-industrial-revolution/
[14] Are you ready for Industry 4.0. [online] http://www.delivered.dhl.com/en/articles/2017/02/skills-for-industry-4-0.html
[15] e-CF3.0, 2016 European e-Competence Framework 3.0. A common European Framework for ICT Professionals in all industry sectors. CWA 16234:2014 Part 1. Available at http://ecompetences.eu/wp-content/uploads/2014/02/European-e-Competence-Framework-3.0_CEN_CWA_16234-1_2014.pdf
[16] CCS, 2012 The 2012 ACM Computing Classification System. Available at http://www.acm.org/about/class/class/2012
[17] European Skills, Competences, Qualifications and Occupations (ESCO) framework. Available at https://ec.europa.eu/esco/portal/#modal-one
[18] European ICT Professional Profiles CWA 16458 (2012) (Updated by e-CF3.0) [online] http://relaunch.ecompetences.eu/wp-content/uploads/2013/12/EU_ICT_Professional_Profiles_CWA_updated_by_e_CF_3.0.pdf
[19] DigComp 2.1, 2017, The Digital Competence Framework for Citizens, by Stephanie Carretero, Riina Vuorikari and Yves Punie, Joint Research Center, 2017, EUR 28558 EN [online] https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/digcomp-21-digital-competence-framework-citizens-eight-proficiency-levels-and-examples-use
[20] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, R. Wirth, CRISP-DM 1.0, Cross Industry Standard Process for Data Mining Reference Model, http://edison... (2000).
[21] IBM, ASUM, Analytics solutions unified method [online] https://www-01.ibm.com/marketing/iwm/iwm/web/pick.do?source=swerpba- basimext&lang=en US
[22] TSDP, Team Data Science Process, https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/overview
[23] P.Winters, M.Berthold, KMF, KNIME Model Factory, [online] https://www.knime.com/blog/the-knime-model-process-factory
[24] DAMA – DMBOK: Data Management Body of Knowledge, 2nd Edition, 2017. DAMA International, Technics Publications Llc
[25] DNVGL-RP-0497 Data quality assessment framework, Edition January 2017 [online] https://rules.dnvgl.com/docs/pdf/DNVGL/RP/2017-01/DNVGL-RP-0497.pdf
[26] Barend Mons, et al, The FAIR Guiding Principles for scientific data management and stewardship [online] https://www.nature.com/articles/sdata201618
[27] Turning FAIR into reality. Final Report and Action Plan from the European Commission Expert Group on FAIR Data, 2018 [online] https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf
[28] EOSCpilot Deliverable D7.5: Strategy for Sustainable Development of Skills and Capabilities [online] https://eoscpilot.eu/content/d75-strategy-sustainable-development-skills-and-capabilities