EDISON Data Science Framework:

Part 3. Data Science Model Curriculum (MC-DS)

Release 4 (EDSF04 or EDSF2022)

EDISON Community Initiative
(Maintaining the H2020 EDISON project outcome)

Release Date	31 December 2022
Document Editor/s	Yuri Demchenko
Version	Release 4, v07
Status	Working document, request for comments

Document Version Control

	Version	Date	Change Made (and if appropriate, the reason for change)	Initials of Commentator(s) or Author(s)
Release 1	01	10/10/2016	Release 1 after ELG03 meeting discussion	YD
Release 2	03	03/07/2017	Release 2 document (updated after multiple discussions and comments, feedback from champion universities, ELG04 comments)	TW, YD
Pre-Release 3	04	07/09/2018	Pre-release 3. Definition of courses templates is revised. New courses included.	YD, TW
Release 3	05	31/12/2018	Release 3. Release document updated based on received comments and feedback from practical implementation.	YD, TW
Pre-Release 4	06	31/12/2020	Pre-release 4. Pre-release document updated based on received comments and feedback from practical implementation and design by November 2019.	YD, TW
Release 4	07	31/12/2022	Release 4. Release document updated based on received comments and feedback from practical implementation and design since 2018	YD, JJCG

Contributors

Document Editors: Yuri Demchenko		
Author Initials	Name of Contributor	Institution
YD	Yuri Demchenko	University of Amsterdam
AB	Adam Belloum	University of Amsterdam
AM	Andrea Manieri	Engineering
TW	Tomasz Wiktorski	University of Stavanger
JJCG	Cuadrado Gallego Juan José	Alcala University

Acknowledgement

Gábor Kismihók, Center of Job Knowledge Research, University of Amsterdam
Oleg Chertov, Professor, Sikorski National Technical University of Ukraine “Kyiv Polytechnic Institute”

Executive summary

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 lack 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.

The presented Data Science Model Curriculum is a part of the EDISON Data Science Framework (EDSF), providing 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), Data Science Professional Profiles definition (DSPP), EDSF Use cases and application (EDSF-UCA).

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 suggested Learning Units are assigned suggested labels, marking their relevance to the core Data Science knowledge areas in the form of Tier 1, Tier 2, or Elective courses. Further MC-DS refinement will be done based on consultation with the university community and experts both in Data Science and scientific or industry domains.

The proposed MC-DS intends to provide guidance to universities and training organisations in the construction of Data Science programmes and individual courses selection that are balanced according to requirements elicited from the research and industry domains. MC-DS can be used for the assessment and improvement of existing Data Science programmes with respect to the knowledge areas and competence groups that are associated with specific professional profiles. When coupled with individual or group competence benchmarking, MC-DS can also be used for building individual training curricula and professional (self/up) skilling for effective career management.

Further work will be required to develop consistent MC-DS that can be used by the academic community and professional training community. The proposed version is intended to initiate community discussion and solicit a contribution from the subject matter expects and practitioners.

The EDSF documents are available for public discussion at the EDISON Community initiative at https://github.com/EDISONcommunity/EDSF/wiki/EDSFhome

TABLE OF CONTENTS

1 Introduction ... 6
2 EDISON Data Science Framework (EDSF) ... 8
3 Overview of Best Practices in Curricula Design ... 10
3.1 Learning models and curriculum design approaches ... 10
3.1.1 Bloom's Taxonomy ... 10
3.1.2 Constructive Alignment and Problem-based Learning ... 13
3.1.3 Competence Based Learning Model ... 13
3.2 ACM Computer Science Curriculum (CS2013) and Body of Knowledge (CS-BoK) ... 14
3.3 ACM/IEEE-CS Curricula Guidelines and Competency Model for Information Technologies ... 15
3.4 ICT professional Body of knowledge and new curricula for e-Leadership skills ... 16
4 Data Science Model Curriculum Design Approach ... 17
4.1 Linking DS-BoK Knowledge Areas and MC-DS Learning Units for target Competence Groups ... 17
4.2 Mastery levels and Learning Outcomes ... 17
4.3 Learning Outcomes definition based on CF-DS ... 19
4.4 Definition of MC-DS Learning Units ... 24
5 Data Science Model Curriculum (MC-DS) ... 25
5.1 Organization and Application of Model Curriculum ... 25
5.1.1 Organization of Model Curriculum ... 25
5.1.2 Application of Model Curriculum ... 26
5.2 Assignments of ECTS points to Competence Groups and Knowledge Areas ... 27
5.3 Data Science Data Analytics (KAG1 – DSDA) related courses ... 28
5.3.1 DSDA/SMDA - Statistical methods and data analysis ... 28
5.3.2 DSDA/ML – Machine Learning ... 29
DSDA/DM - Data Mining ... 29
DSDA/TDM - Text Data Mining ... 30
5.3.3 DSDA/PA - Predictive Analytics ... 30
5.3.4 DSDA/MODSIM - Modelling, simulation and optimization ... 30
5.4 Data Science Engineering (KAG2-DSENG) ... 31
5.4.1 DSENG/BDI - Big Data infrastructure and technologies ... 31
5.4.2 DSENG/DSIAPP - Infrastructure and platforms for Data Science applications ... 32
5.4.3 DSENG/CCT - Cloud Computing technologies for Big Data and Data Analytics ... 32
5.4.4 DSENG/SEC - Data and Applications Security ... 32
5.4.5 DSENG/BDSE - Big Data systems organization and engineering ... 33
5.4.6 DSENG/DSAPPD - Data Science (Big Data) application design ... 33
5.4.7 DSENG/IS - Information Systems ... 33
6 Example of using EDSF for Curricula Design and Evaluation ... 35
6.1 Designing a new programme ... 35
6.2 Assessment of existing programmes and identification of potential gaps ... 39
7 Conclusion and further developments ... 40
7.1 Summary of the recent developments ... 40
7.2 Further developments to formalize MC-DS ... 40
8 References ... 42
9 Acronyms ... 44
Appendix A. Mastery levels ... 45
Appendix B. Subset of ACM/IEEE CCS2012 for Data Science ... 48
B.1. ACM Classification Computer Science (2012) structure and Data Science related Knowledge Areas ... 48
Appendix C. Example ECTS points assignment to different Data Science Professional groups ... 52

1 Introduction

Data Science is an emerging field of science, which requires a multi-disciplinary approach and should be built with a strong link to Big Data and data driven technologies that create transformational effect to all research and industry domains, and consequently require re-thinking and re-design of both traditional educational models and existing courses. However, at the present time most of the existing university curricula and training programs are built based on available courses and cover a limited set of academic subjects related to a full Data Science Body of Knowledge covering only a limited set of knowledge areas and 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 into the real working environment (both in industry and academia).

The Data Science Model Curriculum is the part of the EDISON Data Science Framework that includes the following parts: Data Science Competence Framework (CF-DS or Competence Framework), Data Science Body of Knowledge (DS-BoK or Body of Knowledge), Data Science Model Curriculum (MC-DS or Model Curriculum), Data Science Professional Profiles (DSPP) definition, EDSF Use cases and applications (EDSF-UCA).

The proposed Data Science Model Curriculum reuses the best practices in curriculum design and a new educational model to facilitate the students learning as well as existing staff professional training and re-skilling for data related technologies. Building on insights gathered through thorough analyses of existing Data Science programmes (performed in the EDISON project[1]) and the requirements of targeted educational stakeholders, the Model Curriculum reflects by design the whole data handling/processing lifecycle and organizational or structural processes (such as scientific methods and data driven research cycle, business process management cycle as defined in CF-DS document [1]).

The presented MC-DS Release 3 contains example of the suggested courses (with approximate credits points distribution) for two groups: Data Science and Analytics (DSDA) and Data Science Engineering (DSENG). Further courses definition for Data Management and Stewardship (DSDM) and Research Methods and Project Management (DSRMP) will be provided in the new MC-DS releases.

The definition of the MC-DS can be used as instrumental in defining recommended training for Data Science professional certification programs. From the practical perspective, the Model Curriculum represents a tool for

i) supporting the development of new Data Science programmes (including appraisal/selection of appropriate units/modules) tailored according to proficiency levels required to address competences required for identified Data Science Professional profiles, and
ii) assessing the coverage of existing Data Science programmes, facilitating the elicitation of potential gaps w.r.t. to specific competence groups and knowledge areas implied by targeted professional profiles.

By its design, the Model Curriculum helps match the supply-side and demand-side requirements for Data Science education. The formal definition of the Data Science Model Curriculum intends to create a basis for the Data Science educational and training programmes compatibility and consequently, Data Science related competences and skills transferability.

Further work will be required to develop consistent MC-DS that can be used by the academic community and professional training community. The presented MC-DS version should facilitate Data Science curriculum harmonisation and contribution from the subject matter expects and practitioners. The initial discussions of the MC-DS have been presented to the EDISON Liaison Groups of experts and will undergo further community discussion via the EDISON community forum and through presentations at community oriented workshops and conferences.

The presented document has the following structure. Section 2 provides an overview of the EDISON Data Science Framework and related project activities that support the framework components development and pilot implementation. Section 3 provides overview of existing BoKs related to Data Science knowledge areas. Section 3 also refers to best practices in curriculum design such as Bloom's Taxonomy, problem and

[1] Refer to the EDISON Project Deliverable D2.2 Existing educational and training resources inventory and analysis, 31 May 2016 [6, 38]

competence based learning models. Section 4 briefly discusses the DS-BoK design principles and provides definition of the Learning Outcomes related to CF-DS competence Section 5 describes the MC-DS organisation and provides an example definition of the courses related to main Knowledge Areas Groups and Knowledge Areas as they are defined in the DS-BoK [2]. Section 6 provides examples how the proposed MC-DS can be used in practice for Data Science programmes and courses assessment. Section 7 provides a summary of the achieved results. Appendices contain necessary supplementary information such as Classification Computer Science (CCS2012) and exception from the DS-BoK necessary for MC-DS understanding and use.

2 EDISON Data Science Framework (EDSF)

The EDISON Data Science Framework provides a basis for the definition of the Data Science profession and enables the definition of the other components related to Data Science education, training, organisational roles definition and skills management, as well as professional certification.

Figure 2.1 below illustrates the main components of the EDISON Data Science Framework (EDSF) and their inter-relations that provides the conceptual basis for the development of the Data Science profession:

- CF-DS – Data Science Competence Framework (this document [1])
- DS-BoK – Data Science Body of Knowledge [2]
- MC-DS – Data Science Model Curriculum [3]
- DSPP - Data Science Professional profiles and occupations taxonomy [4]
- Data Science Taxonomy and Scientific Disciplines Classification

The proposed framework provides a basis for other components of the Data Science professional ecosystem[2] , such as

- EDISON Online Education Environment (EOEE)
- Education and Training Directory and Marketplace
- 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

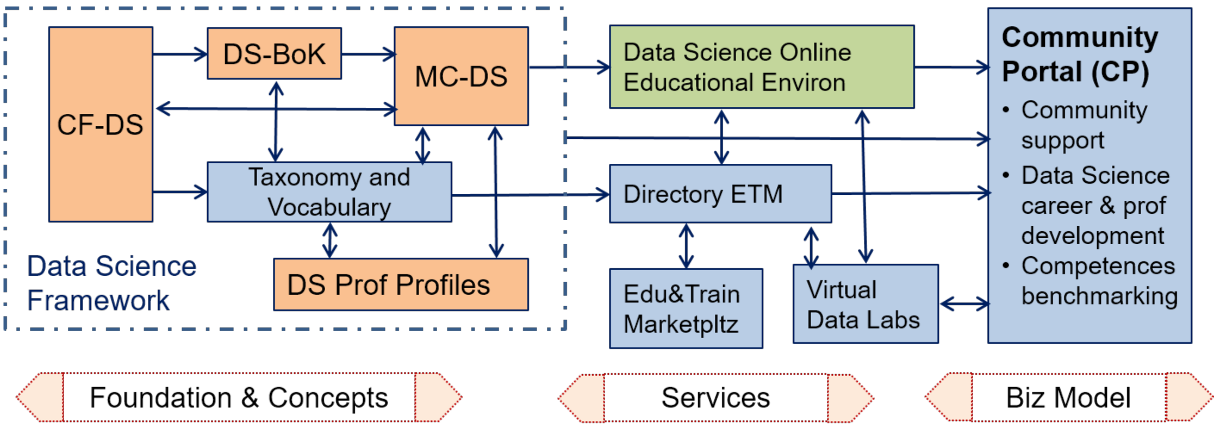

Figure 2.1 EDISON Data Science Framework components and Data Science professional ecosystem.

The EDSF Release 4 includes Part 5 EDSF Use cases and applications [5] which describes a few uses of using EDSF by universities and professional education and training organisations as well as subject domain communities; the guidelines part provides recommendations on using EDSF for practical cases of defining new domain specific competence profiles, knowledge areas and model curricula.

The CF-DS provides the overall basis for the whole EDSF. The core CF-DS includes common competences required for the successful work of a Data Scientist in different work environments in industry and in research and throughout the whole career path. The future CF-DS development may include coverage of the domain specific competences and skills by involving domain and subject matter experts, which may be published as separate CF-DS profiles[3].

[2] The described Data Science ecosystem components are defined and piloted in the EDISON project and constitute the project legacy that can be re-used and followed by the community.
[3] Data Stewardship Professional Competence Framework (CF-DSP) has been developed by the FAIRsFAIR project by extending CF-DS with the Data Stewardship and FAIR related competences and skills and published as a separate document referring to the core EDSF documents [6]

The DS-BoK defines the Knowledge Areas (KA) 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. Knowledge Areas are composed of a number of Knowledge Units (KU) which are currently the lowest component of the DS-BoK. DS-BoK incorporates best practices in Computer Science and domain specific BoKs and includes KAs and KUs defined where possible based on the Classification Computer Science (CCS2012) [7], components taken from other BoKs and proposed new KAs/KUs 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 (LO) 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 proposed Learning Outcomes are enumerated to have a direct mapping to the enumerated competences in CF-DS.

The DSPP professional profiles are defined as an extension to the European Skills, Competences, Qualifications and Occupations (ESCO) taxonomy [30] using the ESCO top classification groups. DSPP definition provides an important instrument to define effective organisational structures and roles related to Data Science positions and can also be used for building individual career paths and corresponding competences and skills transferability between organisations and sectors.

The Data Science Taxonomy and Scientific Disciplines Classification will serve to maintain consistency between four core components of EDSF: CF-DS, DS-BoK, MC-DS, and DSP profiles. To ensure consistency and linking between EDSF components, all individual elements of the framework are enumerated, in particular: competences, skills, and knowledge topics in CF-DS, knowledge groups, areas and units in DS-BoK, learning outcomes and learning units in MC-DS, and professional profiles in DSPP.

It is anticipated that successful acceptance of the proposed EDSF and its core components will require standardisation and interaction with the European and international standardisation bodies and professional organisations. This work is being done as a part of the EDSF sustainability support by the EDISON community initiative provided by the University of Amsterdam[4].

The EDISON Data Science professional ecosystem illustrated in Figure 2.1 shows how the core EDSF components may be related to the potential services that can be offered for the professional Data Science community and provide a basis for sustainable Data Science competences and skills management by organisations, in particular in conditions of emerging Industry 4.0, growing digitalisations and Artificial Intelligence development. As an example of practical use, CF-DS and DS-BoK can be used for individual competences and knowledge benchmarking and play an instrumental role in constructing personalised learning paths and professional (up/re-) skilling programs based on MC-DS.

[4] EDISON Community Initiative website https://edisoncommunity.github.io/EDSF/

3 Overview of Best Practices in Curricula Design

This section provides background information and best practices in building effective professional curricula for specific domains of knowledge, target groups and purposes. The reviewed selected learning model and curricula design models are used to develop the EDISON approach that is targeted to provide quality education and training for specific groups of Data Science related professions to acquire necessary competences and skills.

The following curricula and bodies of knowledge have been reviewed to identify best practices and components to be used for the initial definition of the MC-DS structure and content:

- ACM Computer Science Curriculum and Body of Knowledge (ACM CS2013 and CS-BoK) [8]
- Information Technology Competency Model of Learning Outcome ACM CCECC2014 [9]
- ICT professional Body of Knowledge and ICT leadership curriculum (ICT-BoK) [10]

Other relevant BoKs that were used in defining the DS-BoK are reviewed in the corresponding DS-BoK document [2];

- Data Management Body of Knowledge (DM-BoK) by Data Management Association International (DAMAI) [11]
- Software Engineering Body of Knowledge (SWEBOK) [12]
- Business Analytics Body of Knowledge (BABOK) [13]
- Project Management Professional Body of knowledge (PM-BoK) [14]

It is important to mention that due to complex nature of the Data Science profession consisting of few quite different knowledge areas, the MC-DS definition requires combination of the elements form different BoKs and using different approaches to the curriculum definition, moreover the MC-DS should allow different learning models and adaptation to different subject domains. The final curriculum definition will depend on local conditions defined by the job market demand side (i.e. employers, industry), available teaching staff and expertise, and available educational base and infrastructure.

3.1 Learning models and curriculum design approaches

To define the MC-DS consistently, we need to understand the commonly accepted approaches to defining education and training programmes and put them in the context of the European education system and policies, and also consider alignment with international practices. Two approaches to education and training are followed in practice, the traditional approach which is based on defining the time students have to spend learning given topics or concepts like the European Credit Transfer and Accumulation System (ECTS) [15] or Carnegie unit credit hour [16]. The former is also known as competence-based education or outcomes-based education (OBE), it focuses on the outcome assessing whether students have mastered the given competences, namely the skills, abilities, and knowledge. There is no specified style of teaching or assessment in OBE; instead, classes, opportunities, and assessments should all help students achieve the specified outcomes. In 2012, the EC called for a rethinking of education towards the OBE approach. The motivation for such a rethinking is to ensure that education is more relevant to the needs of students and the labour market, and assessment methods need to be adapted and modernised. Not like the traditional BoK which is defined in term of Knowledge Areas (KA), in OBE, the BoK and curriculum are defined in terms of the core learning outcomes, which are grouped into technical competence areas and workplace skills.

3.1.1 Bloom's Taxonomy

Bloom's taxonomy [17] provides a conceptual framework to organize levels of learning of a topic or subject, and assigns action verbs to each level that help to understand activities related to a particular level of learning. Figure 3.1 below illustrates Bloom's Taxonomy learning levels. For instance, students start at the *knowledge* level when they can *name* and *identify* relevant technologies. To further move to the *comprehension* level when they can *explain* how technologies work. They can then move to the *application* level when they can *choose the* right technology to *solve* a problem. Further, they can progress to *analysis*, *synthesis*, and finally *evaluation* levels.

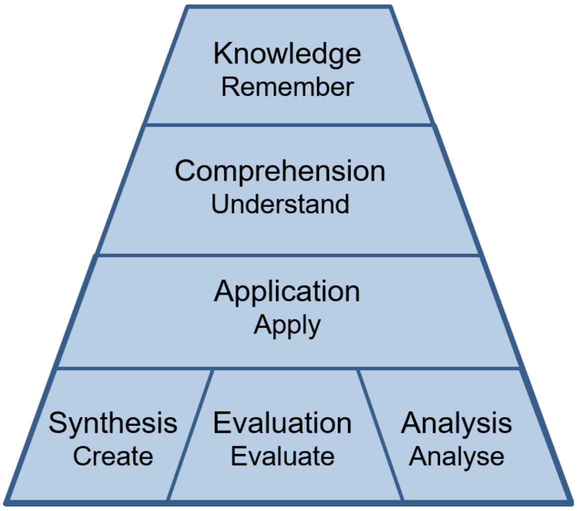

Figure 3.1. Simple Bloom's taxonomy: Learning levels and action verbs.

The below example shows typical attributes of the different levels of learning and example questions to test these levels.

Knowledge
Exhibit memory of previously learned materials by recalling facts, terms, basic concepts and answers
Knowledge of specifics - terminology, specific facts
Knowledge of ways and means of dealing with specifics - conventions, trends and sequences, classifications and categories, criteria, methodology
Knowledge of the universals and abstractions in a field - principles and generalizations, theories and structures
Questions like: What are the main benefits of implementing Big Data and data analytics methods for organisation? What are the main methods and algorithms in supervised Machine Learning.

Comprehension
Demonstrate understanding of facts and ideas by organizing, comparing, translating, interpreting, describing, and stating the main ideas
Translation, Interpretation, Extrapolation
Questions like: Compare the business and operational models of private clouds and hybrid clouds.

Application
Using new knowledge. Solve problems in new situations by applying acquired knowledge, facts, techniques and rules in a different way
Questions like: What data analytics methods should be applied for specific data types analysis or for specific business processes and activities? Which Big Data services architecture is best suited for medium size research organisation or company, and why?

Analysis
Examine and break information into parts by identifying motives or causes. Make inferences and find evidence to support generalizations
Analysis of elements, relationships, organizational principles
Questions like: What data analytics methods and services are required to support typical business processes of a web trading company? Give suggestions on how these services can be implemented with the selected data analytics platform, including on-premises or outsourced to the cloud. Provide references to support your statements.

Synthesis
Compile information together in a different way by combining elements in a new pattern or proposing alternative solutions

Production of a unique communication, a plan, or a proposed set of operations, derivation of a set of abstract relations
Questions like: Describe the main steps and tasks for implementing data analytics and data management services for an example company or research organisation? What services and data analytics can be moved to the clouds and which will remain at the enterprise premises and be run by the company's personnel?

Evaluation
Present and defend opinions by making judgments about information, the validity of ideas or quality of work based on a set of criteria
Judgments in terms of internal evidence or external criteria
Questions like: Do you think that implementing the Agile Data Driven Enterprise model creates benefits for enterprises, short term and long term?

Figure 3.2. provides a consolidated presentation of Bloom's Taxonomy [17, 18] structure, attributes and action verbs that can be effectively used for designing effective curricula and knowledge evaluation. When designing Learning Outcomes for a course or program, it is essential to ensure that all levels will be adequately covered. Consideration of Bloom's taxonomy assists instructors both on the design phase of a course or program, and during the grading process. It is a reliable and simple method to distinguish e.g. between familiarity with many concepts and actually being able to use them in a practical setting.

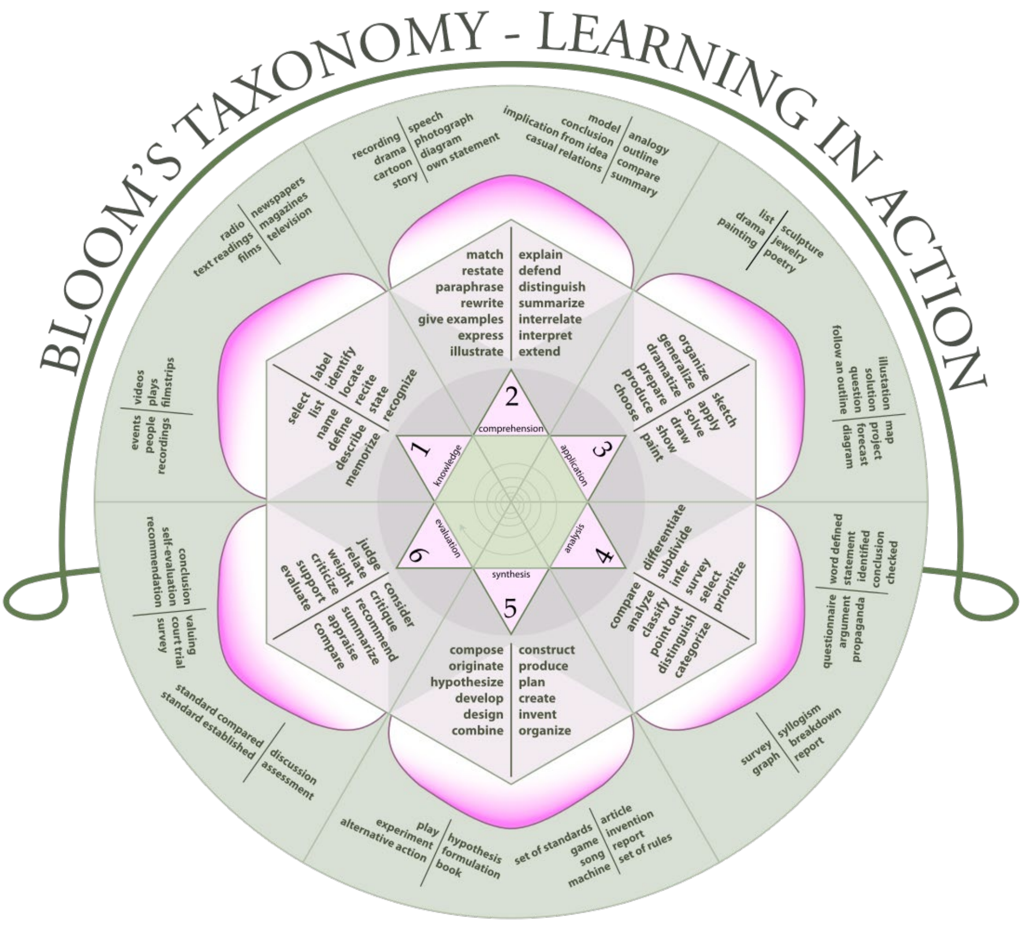

Figure 3.2. **Extended Bloom's taxonomy[5]: consolidated presentation of learning levels, action verbs, and associated learning instruments** (https://en.wikipedia.org/wiki/Bloom%27s_taxonomy)

[5] CC BY-SA 3.0 K. Aainsqatsi

3.1.2 Constructive Alignment and Problem-based Learning

The traditional and still usual approach in science and engineering education is based on a behaviorist or objectivist epistemology, in which the student is passively imparted with knowledge by the teacher. Students' participation in the learning process is limited to memorizing schemes given by the instructor, which are assessed through instruments such as examinations and quizzes that measure the degree of conformance to a norm instead of actual competences [19]. In contrast, a constructivist epistemology puts the student in the center of the learning process as an active participant in constructing knowledge [20].

Problem Based Learning (PBL) [21, 22] is an alternative approach to instruction based on providing student with a non-trivial problem to solve, and guidance in obtaining the necessary competencies. PBL is underlined by a constructivist epistemology that emphasizes active student participation in the construction of their knowledge from learning activities and motivating them through careful alignment of evaluation activities, leading to a concept called Constructive Alignment described by Biggs [23]. Ben-Ari [24] describes the applicability of constructivism to computer science education. Despite certain differences in epistemology between computer science and other sciences, constructivism is a useful approach to computer science education.

From the perspective of a whole education program, constructive alignment and problem-based learning can be implemented in the form of project-based learning. In such a model, regular classes provide students with competences related to specific knowledge areas, while additional project classes allow to establish a link between these competences. In addition, project classes provide an opportunity to reach higher levels of learning. An example of such an approach on an institutional scale is the University of Aalborg [25].

These education concepts provide guidance for further definition of Learning Outcomes and finally Model Curricula, and can be used for the evaluation of the existing programmes.

3.1.3 Competence Based Learning Model

Competency Based Learning (CBL) or Competence Based Education (CBE), also known as outcomes based learning 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 CBL assesses whether students have mastered the given competencies, namely the knowledge, skills, and abilities [9]. 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 CBL 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 the 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 the 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 lifelong learning and self-re-skilling dictated by current fast technologies development. The CBL programmes should offer the following features [26]:

- Self-pacing
- Modularization
- Effective assessments
- Intentional and explicit learning objectives shared with the student,
- Anytime/anywhere access to learning objects and resources,
- Personalized, adaptive or differentiated instruction
- Learner supports through instructional advising or coaching.

Although there are many examples of universities using CBL/CBE model, its practical implementation may create problems in some universities. Paper [27] formulates the following principles that would allow integrating CBE into existing campus structures:

- The degree reflects robust and valid competencies.
- Students are able to learn at a variable pace and are supported in their learning.
- Effective learning resources are available any time and are reusable.
- Assessments are secure and reliable.

It is apparent that CBL is well suited for the professional education and training of one of the EDISON target groups, the self-made or practicing Data Scientists. It is admitted [26] that the CBL was actually created to address the needs of non-traditional students who cannot devote their full time to traditional academic study as well as an effective model for companies to provide (re/up) skilling their staff.

3.2 ACM Computer Science Curriculum (CS2013) and Body of Knowledge (CS-BoK)

In the ACM-CS2013-final report [8] the Body of Knowledge is defined as a specification of the content to be covered in a curriculum as an implementation. The ACM-BoK describes and structures the knowledge areas needed to define a curriculum in Computer Science, it includes 18 Knowledge Areas (where 6 KAs were newly introduced in ACM CS2013):

AL - Algorithms and Complexity
AR - Architecture and Organization
CN - Computational Science
DS - Discrete Structures
GV - Graphics and Visualization
HCI - Human-Computer Interaction
IAS - Information Assurance and Security (new)
IM - Information Management
IS - Intelligent Systems
NC - Networking and Communications (new)
OS - Operating Systems
PBD - Platform-based Development (new)
PD - Parallel and Distributed Computing (new)
PL - Programming Languages
SDF - Software Development Fundamentals (new)
SE - Software Engineering
SF - Systems Fundamentals (new)
SP - Social Issues and Professional Practice

Knowledge areas should not directly match a particular course in a curriculum (this practice is strongly discouraged in the ACM report), often courses address topics from multiple knowledge areas. The ACM-CS2013-final report distinguishes between two types of topics: Core topics subdivided into “Tier-1” (that are mandatory for each curriculum) and “Tier-2” (that are expected to be covered at 90-100% with minimum advised 80%), and elective topics. The ACM classification suggests that a curriculum should include all topics in Tier-1 and all or almost the topics in Tier 2. Tier 1 and Tier 2 topics are defined differently for different programmes and specialisations. To be complete, a curriculum should cover in addition to the topics of Core Tier 1 and 2 a significant amount of elective material. The reason for such a hierarchical approach to the structure of the Body of Knowledge is a useful way to group related information, not as a structure for organizing material into courses.

The ACM Curriculum for computing Education in Community Colleges [8] defines a BoK for IT outcome-based learning/education, which identifies 6 technical competency areas and 5 workplace skills[6]. While the technical areas are specific to IT competences and specify a set of demonstrable abilities of graduates to perform some specific functions, the so-called workplace skills describe the ability of the student/trainee to:

(1) function effectively as a member of a diverse team,

[6] Importance of work place skills (that also referred to as attitude skills, “soft” skills, or 21st Century skills). These skills are defined as a part of the Data Science Competence Framework, refer to CF-DS document section 4.6 [1].

(2) read and interpret technical information,
(3) engage in continuous learning,
(4) professional, legal, and ethical behavior, and
(5) demonstrate business awareness and workplace effectiveness

The ACM steering committee agrees on set principles to guide the development of CS2013 model curriculum. These principles aim at providing students with the necessary flexibility to work across disciplines and prepare the graduates for a variety of disciplines. Following is the summary of the most important principles:

(1) CS2013 should provide guidance for the expected level of mastery of topics by the graduate
(2) CS2013 should provide realistic, adoptable recommendations that provide guidance and flexibility, allowing curricula designs that are innovative and track recent developments in the field
(3) Size of the essential knowledge must be manageable
(4) Computer science curricula should prepare graduates to succeed in a rapidly changing area
(5) CS2013 should identify the fundamental skills and knowledge that all computer Science graduates should possess while providing the greatest flexibility in selecting topics
(6) CS2013 should provide great flexibility in organizing topics into courses and curricula.

Through these principles, ACM provides graduates with fundamental knowledge in the areas described in the ACM-BoK and a style of thinking and problem solving. The latter is achieved through defining the expected characteristics of computer science graduates, namely:

- Technical understanding of computer science
- Familiarity with common themes and principals
- Appreciation of interplay between theory and practice
- System-level perspective
- Problem solving skills
- Project experience
- Commitment to life-long learning
- Commitment to professional responsibility
- Communication and organization skills
- Appreciation of domain specific knowledge

ACM follows a simple straightforward approach to design the ACM Model Curriculum. It starts from the CS2013 based CS-BoK, which is structured into Knowledge areas (KA), organized in topical themes rather than by courses boundary. Each KA is further organized into a set of Knowledge Units (KU). In the final step, each KU lists a set of topics and learning outcomes (LO). The LO are associated with a level of mastery derived from the Bloom taxonomy (familiarity, usage, and assessment).

The CS-BoK uses ACM Computing Classification System (CCS2012) for defining BoK topics and academic subjects in the curriculum. Necessary extensions/KAs related to identified Data Science competence groups are provided as CCS2012 extension points (see DS-BoK [2] document and Appendix B).

3.3 ACM/IEEE-CS Curricula Guidelines and Competency Model for Information Technologies

The ACM Committee for Computing Education in Community Colleges (CCECC) and its partner professional societies (in particular, IEEE Computer Society) have jointly produced curricular recommendations and guidelines for baccalaureate computing programs, known collectively as the ACM Computing Curricula series. One of these guidelines is the Curriculum Guidelines for Undergraduate Degree Programs in Information Technology (IT2008) and its later published companion document ACM Competency Model of Core Learning Outcomes and Assessment for Associate-Degree Curriculum in Information Technology (IT2014) [9]. The guidelines use the competence-based learning model that focuses on the extent that students learn given competencies (knowledge, skills, qualifications), instead of focusing on so-called „seat time“, commonly expressed by credit points. The proposed competency model for constructing Information Technology curricula is based on defining measurable learning outcomes. The CCECC identified the Body of Knowledge as a set of fifty student learning outcomes that span the first three levels of Bloom’s Revised Taxonomy (see above), and each outcome is accompanied by a three-tier assessment rubric that provides additional clarity and a measurable evaluation metric [9].

3.4 ICT professional Body of knowledge and new curricula for e-Leadership skills

The ICT-BoK [10] is an effort promoted by the European Commission, under the eSkills initiative (http://eskills4jobs.ec.europa.eu/) to define and organise the core knowledge of the ICT discipline. In order to foster the growth of digital jobs in Europe and to improve ICT Professionalism, a study has been conducted to provide the basis of a "Framework for ICT professionalism" (http://ictprof.eu/). This framework consists of four building blocks (also called pillars) which are also found in other professions:

i) body of knowledge (BoK);
ii) competence framework;
iii) education and training resources; and
iv) code of professional ethics.

A competence framework already exists and consists in the e-Competence Framework (now in its version 3.0 and promoted by CEN). However, an ICT Body of Knowledge that provides the basis for a common understanding of the foundational knowledge an ICT professional should possess is not yet available.

The ICT-BoK is suggested to be structured in 5 *Process Groups*, defining the various phases of the project development or organisational workflow: *Initiating*, *Planning*, *Executing*, *Monitoring and Controlling*, *Closing*.

The ICT-BoK aims at informing about the level of knowledge required to enter the ICT profession and acts as the first point of reference for anyone interested in working in ICT. Even if the ICT-BoK does not refer to Data Science competences explicitly, the identified ICT processes can be applied to data management processes both in industry and academia in the context of well-defined and structured projects.

Further ICT BoK development was focused on developing the new curricula for e-leadership skills in Europe. (refer to the SCALE Project report [29] for details).

4 Data Science Model Curriculum Design Approach

This section presents the definition of the EDISON Data Science Model Curriculum that is primarily based on the mapping between DS-BoK Knowledge Areas and MC-DS Learning Units, that may represent academic courses and training modules, for required competence groups using competence bases learning model.

The proposed MC-DS can be used for defining individual curricula for specific Data Science professional profiles or customized individual curricula for practitioners that want to obtain a Data Science qualification or certification. The example of applying competence based approach to selecting a set of Learning Units for different DSP profiles is given in Chapter 6. The proposed methods can be used for developing tools for customizing or profiling the training and/or education programmes for students or individual trainees.

4.1 Linking DS-BoK Knowledge Areas and MC-DS Learning Units for target Competence Groups

In general, a Model Curriculum can be regarded as a blueprint that can be used by educators and trainers to develop curricula at various educational institutions. There are several concepts that can guide the development of a curriculum like: Alignment and Coherence, Scope, Sequence, Continuity, and Integration [30]. These 5 basic concepts help to develop a logically consistent curriculum in which components (courses, and learning units) complement each other and are ordered in such a way that it forms 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 concepts 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 (OBL). In this case, well-defined learning outcomes are specified for all academic activities or classes are specified, and students' progress is assessed against those learning outcomes.

The Model Curriculum is organized as core and elective topics, following the ACM definition [8]. Core topics are required for every Data Science program, while Elective topics aim to cover in-depth knowledge of a specific area of data science. The last step identifies the Learning outcomes associated with each core or elective topic.

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

1. For each enumerated competence from CF-DS, define Learning Outcome according to knowledge or mastery level (defined as Familiarity, Usage, Assessment for current MC-DS version)
2. Each Knowledge Area Group of DS-BoK (that includes both KAGs from existing BoKs and those defined based on the ACM Classification Computer Science CCS2012 is mapped to existing academic subject classification groups that is primarily based on ACM CS2012 complemented with the domain or technology specific classifications such as BABOK, ACM-BOK, DAMA-BOK, PM-BOK, and others to be defined by subject matter experts.
3. For each KAG or Knowledge Unit, specify related Learning Units defined according to academic subject classification or following current practices by universities
4. For each Learning Unit, assign/suggest its category as core/mandatory (Tier1 or Tier 2), elective or prerequisite
5. For both Core or Elective, define a list of Learning Outcomes

4.2 Mastery levels and Learning Outcomes

In this section, we compare mastery levels as used in the European Qualifications Framework (EQF) [30], The European e-Competence Framework (e-CFv3.0) [31], ACM/IEEE guidelines for Computer Science curriculum [8] and Bloom's taxonomy [17]. It leads to the definition of mastery levels (also called proficiency levels in e-CF) necessary to define Learning Outcomes in MC-DS. The e-CFv3.0 uses EQF for defining the proficiency level of knowledge and skills related to specific competences.

The European Qualification Framework (EQF) [30] defines eight levels of knowledge achieved through stages of education. Level 6 is considered to be achieved through a bachelor degree, level 7 through a master's degree and level 8 through a PhD degree. Levels 3-8 are mapped to 5 levels in e-CF dimension 3.

EQF descriptions provide reference both to actual levels of knowledge but also to additional skills related to knowledge application, analysis, synthesis and evaluation. It is quite similar to Bloom's approach. At the same time, levels in EQF do not only correspond to higher levels of conceptualization, but also to more specialized knowledge, experience and interpersonal skills related to people management, and professional integrity and responsibility. e-CFv3.0 adds to its description of typical tasks regarding their complexity and autonomy. Therefore, higher levels of EQF and e-CFv3.0 should not just be seen directly as the same higher levels in Bloom. At the same time, higher levels in Bloom's taxonomy are necessary to move up in e-CFv3.0 and EQF.

EQF has 8 levels, e-CFv3.0 has 5 levels and Bloom's Taxonomy has 6 levels. Designing LOs of whole programs is a balance between precision and avoiding micromanagement of further definition of courses, especially when designing a guideline for programs instead of a specific program. It might be useful to limit the number of levels on which LOs are considered. Such an approach is used in ACM/IEEE Computer Science and Information Technology curricula guidelines. Information Technology guidelines [9] define the three levels as: emerging, developed and highly developed. Computer Science guidelines [8] define the three levels as: familiarity, usage, and assessment. Bloom's taxonomy defines the six levels: knowledge, comprehension, application, analysis, synthesis and evaluation.

The three levels, as used in ACM/IEEE Computer Science guidelines, are of particular importance because significant parts of a related ACM/IEEE taxonomy and BoK is used in the definition of CF-DS and BoK-DS in EDSF. The verb usage is not fully consistent with the original Bloom's taxonomy [17] or the revised version, which is acknowledged in the document.

The comparison of the mastery levels definition used in EQF, e-CFv3.0, ACM/IEEE guidelines for Computer Science curriculum and Bloom's taxonomy is provided in Appendix A. Mastery levels.

While not required in undergraduate curricula, the holistic definition covering all EQF, e-CF levels, would also require a closer link between mastery levels and levels in Bloom's taxonomy[7]. At the same time, limitation to 3 levels should be maintained to preserve simplicity and compatibility. For the proposed MC-DS we will use the following three levels: familiarity as understood by knowledge and comprehension in Bloom's taxonomy, usage as understood by application and analysis in Bloom's taxonomy, and creation as understood by synthesis and evaluation in Bloom's taxonomy. We present the three levels again in this document for reference in Table 4.1. Details on the relation to EQF and e-CF levels can be found in Appendix A. Mastery levels. Action verbs were defined based on the original and revised Bloom's taxonomy with adjustments tailored to Data Science curricula.

Table 4.1 Knowledge levels for learning outcomes in Data Science model curricula (MC-DS)

Level	Action Verbs
Familiarity	Choose, Classify, Collect, Compare, Configure, Contrast, Define, Demonstrate, Describe, Execute, Explain, Find, Identify, Illustrate, Label, List, Match, Name, Omit, Operate, Outline, Recall, Rephrase, Show, Summarize, Tell, Translate
Usage	Apply, Analyze, Build, Construct, Develop, Examine, Experiment with, Identify, Infer, Inspect, Model, Motivate, Organize, Select, Simplify, Solve, Survey, Test for, Visualize
Assessment	Adapt, Assess, Change, Combine, Compile, Compose, Conclude, Criticize, Create, Decide, Deduct, Defend, Design, Discuss, Determine,

[7] Note, there may not be a direct mapping between mastery, proficiency or qualification levels and Bloom's Taxonomy as it is primary defined for cognitive processes during learning in the time frame of curricula or individual courses, while proficiency level related to competences is combination of knowledge and experience related to professional activity.

Disprove, Evaluate, Imagine, Improve, Influence, Invent, Judge, Justify, Optimize, Plan, Predict, Prioritize, Prove, Rate, Recommend, Solve

4.3 Learning Outcomes definition based on CF-DS

Table 4.2 presented below provides a template and examples for defining the Learning Outcomes related to enumerated CF-DS competences and different knowledge/proficiency levels defined based on Bloom's Taxonomy. The table contains the general Learning Outcomes defined after CF-DS competences that are, in most cases, split into 3 knowledge levels and use specific verbs that reflect the necessary comprehension or mastery level.

Table 4.2 Learning outcomes defined for CF-DS competences and different mastery/proficiency levels

LO ID	Data Science Competence	LO by Knowledge levels (compliant to ACM CSC 2013) and key verbs		
		Familiarity	**Usage**	**Assessment**
		Choose, Classify, Collect, Compare, Configure, Contrast, Define, Demonstrate, Describe, Execute, Explain, Find, Identify, Illustrate, Label, List, Match, Name, Omit, Operate, Outline, Recall, Rephrase, Show, Summarize, Tell, Translate	Apply, Analyze, Build, Construct, Develop, Examine, Experiment with, Identify, Infer, Inspect, Model, Motivate, Organize, Select, Simplify, Solve, Survey, Test for, Visualize	Adapt, Assess, Change, Combine, Compile, Compose, Conclude, Criticize, Create, Decide, Deduct, Defend, Design, Discuss, Determine, Disprove, Evaluate, Imagine, Improve, Influence, Invent, Judge, Justify, Optimize, Plan, Predict, Prioritize, Prove, Rate, Recommend, Solve
Data Science Data Analytics (DSDA)				
LO1-DA	**DSDA-DA** **Use appropriate data analytics and statistical techniques on available data to discover new relations and deliver insights into research problem or organizational processes and support decision-making.**	**Choose the appropriate existing analytical method and operate existing tools to do specified data analysis. Present data in the required form.**	**Develop data analysis application for specific data sets and tasks or processes. Identify necessary methods and use them in combination if necessary. Identify relations and provide consistent reports and visualizations.**	**Create a formal model for the specific organizational tasks and processes and use it to discover hidden relations, propose optimization and improvements. Develop new models and methods if necessary. Recommend and influence organizational improvement based on continuous data analysis.**
LO1.01	DSDA01 Effectively use a variety of data analytics techniques, such as Machine Learning (including supervised, unsupervised, semi-supervised learning), Data Mining, Prescriptive and Predictive Analytics, for complex data analysis through the whole data lifecycle	Choose and execute existing data analytics and predictive analytics tools.	Identify existing requirements and develop predictive analysis tools.	Design and evaluate predictive analysis tools to discover new relations.
LO1.02	DSDA02 Apply designated quantitative techniques, including statistics, time series analysis, optimization, and simulation	Choose and execute standard methods from existing statistical libraries to provide overview.	Select the most appropriate statistical techniques and model available data to deliver insights.	Assess and optimize organization processes using statistical techniques.

	to deploy appropriate models for analysis and prediction			
LO1.03	DSDA03 Identify, extract, and pull together available and pertinent heterogeneous data, including modern data sources such as social media data, open data, governmental data	Operate tools for complex data handling.	Analyze available data sources and develop tool that work with complex datasets.	Assess, adapt, and combine data sources to improve analytics
LO1.04	DSDA04 Understand and use different performance and accuracy metrics for model validation in analytics projects, hypothesis testing, and information retrieval	Name and use basic performance assessment metrics and tools.	Use multiple performance and accuracy metrics, select and use the most appropriate for a specific type of data analytics application.	Evaluate and recommend the most appropriate metrics, and propose new for new applications.
LO1.05	DSDA05 Develop required data analytics for organizational tasks, integrate data analytics and processing applications into organization workflow and business processes to enable agile decision making	Define data elements necessary to develop specified data analytics.	Develop specialized analytics to enable decision-making.	Design specialized analytics to improve decision-making.
LO1.06	DSDA06 Visualise results of data analysis, design dashboard and use storytelling methods	Choose and execute standard visualization.	Build visualizations for complex and variable data.	Create and optimize visualizations to influence executive decisions.
Data Science Engineering				
LO2-ENG	**DSENG -** Use engineering principles and modern computer technologies to research, design, implement new data analytics applications; develop experiments, processes, instruments, systems, and infrastructures to support data handling during the whole data lifecycle.	**Identify and operate instruments and applications for data collection, analysis and management**	**Model problems and develop new instruments and applications for data collection, analysis and management following established engineering principles.**	**Evaluate instruments and applications to optimize data collection, analysis and management.**
LO2.01	DSENG01 Use engineering principles (general and software) to research, design, develop and implement new instruments and applications for data collection, storage, analysis and visualisation	Choose potential technologies to develop, structure, instruments, machines, experiments, processes, and systems.	Model data analytics applications to better develop suitable instruments, machines, experiments, processes, and systems.	Create an innovative solution to research and design data analytics
LO2.02	DSENG02 Develop and apply computational and data driven solutions to domain related problems using wide range of data analytics platforms, with the special focus on Big Data technologies for large datasets and cloud based data analytics platforms	Name computational solution and identify potential data analytics platform	Apply existing computational solutions to data analytic platform.	Adapt and optimize existing computational solutions to better fit a given data analytics platform.
LO2.03	DSENG03 Develop and prototype specialised data analysis	Identify a set of potential data analytics	Survey various specialized data	Evaluate and recommend optimal data analytics tools to

	applicaions, tools and supporting infrastructures for data driven scientific, business or organisational workflow; use distributed, parallel, batch and streaming processing platforms, including online and cloud based solutions for on-demand provisioned and scalable services	tools to fit specification.	analytics tools and identify the best option.	influence decision making.
LO2.04	DSENG04 Develop, deploy and operate large scale data storage and processing solutions using different distributed and cloud based platforms for storing data (e.g. Data Lakes, Hadoop, Hbase, Cassandra, MongoDB, Accumulo, DynamoDB, others)	Find possible database solutions, including both relational and non-relational databases.	Model the problem to apply database technology.	Predict the difference in terms of performance between relational and non-relational databases and recommend a solution.
LO2.05	DSENG05 Consistently apply data security mechanisms and controls at each stage of the data processing, including data anonymisation, privacy and IPR protection.	Identify security issues related to reliable data access.	Analyze security threats and solve them using known techniques.	Evaluate security threats and recommend adequate solutions.
LO2.06	DSENG06 Design, build, operate relational and non-relational databases (SQL and NoSQL), integrate them with modern Data Warehouse solutions, ensure effective ETL (Extract, Transform, Load), OLTP, OLAP processes for large datasets	Define technical requirements for SQL/NoSQL databases, Data Warehouse technologies for data ingest.	Apply existing SQL/NoSQL databases, Data Warehouse technologies for creating data pipelines.	Combine several techniques and optimize them to design new or custom environments to integrate existing DW and database technologies for a new type of data and analytic applications.
Data Science Data Management (DSDM)				
LO3-DM	**DSDM-DM** Develop and implement data management strategy for data collection, storage, preservation, and availability for further processing.	**Execute data strategy in the form of a Data Management Plan and illustrate how available software can help to promote data quality and accessibility.**	**Develop components of data strategy and methods that improve the quality, accessibility and publications of data.**	**Create Data Management Plan aligned with the organizational needs, and evaluate IPR and ethical issues.**
LO3.01	DSDM01 - Develop and implement data strategy, in particular, in the form of Data Management Plan (DMP).	Explain and execute data strategy in the form of Data Management Plan.	Develop components of data strategy in the form of Data Management Plan.	Assess various data strategies and create strategy, in the form of Data Management Plan, aligned with organizational needs.
LO3.02	DSDM02 - Develop and implement relevant data models, including metadata.	Operate data models including metadata.	Experiment with data models and model relevant metadata.	Evaluate and design data models, including metadata.
LO3.03	DSDM03 - Collect and integrate different data sources and provide them for further analysis.	Collect different data sources.	Survey and visualize the connection between different data sources.	Compose different data sources to enable further analysis.
LO3.04	DSDM04 - Develop and maintain a historical data repository of analysis results (data provenance).	Operate a historical data repository.	Construct a historical data repository.	Improve or design a historical data repository.

LO3.05	DSDM05 - Ensure data quality, accessibility, publications (data curation).	Illustrate how available software can help to promote data quality, accessibility and publications.	Develop methods that improve the quality, accessibility and publications of data.	Improve quality, accessibility and publications of data.
LO3.06	DSDM06 - Manage IPR and ethical issues in data management.	Configure data management software to manage IPR and ethical issues.	Identify IPR and ethical issues in the data repository.	Evaluate IPR and ethical issues in the data repository.
Data Science Research Methods and Project Management (DSRMP)				
LO4-RMP	DSRM Create new understandings and capabilities by using the scientific method (hypothesis, test/artefact, evaluation) or similar engineering methods to discover new approaches to create new knowledge and achieve research or organisational goals	**Match elements of the scientific or similar method and identify appropriate actions for organizational strategy to create new capabilities.**	**Apply scientific or similar methods and develop action plans to translate organizational strategies to create new capabilities.**	**Evaluate methodologies to optimize the development of organizational objectives.**
LO4.01	DSRM01 Create new understandings by using the research methods (including hypothesis, artefact/experiment, evaluation) or similar engineering research and development methods	Match elements of a scientific or similar method to a given problem	Apply the scientific method to create new understandings and capabilities.	Evaluate various methods and predict which method can optimize the creation of new understandings and capabilities.
LO4.02	DSRM02 Direct systematic study toward understanding the observable facts, and discovers new approaches to achieve research or organisational goals	Choose observable facts from an existing study for a better understanding.	Apply systematic study toward fuller knowledge or understanding of the observable facts.	Combine several methods to discover new approaches to achieve organizational goals.
LO4.03	DSRM03 Analyse domain related research process model, identify and analyse available data to identify research questions and/or organisational objectives and formulate sound hypothesis	Formulate and test a hypothesis for a specified task or research question.	Create a full experiment to test a hypothesis for domain specific task or experiment	Analysis domain related models and propose analytics methods, suggest new data or improve the quality of used data.
LO4.04	DSRM04 Undertake creative work, making systematic use of investigation or experimentation, to discover or revise knowledge of reality, and uses this knowledge to devise new applications, contribute to the development of organizational objectives	Show creativity under the guidance of senior staff in discovering and revising knowledge.	Develop creative solutions using systematic investigation or experimentation to revise and discover knowledge.	Adapt common systematic investigation to design and plan creative work to discover or revise knowledge.
LO4.05	DSRM05 Design experiments that include data collection (passive and active) for hypothesis testing and problem solving	Illustrate outstanding ideas to solve complex problems.	Identify non-standard solutions to solve complex problems.	Recommend cost effective solution to a complex problem.
LO4.06	DSRM06	Identify appropriate actions for a given	Develop actions and action plan to translate	Recommend effective action plans to translate

	Develop and guide data driven projects, including project planning, experiment design, data collection and handling	project plan or experiment.	strategies into an actionable plan.	strategies, suggest new data to improve effectiveness.
Business Process Management				
LO5-BA	**DSDK** **Use domain knowledge (scientific or business) to develop relevant data analytics applications; adopt general Data Science methods to domain specific data types and presentations, data and process models, organisational roles and relations**	**Match elements of a mathematical framework to a given business problem and operate data support services for other organizational roles.**	**Model business problems into an abstract mathematical framework and identify critical points which influence the development of organizational objectives.**	**Evaluate various methods to predict which method can optimize solving business problems and recommend strategies that optimize the development of organizational objectives.**
LO5.01	DSBA01 Analyse information needs, assess existing data and suggest/identify new data required for specific business contexts to achieve organizational goal, including using social network and open data sources	Match elements of a mathematical framework to a given business problem.	Model an unstructured business problem into an abstract mathematical framework.	Evaluate various methods and predict which method can optimize solving business problems.
LO5.02	DSBA02 Operationalise fuzzy concepts to enable key performance indicators measurement to validate the business analysis, identify and assess potential challenges	Match data to the specification of services.	Analyze services to develop data specifications.	Assess and improve the use of data in services.
LO5.03	DSBA03 Deliver business focused analysis using appropriate BA/BI methods and tools, identify business impact from trends; make a business case as a result of organisational data analysis and identified trends	Identify appropriate actions for management and organizational decisions.	Identify critical points which influence the development of organizational objectives.	Recommend strategies that optimize the development of organizational objectives.
LO5.04	DSBA04 Analyse opportunity and suggest the use of historical data available at an organisation for organizational processes optimization	Operate data support services for other organizational roles.	Develop data support services for other organizational roles.	Optimize data support services for other organizational roles.
LO5.05	DSBA05 Analyse customer relations data to optimise/improve interacting with specific user groups or in the specific business sectors	Summarize customer data.	Survey and visualize customer data.	Recommend actions based on data analysis to improve customer relations.
LO5.05	DSBA06 Analyse multiple data sources for marketing purposes; identify effective marketing actions	Access and use external open data and social network data.	Identify data that bring value to used analytics for marketing. Use cloud based solutions.	Suggest new marketing models based on existing and external data.

4.4 Definition of MC-DS Learning Units

The MC-DS Learning Units (LU) or courses can be defined based on the Knowledge Areas Groups and Knowledge Units defined in the DS-BoK (refer to DS-BoK [2] or excerption in Appendix C of the current document). The following Section 5 provides examples defining courses or modules related to the DS-BoK Knowledge Area Group on Data Science and Analytics KAG1-DSDA and Knowledge Area Group on Data Science Engineering KAG2-DSENG[8]. The individual learning units or courses are defined in accordance with the existing classification of academic disciplines, in particular, the ACM Classification Computer Science (2012) [12] and are verified with the existing offered courses at universities.

The proposed LUs are grouped according to CCS2012 classification or DS-BoK knowledge groups/units that can be used as context information for future Data Science curricula development, modification or enhancement with the linked courses and disciplines.

The further development will intend to provide a flexible mapping between Learning Outcomes including proficiency or mastery level, competences related to professional profiles, and knowledge units; however, this will require wider involvement of subject matter experts and practitioners. This will allow constructing a customized MC-DS curriculum for individual learner groups or organizational needs.

[8] Acronyms and enumeration used here are according to CF-DS and DS-BoK that are explained in the corresponding documents. Full list of enumerated values and vocabulary currently is maintained in the Excel workbooks at https://github.com/EDISONcommunity/EDSF

5 Data Science Model Curriculum (MC-DS)

The proposed MC-DS intends to provide guidance to universities and training organisations in the construction of Data Science programmes and individual courses selection that are balanced according to the requirements elicited from the research and industry domains. MC-DS can be used for the assessment and improvement of existing Data Science programmes with respect to the knowledge areas and competence groups that are associated with specific professional profiles. When coupled with individual or group competence benchmarking, MC-DS can also be used for building individual training curricula and professional (self/up/re-) skilling for effective career management.

MC-DS follows the competence-based curriculum design approach grounded in the Data Science competences defined in CF-DS and correspondingly defined Learning Outcomes (LO). The DS-BoK provides a basis for structuring the proposed MC-DS by Knowledge Area Groups (KAG) and Knowledge Areas (KA) defined in correspondence with the CF-DS competence groups and individual competences. MC-DS design supports the design of programs and courses that make use of best educational practices, such as Constructive Alignment, Problem- and Project-based Learning, Bloom's Taxonomy.

This chapter presents a short overview of the MC-DS organization and its application to defining knowledge topics (knowledge units) and learning outcomes for two main Knowledge Area Groups: Data Science Analytics and Data Science Engineering[9]. It also provides suggestions for ECTS points specification for main professional profiles group: Data Science Professionals DSP04-DSP09 (refer to sections 5 and 6 of the DSPP document [4]).

5.1 Organization and Application of Model Curriculum

In this section, we start by describing organization of MC-DS and the relation between its elements and other elements of EDSF. Further, we explain how to use MC-DS together with other EDSF components to design a new education program in Data Science.

5.1.1 Organization of Model Curriculum

MC-DS organisation is based on the Data Science Competence Framework, Professional Profiles and Body of Knowledge. For each enumerated competence, MC-DS defines Learning Outcomes according to knowledge or mastery level (defined as Familiarity, Usage, Assessment). Each Knowledge Area Group of DS-BoK is mapped to existing academic subject classification groups that is primarily based on ACM Classification Computer Science CCS2012 [12] complemented with the domain or technology specific classifications such as defined in the existing BoK's ACM CS-BOK [15], BABOK [16], SWEBOK [17], DM-BoK [18], PM-BOK [19], and others that should to be defined by subject matter experts. For each KAG, MC-DS specifies Learning Outcomes and mastery levels following Bloom's Taxonomy verb usage. Learning Outcomes are also linked to a set of Learning Units, which are examples of the practical application of Knowledge Units. ECTS points are provided for Professional Profile groups and divided into Tier-1, Tier-2, Elective and Prerequisite categories to help create detailed tracks and specializations for academic programs and professional training.

Figure 5.1 illustrates the relationship between different EDSF components when defining specific academic or professional training programme that can be tailored for specific target Data Science professional groups.

[9] DSDA and DSENG are the most distinguishing for the Data Science curricula, and in particular for the target professional profiles DSP04-DSP09.

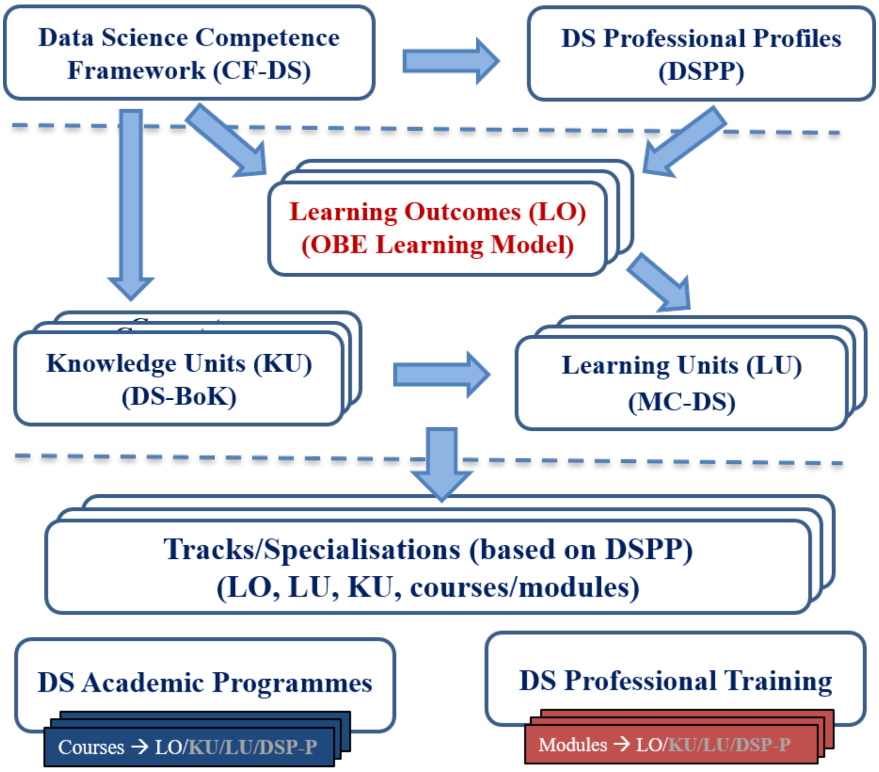

Figure 5.1. Interaction between different components of EDSF when using Model Curriculum for defining academic of a professional training programme for target professional group.

5.1.2 Application of Model Curriculum

This section describes a general approach to application of the Model Curriculum to create an educational program that is illustrated in Figure 5.2.

The work starts by deciding on a target Data Science professional profiles group the program should cover and the level of the program, usually Bachelor or Master. These elements allow to identify a set of competencies to be addressed in the program. To identify relevant Knowledge Units and to what extent they should be covered in the new program, the program designer can consult tables with ECTS points, which are defined for each Professional Profile. ECTS points specifications include a degree of flexibility to adjust to particular needs. For each Knowledge Area, MC-DS defines a set of knowledge units based on BoK and a set of learning outcomes based on the Competence Framework. Topics and learning outcomes become a base for the definition of new courses or use of existing courses. It is important to note that when designing a specific course, it may include elements from several Knowledge Areas to ensure consistency of the whole Data Science programme.

Adjustment of learning outcomes levels for different proficiency levels can be made based on the full list of learning outcomes for all CF-DS competences and for all mastery/proficiency levels provided in Table 4.2 of this document. Learning outcomes can repeat between subgroups within the same KAG, however adjusted to a specific course and topics context.

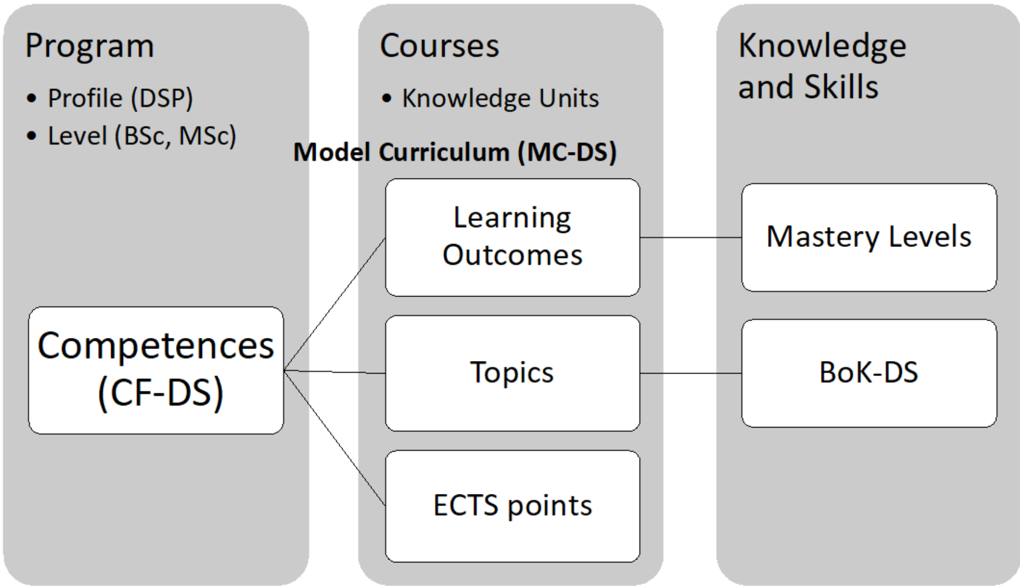

Figure 5.2. Visualization of Model Curriculum application for programs and courses.

5.2 Assignments of ECTS points to Competence Groups and Knowledge Areas

This section presents an example ECTS points specification for the main professional profile group: Data Science Professionals (DSP04-DSP09). Table 5.1 contains an example specification for a program on a Bachelor level, while Table 5.2 provides an example specification for a program on a Master level.

Points for each Knowledge Area are divided into four categories: Tier-1, Tier-2, Elective and Prerequisite. For each program 100% of Tier-1 should be covered, 80% of Tier-2 and 50% of Elective, with minor adjustments if necessary. Such system ensure that each program based on MC-DS covers basic competence and knowledge, but at the same time allowing for a necessary degree of flexibility. No prerequisites are expected for a Bachelor program, while for a Master program we set prerequisite at around 50% of combined Tier-1 and Tier-2. The goal is to ensure that students entering a program have at least the basic competences necessary to succeed in Master education, but at the same time, it allows students from a relatively wide set of backgrounds to participate. Students that do not possess the required competences, should be able to make up the difference by engaging in additional courses or bootcamps. In case, the program wants to accept a student with a different profile, e.g. pure Computer Science or pure Statistics, we recommend that the distribution of points in the program is adjusted to balance that. For instance, students with BSc in Computer Science come with a strong background in Software Development and Databases, but limited knowledge of statistics. In such a case, ECTS points should be moved between these areas.

Suggested ECTS specification for Data Analytics and Data Science Engineering Knowledge Area groups is presented here. Points for Data Management and Research methods will be presented in the MC-DS development. They should complement the ECTS points from the two groups presented here to provide 180 ECTS for Bachelor programs and 120 ECTS for Master programs. Actual curriculum design and courses selection should be done by the program coordinator at the hosting organisations to reach the total required number of credits and contain courses from Tier 1, Tier 2, and Elective.

Table 5.1. ECTS credit points for BSc program for profiles DSP04-09

Course related to DS-BoK Knowledge Areas	Tier - 1	Tier - 2	Elective	Prerequisite
DSDA/SMA (Statistical methods and data analysis)	7	4	6	NA
DSDA/ML (Machine learning)	9	8	8	NA
DSDA/DM (Data Mining)	5	4	3	NA

DSDA/TDM (Text Data Mining)	4	3	3	NA
DSDA/PA (Predictive analytics)	6	7	6	NA
DSDA/MSO (Modeling, simulation, and optimization)	5	3	4	NA
DSENG/BDI (Big Data infrastructure and technologies)	4	3	4	NA
DSENG/IPDS (Infrastructure and platforms for Data Science)	8	5	4	NA
DSENG/CCT (Cloud Computing technologies for BD and DA)	6	5	5	NA
DSENG/SEC (Data and Applications security)	2	2	2	NA
DSENG/BDSE (Big Data systems organization and engineering)	9	5	5	NA
DSENG/DSAD (Data Science/Big Data application design)	9	5	5	NA
DSENG/SE (Information Systems)	4	6	5	NA

Table 5.2. ECTS credit points for MSc program for profiles DSP04-09

Course related to DS-BoK Knowledge Areas	Tier - 1	Tier - 2	Elective	Prerequisite
DSDA/SMA (Statistical methods and data analysis)	6	2	4	6
DSDA/ML (Machine learning)	6	5	5	9
DSDA/DM (Data Mining)	4	2	4	5
DSDA/TDM (Text Data Mining)	3	2	4	4
DSDA/PA (Predictive analytics)	4	4	4	7
DSDA/MSO (Modeling, simulation, and optimization)	2	2	4	4
DSENG/BDI (Big Data infrastructure and technologies)	3	3	3	4
DSENG/IPDS (Infrastructure and platforms for Data Science)	5	3	4	7
DSENG/CCT (Cloud Computing technologies for BD and DA)	5	3	4	6
DSENG/SEC (Data and Applications security)	1	2	2	2
DSENG/BDSE (Big Data systems organization and engineering)	5	3	4	7
DSENG/DSAD (Data Science/Big Data application design)	5	3	4	7
DSENG/SE (Information Systems)	2	3	3	5

5.3 Data Science Data Analytics (KAG1 – DSDA) related courses

Data Science Analytics Knowledge Group builds the ability to use appropriate statistical and data analytics techniques on available data to deliver insights and discover information, providing recommendations, and supporting decision-making. It includes Knowledge Areas that cover: data mining, supervised and unsupervised machine learning, statistical modelling, and predictive analytics.

The following are commonly defined Data Science Analytics Knowledge Areas Group (KAG01-DSDA)[10]:

- KA01.01 (DSDA/SMDA) Statistical methods, including Descriptive statistics, exploratory data analysis (EDA) focused on discovering new features in the data, and confirmatory data analysis (CDA) dealing with validating formulated hypotheses;
- KA01.02 (DSDA/ML) Machine learning and related methods for information search, image recognition, decision support, classification;
- KA01.03 (DSDA/DM) Data mining as a particular data analysis technique that focuses on modelling and knowledge discovery for predictive rather than purely descriptive purposes;
- KA01.04 (DSDA/TDM) Text analytics applies statistical, linguistic, and structural techniques to extract and classify information from textual sources, a species of unstructured data;
- KA01.05 (DSDA/PA) Predictive analytics focuses on the application of statistical models for predictive forecasting or classification;
- KA01.06 (DSDA/MODSIM) Computational modelling, simulation and optimisation.

5.3.1 DSDA/SMDA - Statistical methods and data analysis

Statistics and probability theory are foundational components of data analytics and constitute a significant part of Data Science competences and knowledge. This module provides insight into major statistical and data

[10] All enumerated values and acronyms related to knowledge areas and units are according to the DS-BoK [2]

analytics paradigms and schools of thought. They can be taught separately or as a part of other Data Analytics related modules or courses.

Topics:
- Statistical paradigms (regression, time series, dimensionality, clusters)
- Probabilistic representations (causal networks, Bayesian analysis, Markov nets)
- Frequentist and Bayesian statistics
- Exploratory and confirmatory data analysis
- Information theory
- Graph theory

Learning Outcomes:
- Choose and execute standard methods from existing statistical libraries to provide an overview (LODA.02 L1)
- Select the most appropriate statistical techniques and model available data to deliver insights (LODA.02 L2)
- Identify requirements and develop analysis approaches (LODA.01 L2)
- Assess and optimize organization processes using statistical techniques and simulation (LODA.02 L3)

5.3.2 DSDA/ML – Machine Learning

Data Scientists have a wide range of ready machine learning libraries available. Nevertheless, they also need to go beyond the simple application of algorithms to achieve the expected results. New problems they face might require in-depth understanding of the theoretical underpinning of both simple and advanced algorithms. This module covers the use, analyze and design of machine learning algorithms.

Topics:
- Machine learning theory (supervised, unsupervised, reinforced learning, deep learning, kernel methods, Markov decision processes)
- Design and analysis of algorithms (graph algorithms, data structures design and analysis, online algorithms, bloom filters and hashing, MapReduce algorithms)
- Game theory and mechanism design
- Classification methods
- Ensemble methods
- Cross-validation

Learning Outcomes:
- Choose and execute existing analytic techniques and tools (LODA.01 L1)
- Identify requirements and develop analysis approaches (LODA.01 L2)
- Develop specialized analytics to enable agile decision-making and integrate them into organizational workflows (L0DA.05 L2)
- Design and evaluate analysis techniques and tools to discover new relations (LODA.01 L3)

DSDA/DM - Data Mining

Mathematical and theoretical aspects of data analytics must be implemented in a computational form appropriate for both problems at hand and the data size. This module builds familiarity with the most relevant data mining algorithms and related methods for knowledge representation and reasoning.

Topics:
- Data mining and knowledge discovery
- Knowledge Representation and Reasoning
- CRISP-DM and data mining stages
- Anomaly Detection
- Time series analysis
- Feature selection, Apriori algorithm
- Graph data analytics

Learning Outcomes:
- Choose and execute standard methods from statistical libraries to provide an overview (LODA.02 L1)
- Select the most appropriate statistical techniques and model available data to deliver insights (LODA.02 L2)
- Analyze available data sources and development tools that work with complex datasets (LODA.03 L2)
- Develop specialized analytics to enable agile decision-making and integrate them into organizational workflows (LODA.05 L2)
- Evaluate and recommend data analytics w.r.t. organizational strategy (LODA.05 L3)

DSDA/TDM - Text Data Mining

Text data mining can be considered a subset of data mining, but it is worth a separate consideration due to the amount of text data available and the particular methods developed over the years to analyze it.

Topics
- Text analytics including statistical, linguistic, and structural techniques to analyse structured and unstructured data
- Data mining and text analytics
- Natural Language Processing
- Predictive Models for Text
- Retrieval and Clustering of Documents
- Information Extraction
- Sentiments analysis

Learning outcomes
- Choose and execute standard methods from statistical libraries to provide an overview (LODA.02 L1)
- Analyze available data sources and development tools that work with complex datasets (LODA.03 L2)
- Evaluate and recommend data analytics w.r.t. organizational strategy (LODA.05 L3)

5.3.3 DSDA/PA - Predictive Analytics

Predictive analytics is commonly used to foresee future events in order to avoid them or act ahead. This module covers both traditional approaches based on time series and newer approaches based on deep learning. Anomaly detection is a particular focus since it is one of the most common application areas.

Topics
- Predictive modeling and analytics
- Inferential and predictive statistics
- Machine Learning for predictive analytics
- Regression and Multi Analysis
- Generalised linear models
- Time series analysis and forecasting
- Deploying and refining predictive models

Learning outcomes
- Choose and execute existing analytic techniques and tools (LODA.01 L1)
- Identify requirements and develop analysis approaches (LODA.01 L2)
- Create stories and optimize visualizations to influence executive decisions (LODA.06 L3)

5.3.4 DSDA/MODSIM - Modelling, simulation and optimization

Modeling and simulation are essential approaches to handling complexity of some systems and event chains. This module provides an introduction to both theoretical and practical aspects of model development and simulation techniques.

Topics:
- Modelling and simulation theory and techniques (general and domain oriented)

- Operations research and optimisation
- Large scale modelling and simulation systems
- Network oprtimisation
- Risk simulation and queuing

Learning Outcomes:
- Describe and execute different performance and accuracy metrics (LODA.04 L1)
- Compare and choose performance and accuracy metrics (LODA.04 L2)
- Assess and optimize organization processes using statistical techniques and simulation (LODA.02 L3)

5.4 Data Science Engineering (KAG2-DSENG)

Data Science Engineering Knowledge Group builds the ability to use engineering principles to research, design, develop and implement new instruments and applications for data collection, analysis and management. It includes Knowledge Areas that cover: software and infrastructure engineering, manipulating and analysing complex, high- volume, high- dimensionality data, structured and unstructured data, Cloud based data storage and data management.

Data Science Engineering includes software development, infrastructure operations, and algorithms design with the goal of supporting Big Data and Data Science applications in and outside the Cloud. The following are commonly defined Data Science Engineering Knowledge Areas Group (KAG02-DSENG):
- KA02.01 (DSENG/BDI) Big Data infrastructure and technologies, including NOSQL databased, platforms for Big Data deployment and technologies for large-scale storage;
- KA02.02 (DSENG/DSIAPP) Infrastructure and platforms for Data Science applications, including typical frameworks such as Spark and Hadoop, data processing models and consideration of common data inputs at scale;
- KA02.03 (DSENG/CCT) Cloud Computing technologies for Big Data and Data Analytics;
- KA02.04 (DSENG/SEC) Data and Applications security, accountability, certification, and compliance;
- KA02.05 (DSENG/BDSE) Big Data systems organization and engineering, including approach to big data analysis and common MapReduce algorithms;
- KA02.06 (DSENG/DSAPPD) Data Science (Big Data) application design, including languages for big data (Python, R), tools and models for data presentation and visualization;
- KA02.07 (DSENG/IS) Information Systems, to support data-driven decision making, with a focus on data warehouse and data centers.

5.4.1 DSENG/BDI - Big Data infrastructure and technologies

Big data infrastructures and technologies drive many Data Science applications. Systems and platforms behind big data differ significantly from traditional ones due to specific challenges of volume, velocity, and variety of data. This module addresses these aspects with a focus on underlying storage technologies and distributed architectures.

Topics:
- Big Data Cloud platforms (Azure, AWS)
- Approaches to data ingestion at scale
- Parallel and distributed computer architectures (Cloud Computing, client/server, grid)
- Large scale storage systems, SQL and NoSQL databases
- Computer networks architectures and protocols
- Storage for big data infrastructures and high-performance computing (HDFS, Ceph)

Learning Outcomes:
- Find possible data storage and processing solutions including both traditional and NOSQL databases (LOENG.06 L1)
- Survey various specialized data-driven tools and identify the best option (LOENG.03 L2)
- Evaluate the difference in performance between various distributed and Cloud-based platforms and recommend a solution (LOENG.01 L3)

5.4.2 DSENG/DSIAPP - Infrastructure and platforms for Data Science applications

Deployment of Data Science applications is usually tied to one of the most common platforms, such as Hadoop or Spark, hosted either on private or public Cloud. The application must also be tied to a whole data processing pipeline including ingestion and storage. This module covers these aspects with an additional focus on handling the most common types of data inputs at scale.

Topics:

- Big data frameworks (Hadoop, Spark, HortonWorks, others)
- Big data infrastructures (ingestion, storage, streaming, enabling analytics, Lambda Architecture)
- Data processing models (batch, streaming, parallelism)
- Large-scale data storage and management (data inputs: graph, text, image, table, time series)

Learning Outcomes:

- Define technical requirements for new distributed and Cloud-based applications for a given high-level design (LOENG.04 L1)
- Apply existing data-driven solutions to data analytic platform (LOENG.02 L2)
- Evaluate the difference in performance between various distributed and Cloud-based platforms and recommend a solution (LOENG.04 L3)

5.4.3 DSENG/CCT - Cloud Computing technologies for Big Data and Data Analytics

Cloud Computing technologies are the most common way to deploy Big Data and Data Analytics applications. This module provides an introduction to various levels of Cloud Computing services, such as IaaS or PaaS, on practical examples. It is also important to consider both private and Public Cloud.

Topics

- Cloud Computing architecture and services
- Cloud Computing engineering (design, management, operation)
- Cloud-enabled applications development (IaaS, PaaS, SaaS, autoscaling)
- Capex vs Opex consideration

Learning outcomes

- Choose potential technologies to implement new applications for data collection and storage (LOENG.01 L1)
- Model a problem to apply distributed and Cloud-based platforms (LOENG.04 L2)
- Evaluate the difference in performance between various distributed and Cloud-based platforms and recommend a solution (LOENG.04 L3)

5.4.4 DSENG/SEC - Data and Applications Security

Data Scientists should have a general understanding of data and application security aspects in order to properly plan and execute data-driven processing in the organization. This module provides an overview of the most important aspects, including sometimes omitted concepts of accountability, compliance and certification.

Topics

- Data security, accountability, protection
- Blockchain, and corresponding infrastructure
- Access control and Identity management
- Compliance and certification
- Data anonymization and privacy

Learning outcomes

- Identify security issues related to reliable data access (LOENG.05 L1)
- Analyze security threats and solve them using known techniques (LOENG.05 L2)

5.4.5 DSENG/BDSE - Big Data systems organization and engineering

Systems and platforms behind big data differ significantly from traditional ones due to specific challenges of volume, velocity, and variety of data. They require specialized approaches to data processing and algorithm engineering. This module addresses aspects both in general and based on common MapReduce algorithms.

Topics

- Big data frameworks (Hadoop, Spark, HortonWorks, others)
- Algorithms for large scale data processing
- Methods for pre-processing data implemented in MapReduce, including problems of correct data splitting in clusters
- Approaches to Big Data analysis (Functional abstraction for data processing, MapReduce, Lambda Architecture)
- Algorithms for visualization of large data sets, including subsampling with different distributions
- Big Data systems for applications domains

Learning outcomes

- Choose potential technologies to implement new applications for data collection and storage (LOENG.01 L1)
- Find possible data storage and processing solutions including both traditional and NOSQL databases (LOENG.06 L1)
- Model data-driven application following engineering principles (LOENG.01 L2)
- Adapt and optimize existing data-driven solutions to better fit to a given data analytics platform (LOENG.02 L3)

5.4.6 DSENG/DSAPPD - Data Science (Big Data) application design

Data Scientists are often tasked with developing new applications and systems. Certain languages and tools are more suitable in a data science context than others. This module covers the most common languages for data science and big data processing together with the most common tools for data presentation.

Topics:

- Languages for big data (Python, R)
- Tools and models for data presentation and visualization (Jupyter, Zeppelin)
- Software requirements and design
- Software engineering models and methods
- Software quality assurance
- Agile development methods, platforms, tools
- DevOps and continuous deployment and improvement paradigm

Learning Outcomes:

- Identify a set of potential data analytics tools to fit specification (LOENG.03 L1)
- Define technical requirements for new distributed and Cloud-based applications for a given high-level design (LOENG.06 L1)
- Model data-driven application following engineering principles (LOENG.01 L2)
- Apply existing techniques to develop new data analytics applications (LOENG.02 L2)
- Combine several techniques and optimize them to design new data analytic applications (LOENG.06 L3)

5.4.7 DSENG/IS - Information Systems

All organizations rely on some form of Information Systems to preserve knowledge and drive decision processes. This module focuses on the basics of well-established data warehouses, expert systems and decision support systems. Big data influence on such systems is also of interest, but related technical details are covered by other KAs.

Topics:

- Decision support systems
- Data warehousing and expert systems

- Enterprise information systems (data centers, intra/extra-net)
- Multimedia information systems

Learning Outcomes:
- Identify a set of potential data-driven tools to fit specification (LOENG.03 L1)
- Model the problem to apply traditional or NOSQL database technology (LOENG.06 L2)
- Evaluate and recommend optimal data-driven tools to influence decision making (LOENG.03 L3)

6 Example of using EDSF for Curricula Design and Evaluation

This section provides an example how the proposed EDISON Data Science Framework, in particular its components CF-DS, DS-BoK, MC-DS, and DSP profiles, can be used for designing a new Data Science curriculum or evaluating the existing curriculum for compliance to the selected Data Science professional profiles.

6.1 Designing a new programme

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 usually gives 30 points per semester. At American institutions, credit hours systems are used, and they are not fully uniform between institutions. Therefore, we do not provide an explicit recalculation to this system here. It can be easily done for each institution, depending on the typical semester load and its proportion to 30 ECTS points.

Table 6.1 visually illustrates an example of the required proficiency in each competence group for each professional profile. Data Science Professional profiles are described in the DSPP [4] which also provides mapping between the profiles and the competence groups described in CF-DS {1]. This creates a basis for the division of ECTS points between Learning Outcomes and related Learning Units. 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.

Table 6.1 Proficiency/mastery level needed by different Data Science Profiles for each of the Data Science competence groups

Legend:

1. Bars represent individual DSP profiles
2. color represents mastery level: familiarity –light blue; usage- blue; assessment – dark blue; white – not qualified.

The following Table 6.2 provides an example distribution of ECTS points between competence groups for Data Science professional profiles.

Table 6.2 ECTS point assignment to competence groups for professional profile groups (example)

Competence Group	DSP01-03 (Managers)		DSP04-09 (Professionals Data Science)		DSP10-13 (Professionals Data Handling/ Management)		DSP14-16 (Professionals databases)		DSP17-19 (Technician and Associate)	
	BSc	MSc	BSc	MSc	BSc	MSc	BSc	MSc	BSc	MSc
DSDA Data Analytics		30	55	35	30	20	25	15	15	
DS-ENG Data Science Engineering		20	55	35	50	30	115	75	135	
DSDK Domain Knowledge		20	55	35	80	50	25	15	15	
DSDM Data Management		30	5	5	10	10	10	10	10	
DSRM Scientific Research Methods/ DSBPM Business Process		10	10	10	10	10	5	5	5	
		120	180	120	180	120	180	120	180	

Table 6.4 presents an exemplary distribution of ECTS points between specific Learning Outcomes and related Learning Units for the Data Science Professional group DSP04-DSP09. The total amount of ECTS points for all learning outcomes in a specific competence group is based on the high level distribution in Table 6.1. Distribution to specific Learning Outcomes results from the importance of related Learning Units, which can belong to different tiers (Tier-1, Tier-2, Elective).

Details for other DSP professional groups can be found in Appendix C. Example ECTS points assignment to different Data Science Professional groups.

Table 6.4 Distribution of ECTS credit points between specific learning outcomes for profiles DSP04-09 Data Science Professionals

LO ID	Data Science Competence	ECTS credit points by Knowledge levels.					
		Familiarity		**Usage**		**Creation**	
		BSc	**MSc**	**BSc**	**MSc**	**BSc**	**MSc**
Data Science Data Analytics							
LO1-DA	**DSDA-DA - Use appropriate statistical techniques and predictive analytics on available data to deliver insights and discover new relations.**	**30**		**25**	**30**		**25**
LO1.01	DSDA01 - Use predictive analytics to analyze big data and discover new relations.	5		5	5		5
LO1.02	DSDA02 - Use appropriate statistical techniques on available data to deliver insights.	5		5	5		

LO ID	Data Science Competence	ECTS credit points by Knowledge levels.					
		Familiarity		**Usage**		**Creation**	
		BSc	**MSc**	**BSc**	**MSc**	**BSc**	**MSc**
LO1.03	DSDA03 - Develop specialized analytics to enable agile decision making.	5		5	5		5
LO1.04	DSDA04 - Research and analyze complex data sets, combine different sources and types of data to improve analysis.	5		5	5		5
LO1.05	DSDA05 - Use different data analytics platforms to process complex data.	5		5	5		5
LO1.06	DSDA06 - Visualise complex and variable data.	5			5		5
Data Science Data Management							
LO2-DM	**DSDM-DM - Develop and implement a data management strategy for data collection, storage, preservation, and availability for further processing.**			**5**			**5**
LO2.01	DSDM01 - Develop and implement data strategy, in particular, in the form of Data Management Plan (DMP).						
LO2.02	DSDM02 - Develop and implement relevant data models, including metadata.			2			2
LO2.03	DSDM03 - Collect and integrate different data sources and provide them for further analysis.			2			2
LO2.04	DSDM04 - Develop and maintain a historical data repository of analysis results (data provenance).			1			1
LO2.05	DSDM05 - Ensure data quality, accessibility, and publication (data curation).						
LO2.06	DSDM06 - Manage IPR and ethical issues in data management.						
Data Science Engineering							
LO3-ENG	**DSENG-ENG - Use engineering principles to research, design, develop and implement new instruments and applications for data collection, analysis and management.**	**25**		**30**	**25**		**30**
LO3.01	DSENG01 - Use engineering principles to research, design, prototype data analytics applications, or develop structures, instruments, machines, experiments, processes, systems.	5		10	5		10
LO3.02	DSENG02 - Develop and apply computational solutions to domain related problems using a wide range of data analytics platforms.	5		5	5		5
LO3.03	DSENG03 - Develops specialized data analysis tools to support executive decision making.	5		5	5		5
LO3.04	DSENG04 - Design, build, operate database technologies.	5		5	5		5
LO3.05	DSENG05 - Develop solutions for secure and reliable data access.						
LO3.06	DSENG06 - Prototype new data analytics applications.	5		5	5		5
Data Science Research Methods							

LO ID	Data Science Competence	ECTS credit points by Knowledge levels.					
		Familiarity		**Usage**		**Creation**	
		BSc	**MSc**	**BSc**	**MSc**	**BSc**	**MSc**
LO4-RM	**DSRM-RM - Create new understandings and capabilities by using the scientific method (hypothesis, test/artefact, evaluation) or similar engineering methods to discover new approaches to create new knowledge and achieve research or organizational goals.**	**5**		**5**	**5**		**5**
LO4.01	DSRM01 - Create new understandings and capabilities by using the scientific method (hypothesis, test, and evaluation) or similar engineering research and development methods.	2		2			2
LO4.02	DSRM02 - Direct systematic study toward a fuller knowledge or understanding of the observable facts, and discovers new approaches to achieve research or organizational goals.	2		2	2		2
LO4.03	DSRM03 - Undertakes creative work, making systematic use of investigation or experimentation, to discover or revise knowledge of reality, and uses this knowledge to devise new applications				2		
LO4.04	DSRM04 - Ability to translate strategies into action plans and follow through to completion.						1
LO4.05	DSRM05 - Contribute to and influence the development of organizational objectives.				1		
LO4.06	DSRM06 - Apply ingenuity to complex problems, develop innovative ideas	1		1			
Business Process Management							
LO5-BPM	**DSBPM-BPM - Use domain knowledge (scientific or business) to develop relevant data analytics applications, and adopt general Data Science methods to domain specific data types and presentations, data and process models, organisational roles and relations.**	**5**		**5**	**5**		**5**
LO5.01	DSBPM01 - Understand business and provide insight, translate unstructured business problems into an abstract mathematical framework.	2		2	2		2
LO5.02	DSBPM02 - Use data to improve existing services or develop new services.	2		2	1		2
LO5.03	DSBPM03 - Participate strategically and tactically in financial decisions that impact management and organizations.						
LO5.04	DSBPM04 - Provides scientific, technical, and analytic support services to other organizational roles.						
LO5.05	DSBPM05 - Analyse customer data to identify/optimise customer relations actions.	1		1	2		1
LO5.06	DSBPM06 - Analyse multiple data sources for marketing purposes.						

6.2 Assessment of existing programmes and identification of potential gaps

Another important and useful use of the presented MC-DS is the possibility to assess the existing Data Science programmes for compliance with the proposed MC-DS and make their fine tuning for target Data Science professional profiles or target training groups. Such use of the DS-MC will help to close the gap between the offered Data Science education and demand from the job market.

Initial study of existing Data Science programs at the initial stage of the EDISON project (refer to the project deliverable D2.2) which are collected in the EDISON Data Science programs inventory and the programmes developed by the EDISON Champion universities allowed us to make a few observations (the results are published in [37, 38]). The best existing programmes and those developed by the Champions are primarily covering the required competences profiles DSP04-DSP09 for Data Science Professionals and profiles DSP01-DSP03 for Data Science Managers (as defined in DSPP [4]). However, competences related to Data Management are not explicitly covered in most of the existing Computer Science based programmes which are primarily reviewed in the project[11]. DSP10-DSP13 profiles primarily dealing with data management, curation, digital archiving and digital libraries are offered by non-Computer Science departments and their experience and offerings are still to be studied by the project with the purpose to create consistent Data Science programmes covering both Computer Science based programmes and those educating digital librarians, archivists and curators. Taking into account that Data Management competences will be required for all DSP professional groups, necessary training can be offered at post-graduate stage or at workplace.

Together with the EDISON champion universities, we are trying to identify if their respective DS programmes are covering all DS competencies groups and with the right mastery level. As a result, new courses and trainings will be added to the existing programs. Another approach to close the courses gap considered by the EDISON champions is to establish a Data Science Erasmus exchange program across some of the EDISON champions to enable the DS graduates to move across the different universities to complete the missing competences.

[11] This gap has been recognized in the project and recommendations were provided to universities. Also, efforts have been taken to initiate a reference Data Management curriculum and modular course developments as a part of the Research Data Alliance initiative on Research Data Alliance Interest Group on Education and Training on Handling Research Data (IG-ETHRD) https://www.rd-alliance.org/groups/education-and-training-handling-research-data.html

7 Conclusion and further developments

The presented work on defining the MC-DS and other foundational components of the whole EDISON Data Science Framework has been done with wide consultation and engagement of different stakeholders, primarily from the universities, professional training organisations and research community, but also involving industry experts via standardisation bodies, professional communities and directly via the EDISON project network.

The Data Science Model Curriculum is a core component of the EDISON Data Science Framework that connects all components into a comprehensive tool aimed at supporting universities and professional training organizations in the development of new Data Science programmes, but also in the assessment of existing programmes w.r.t. coverage of competencies and knowledge areas associated with specific professional profiles/occupations.

The presented MC-DS intends to provide guidance and a basis for universities to define their Data Science curricula and help with individual courses selection. Together with DSP competence profiles, the MC-DS will help companies to correctly specify requirements to their staff knowledge and provide the necessary training for the career development of their staff.

The new release of Data Science Model Curriculum summarises the developments since the published Release 2 in July 2017. The initial MS-DS versions have been created based on extensive analysis of available information that includes existing best practices and guidelines on curriculum development, analysis of existing (at the time of initial research and study) Data Science programs and courses. This was also supported by the overview of academic publications on education models and approaches, as well as incorporating feedback from the Champion universities cooperating with the EDISON Project. The project partners themselves were actively involved in the curricula and Data Science courses development.

The current and ongoing list of suggested updates and extensions has been discussed at the EDSF Release 3 Design workshop that took place on 18-19 July 2018 in Amsterdam and was hosted by the University of Amsterdam.

7.1 Summary of the recent developments

The presented MC-DS is built around the DS-BoK and uses the existing classification of the academic disciplines, at this stage, mostly defined by the ACM Classification of Computer Science CCS2012.

The presented MC-DS Release 4 document has been updated based on review and contributions from individual experts, discussions at a number of workshops and conferences where the EDSF development has been presented, it also incorporated feedback from practitioners that assessed usability and practically used EDSF and MC-DS for curriculum design or review.

The Release 4 updates were mostly concerned with correcting errors and consistency between different parts of the document, including updates of some figures and tables.

The learning outcomes and their link to competences, mastery/proficiency levels related to Data Science professional profiles have been reviewed and verified.

7.2 Further developments to formalize MC-DS

It is anticipated that the presented MC-DS Release 4 will require further development and validation by experts and communities of practice. This will require community contribution and will include the following activities:

- Provide support on the Data Science Model Curriculum implementation and collect feedback from practitioners and suggestions for further improvement and extension.
- Engage with the partner and cooperating universities in the practical implementation of the MC-DS and DS-BoK, gain experience and collect feedback.

- Define specific knowledge areas related to the identified knowledge area groups by involving experts in the related knowledge areas, possibly also engaging with the specific professional communities such as IEEE, ACM, DAMA, IIBA, etc.
- Finalise the taxonomy of Data Science related knowledge areas and scientific disciplines based on ACM CCS (2012), provide suggestions for new knowledge areas and classifications classes.
- Work with the IEEE and ACM curriculum workshop to define Data Science Curriculum and extend the current CCS2012 (Classification Computer Science 2012)

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 EDSF recommendations and adapt them to available expertise, resources, and demand of the Data Science competences and skills.

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

Future support for EDSF and DS-BoK in particular will be provided in the framework of the EDISON Community via github project space https://github.com/EDISONcommunity/EDSF/wiki/EDSFhome.

8 References

[1] Data Science Competence Framework, EDSF Part 1 [online] https://github.com/EDISONcommunity/EDSF/tree/master/data-science-competence-framework
[2] Data Science Body of Knowledge, EDSF Part 2 [online] https://github.com/EDISONcommunity/EDSF/tree/master/data-science-body-of-knowledge
[3] Data Science Model Curriculum, EDSF Part 3 [online] https://github.com/EDISONcommunity/EDSF/tree/master/data-science-model-curriculum
[4] Data Science Professional Profiles, EDSF Part 4 [online] https://github.com/EDISONcommunity/EDSF/tree/master/data-science-professional-profile
[5] EDSF Use cases and applications, EDSF Part 5 [online] https://github.com/EDISONcommunity/EDSF/tree/master/data-science-edsf-use-cases-applications
[6] FAIR Competence Framework for Higher Education (Data Stewardship Professional Competence Framework), FAIRsFAIR Project Deliverable D7.3, February 2021 [online] https://zenodo.org/record/4562089#.Y6uctnbMK38
[7] The 2012 ACM Computing Classification System [online] http://www.acm.org/about/class/class/2012
[8] Computer Science 2013: Curriculum Guidelines for Undergraduate Programs in Computer Science http://www.acm.org/education/CS2013-final-report.pdf
[9] Information Technology Competency Model of Core Learning Outcomes and Assessment for Associate-Degree Curriculum (2014) http://www.capspace.org/uploads/ACMITCompetencyModel14October2014.pdf
[10] ICT professional Body of Knowledge (ICT-BoK) [online] http://www.ictbok.eu/images/EU_Foundationa_ICTBOK_final.pdf
[11] Data Management Body of Knowledge (DM-BoK) by Data Management Association International (DAMAI) [online] http://www.dama.org/sites/default/files/download/DAMA-DMBOK2-Framework-V2-20140317-FINAL.pdf
[12] Software Engineering Body of Knowledge (SWEBOK) [online] https://www.computer.org/web/swebok/v3
[13] Business Analytics Body of Knowledge (BABOK) [online] http://www.iiba.org/babok-guide.aspx
[14] Project Management Professional Body of Knowledge (PM-BoK) [online] http://www.pmi.org/PMBOK-Guide-and-Standards/pmbok-guide.aspx
[15] European Credit Transfer and Accumulation System (ECTS) [online] http://ec.europa.eu/education/ects/users-guide/docs/year-2009/ects-users-guide-2009_en.pdf
[16] Carnegie unit credit hour [online] https://www.luminafoundation.org/files/resources/carnegie-unit-report.pdf
[17] 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.
[18] Anderson, L. W., & Krathwohl, D. R. (2001). A taxonomy for learning, teaching, and assessing, Abridged Edition. Boston, MA: Allyn and Bacon.
[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] D. A. Kolb, Experiential learning: experience as the source of learning and development. Prentice-Hall, 1984.
[21] P. C. Blumenfeld, E. Soloway, R. W. Marx, J. S. Krajcik, M. Guzdial, and A. Palincsar, “Motivating Project-Based Learning: Sustaining the Doing, Supporting the Learning,” Educational Psychologist, vol. 26, no. 3–4, pp. 369–398, 1991.
[22] T. W. Malone and M. R. Lepper, “Making learning fun: A taxonomy of intrinsic motivations for learning,” Aptitude, learning, and instruction, vol. 3, pp. 223–253, 1987.
[23] J. Biggs, “Enhancing teaching through constructive alignment,” Higher education, vol. 32, no. 3, pp. 347–364, 1996.
[24] M. Ben-Ari, “Constructivism in computer science education,” Journal of Computers in Mathematics and Science Teaching, vol. 20, no. 1, pp. 45–73, 2001.
[25] The Aalborg Model for Problem Based Learning (PBL) [online] http://www.en.aau.dk/education/problem-based-learning/
[26] A. Sasha Thackaberry, A CBE Overview: The Recent History of CBE [online] http://evolllution.com/programming/applied-and-experiential-learning/a-cbe-overview-the-recent-history-of-cbe/
[27] 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
[28] Aligning and Building Curriculum: General Concepts and Design Principles [online] http://gototheexchange.ca/index.php/curriculum-overview/curriculum-models-and-design-principles
[29] e-Leadership Curricula Design and Development, eLeadership Erasmus+, 2020 [online] https://eleadspeed.eilab.eu/wp-content/uploads/2020/09/e-Leadership-Curricula-Design-and-Development-Manual.pdf
[30] European Qualifications Framework (EQF) [online] https://ec.europa.eu/ploteus/content/descriptors-page
[31] European e-Competences Framework [online] https://itprofessionalism.org/about-it-professionalism/competences/the-e-competence-framework/
[32] European Skills, Competences, Qualifications and Occupations (ESCO) [online] https://ec.europa.eu/esco/portal/home
[33] Standards and Guidelines for Quality Assurance in the European Higher Education Area [online] http://www.enqa.eu/wp-content/uploads/2015/11/ESG_2015.pdf
[34] ENQA - European Association for Quality Assurance in Higher Education [online] http://www.enqa.eu/
[35] EQANIE - the European Quality Assurance Network for Informatics Education [online] http://www.eqanie.eu/

[36] Euro-Inf Framework Standards and Accreditation Criteria [online] http://www.eqanie.eu/media/Euro-Inf%20Framework%20Standards%20and%20Accreditation%20Criteria%20V2015-10-12.pdf

[37] Tomasz Wiktorski, Yuri Demchenko, Adam Belloum and Anoosheh Shirazi, Quantitative and Qualitative Analysis of Current Data Science Programs from Perspective of Data Science Competence Groups and Framework, 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.

[38] Tomasz Wiktorski, Yuri Demchenko, Adam Belloum, Model Curricula for Data Science EDISON Data Science Framework, Proc. 4th IEEE STC CC Workshop on Curricula and Teaching Methods in Cloud Computing, Big Data, and Data Science (DTW2017), part of The 9th IEEE International Conference and Workshops on Cloud Computing Technology and Science (CloudCom2017), 11-14 Dec 2017, Hong Kong.

[39] 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.

[40] Yuri Demchenko, Adam Belloum, Cees de Laat, Charles Loomis, Tomasz Wiktorski, Erwin Spekschoor, Customisable Data Science Educational Environment: From Competences Management and Curriculum Design to Virtual Labs On-Demand, Proc. 4th IEEE STC CC Workshop on Curricula and Teaching Methods in Cloud Computing, Big Data, and Data Science (DTW2017), part of The 9th IEEE International Conference and Workshops on Cloud Computing Technology and Science (CloudCom2017), 11-14 Dec 2017, Hong Kong. (PDF)

9 Acronyms

Acronym	Explanation
ACM	Association for Computer Machinery
BABOK	Business Analysis Body of Knowledge
CCS	Classification Computer Science by ACM
CF-DS	Data Science Competence Framework
CODATA	International Council for Science: Committee on Data for Science and Technology
CRISP-DM	Cross Industry Standard Process for Data Mining
CS	Computer Science
DigComp	Digital Competences for citizens (EU report 2017)
DM-BoK	Data Management Body of Knowledge by DAMAI
DS-BoK	Data Science Body of Knowledge
EDSA	European Data Science Academy
EOEE	EDISON Online E-Learning Environment
ETM-DS	Data Science Education and Training Model
EUDAT	http://eudat.eu/what-eudat
EGI	European Grid Initiative
ELG	EDISON Liaison Group
EOSC	European Open Science Cloud
ERA	European Research Area
ESCO	European Skills, Competences, Qualifications and Occupations
EUA	European Association for Data Science
FAIR	Findable, Accessible, Interoperable, Reusable data principles in Research Data Management
FAIRsFAIR	EU funded H2020 project (EOSC cluster)
HPCS	High Performance Computing and Simulation Conference
ICT	Information and Communication Technologies
IEEE	Institute of Electrical and Electronics Engineers
IPR	Intellectual Property Rights
LERU	League of European Research Universities
LIBER	Association of European Research Libraries
MC-DS	Data Science Model Curriculum
NIST	National Institute of Standards and Technologies of USA
P21C	21st Century Skills Framework
PID	Persistent Identifier
PM-BoK	Project Management Body of Knowledge
PRACE	Partnership for Advanced Computing in Europe
RDA	Research Data Alliance
SWEBOK	Software Engineering Body of Knowledge

Appendix A. Mastery levels

This appendix provides a short overview and compares the definition of mastery levels as used in the European Qualifications Framework (EQF) [31], e-CF [30], ACM/IEEE guidelines for Computer Science curriculum [8] and Bloom's taxonomy. It is used for the definition of mastery levels (also called proficiency levels in e-CF) necessary to define Learning Outcomes in MC-DS.

The European qualification framework [31] defines eight levels of knowledge achieved through stages of education. Level 6 is considered to be achieved through a Bachelor degree, level 7 through a master's degree and level 8 through a PhD degree. Levels 3-8 are mapped to 5 levels in e-CF dimension 3. The mapping and description are presented in Table . By comparing e-CF levels directly with education requirements from EQF we can notice a certain mismatch. It is impossible to achieve a desired e-CF level by simply following an education path based on EQF. It is not enough to get a master's degree to become a Lead Professional. Rather, education requirements should be interpreted as a necessary condition but not sufficient.

Table A.1 Description of EQF and e-CF levels

EQF level	EQF level description	e-CF level	e-CF level description
8	Knowledge at the most advanced frontier, the most advanced and specialized skills and techniques to solve critical problems in research and/or innovation, demonstrating substantial authority, innovation, autonomy, scholarly or professional integrity.	e-5	**Principal** Overall accountability and responsibility; recognized inside and outside the organization for innovative solutions and for shaping the future using outstanding leading edge thinking and knowledge.
7	Highly specialized knowledge, some of which is at the forefront of knowledge in a field of work or study, as the basis for original thinking, critical awareness of knowledge issues in a field and at the interface between different fields, specialized problem-solving skills in research and/or innovation to develop new knowledge and procedures and to integrate knowledge from different fields, managing and transforming work or study contexts that are complex, unpredictable and require new strategic approaches, taking responsibility for contributing to professional knowledge and practice and/or for reviewing the strategic performance of teams.	e-4	**Lead Professional/Senior Manager** Extensive scope of responsibilities deploying specialized integration capability in complex environments; full responsibility for the strategic development of staff working in unfamiliar and unpredictable situations.
6	Advanced knowledge of a field of work or study, involving a critical understanding of theories and principles, advanced skills, demonstrating mastery and innovation in solving complex and unpredictable problems in a specialized field of work or study, management of complex technical or professional activities or projects, taking responsibility for decision-making in unpredictable work or study contexts, for continuing personal and group professional development.	e-3	**Senior Professional/Manager** Respected for innovative methods and use of initiative in specific technical or business areas; providing leadership and taking responsibility for team performances and development in unpredictable environments**.**
5	Comprehensive, specialized, factual and theoretical knowledge within a field of work or study and an awareness of the boundaries of that knowledge, expertise in a comprehensive range of cognitive and practical skills in developing creative solutions to abstract problems, management and supervision in contexts where there is unpredictable change, reviewing and developing performance of self and others.	e-2	**Professional** Operates with capability and independence in specified boundaries and may supervise others in this environment; conceptual and abstract model building using creative thinking; uses theoretical knowledge and practical skills to solve complex problems within a predictable and sometimes unpredictable context.
4	Factual and theoretical knowledge in broad contexts within a field of work or study, expertise in a range of cognitive and practical skills in generating solutions to specific problems in a field of work or study, self-management not within the guidelines of work or study contexts that are usually predictable, but are subject to change, supervising the routine work of others, taking		

EQF level	EQF level description	e-CF level	e-CF level description
	some responsibility for the evaluation and improvement of work or study activities.		
3	Knowledge of facts, principles, processes and general concepts in a field of work or study, a range of cognitive and practical skills in accomplishing tasks. Problem solving with basic methods, tools, materials and information, responsibility for completion of tasks in work or study, adapting own behaviour to circumstances in solving problems.	e-1	**Associate** Able to apply knowledge and skills to solve straightforward problems; responsible for own actions; operating in a stable environment.

EQF descriptions provide reference both to actual levels of knowledge but also to additional skills related to knowledge application, analysis, synthesis and evaluation. It is quite similar to Bloom's approach. At the same time, levels in EQF do not only correspond to higher levels of conceptualization, but also to more specialized knowledge, experience and interpersonal skills related to people management and professional integrity and responsibility. e-CF adds to its description of typical tasks regarding their complexity and autonomy. Therefore, higher levels of EQF and e-CF should not just be seen directly as the same higher levels in Bloom. At the same time, higher levels in Bloom's taxonomy are necessary to move up in e-CF and EQF. It follows the earlier argument about education requirements forming necessary but not sufficient conditions.

EQF has 8 levels, e-CF has 5 levels and Bloom's has 6 levels. Designing LOs of whole programs is a balance between precision and avoiding micromanagement of further definition of courses, especially when designing a guideline for programs instead of a specific program. It might be useful to limit the number of levels on which LOs are considered. Such an approach is used in ACM/IEEE Computer Science and Information Technology curricula guidelines. Information Technology guidelines [9] define the three levels as: emerging, developed and highly developed. Computer Science guidelines [88] define the three levels as: familiarity, usage, and assessment. Bloom's taxonomy defines the six levels: knowledge, comprehension, application, analysis, synthesis and evaluation.

The three levels as used in ACM/IEEE Computer Science guidelines, are of particular importance because significant parts of a related taxonomy and BoK is used in the definition of CF-DS and BoK-DS in EDISON. A description of these three levels is presented in [8]. The verb usage is not fully consistent with the original Bloom's taxonomy [16] or the revised version, which is acknowledged in the document.

In principle, these levels are useful, though the synthesis level of Bloom's taxonomy seems to be somewhat omitted both in the naming of levels and also in their description. Furthermore, the analysis level of Bloom's taxonomy is sometimes mixed with the evaluation level. A deeper inspection suggests that ACM/IEEE's familiarity level maps to knowledge and comprehension levels in Bloom's taxonomy. Further, the usage level in ACM/IEEE maps to the analysis level in Bloom's taxonomy; and finally, assessment level in ACM/IEEE maps to the analysis level in Bloom's taxonomy. As a result, synthesis and evaluation levels from Bloom's taxonomy are to a large extent omitted. Such omission might be acceptable for undergraduate curricula that ACM and IEEE consider in these documents.

Table A.2 ACM/IEEE CS curricula master levels

Level	Description
Familiarity	The student understands what a concept is or what it means. This level of mastery concerns a basic awareness of a concept as opposed to expecting a real facility with its application. It provides an answer to the question "What do you know about this?"
Usage	The student is able to use or apply a concept in a concrete way. Using a concept may include, for example, appropriately using a specific concept in a program, using a particular proof technique, or performing a particular analysis. It provides an answer to the question "What do you know how to do?"
Assessment	The student is able to consider a concept from multiple viewpoints and/or justify the selection of a particular approach to solve a problem. This level of mastery implies more than

	using a concept; it involves the ability to select an appropriate approach from understood alternatives. It provides an answer to the question "Why would you do that?"

While not required in undergraduate curricula, the holistic definition covering all EQF, e-CF levels, requires also full coverage of levels in Bloom's taxonomy. At the same time, the limitation to 3 levels should be maintained to preserve simplicity and compatibility. We suggest the following three levels: familiarity as understood by knowledge and comprehension in Bloom's taxonomy, usage as understood by application and analysis in Bloom's taxonomy, creation as understood by synthesis and evolution in Bloom's taxonomy. We present the three levels together with action verbs in Table . Action verbs were defined based on the original and revised Bloom's taxonomy with adjustments tailored to Data Science curricula.

Table A.3 Knowledge levels for learning outcomes in Data Science model curricula (MC-DS)

Level	Action Verbs
Familiarity	Choose, Classify, Collect, Compare, Configure, Contrast, Define, Demonstrate, Describe, Execute, Explain, Find, Identify, Illustrate, Label, List, Match, Name, Omit, Operate, Outline, Recall, Rephrase, Show, Summarize, Tell, Translate
Usage	Apply, Analyze, Build, Construct, Develop, Examine, Experiment with, Identify, Infer, Inspect, Model, Motivate, Organize, Select, Simplify, Solve, Survey, Test for, Visualize
Assessment	Adapt, Assess, Change, Combine, Compile, Compose, Conclude, Criticize, Create, Decide, Deduct, Defend, Design, Discuss, Determine, Disprove, Evaluate, Imagine, Improve, Influence, Invent, Judge, Justify, Optimize, Plan, Predict, Prioritize, Prove, Rate, Recommend, Solve

Appendix B. Subset of ACM/IEEE CCS2012 for Data Science

The presented taxonomy, although based on ACM CCS (2012) classification, can provide a basis and motivation for its extension with a new classification group related to Data Science and individual disciplines that are currently missing in the current ACM classification. This work will be a subject for future development and the results will be presented in other project deliverables.

B.1. ACM Classification Computer Science (2012) structure and Data Science related Knowledge Areas

The 2012 ACM Computing Classification System (CCS) [5] has been developed as a poly-hierarchical ontology that can be utilized in semantic web applications. It replaces the traditional 1998 version of the ACM Computing Classification System (CCS), which has served as the de facto standard classification system for the computing field for many years (also been more human readable). The ACM CCS (2012) is being integrated into the search capabilities and visual topic displays of the ACM Digital Library. It relies on a semantic vocabulary as the single source of categories and concepts that reflect the state of the art of the computing discipline and is receptive to structural change as it evolves in the future. ACM provides a tool within the visual display format to facilitate the application of 2012 CCS categories to forthcoming papers and a process to ensure that the CCS stays current and relevant.

However, at the moment, none of the Data Science, Big Data or Data Intensive Science technologies are reflected in the ACM classification. The following is an extraction of possible classification facets from ACM CCS (2012) related to Data Science what reflects the multi-subject areas nature of Data Science:

As an example, the Cloud Computing that is also a new technology and closely related to Big Data technologies, currently classified in ACM CCS (2012) into 3 groups:

Networks :: Network services :: Cloud Computing
Computer systems organization :: Architectures :: Distributed architectures :: Cloud Computing
Software and its engineering :: Software organization and properties :: Software Systems Structures :: Distributed systems organizing principles :: Cloud Computing

Taxonomy is required to consistently present information about scientific disciplines and knowledge areas related to Data Science. Taxonomy is an important component to link such components as Data Science competences and knowledge areas, Body of Knowledge, and corresponding academic disciplines. From practical point of view, taxonomy includes vocabulary of names (or keywords) and the hierarchy of their relations.

The presented here initial taxonomy of Data Science disciplines and knowledge areas is based on the 2012 ACM Computing Classification System (ACM CCS (2012)). Refer to the initial analysis of ACM CCS (2012) classification and a subset of data related disciplines in section B.1 and Table B.1. The presented in Table B.2 taxonomy includes ACM CCS (2012) subsets/subtrees that contain scientific disciplines that are related to Data Science Knowledge Area groups as defined in chapter 4 Data Science Body of Knowledge definition:

- KAG1-DSA: Data Analytics group including Machine Learning, statistical methods, and Business Analytics
- KAG2-DSE: Data Science Engineering group including Software and infrastructure engineering
- KAG3-DSDM: Data Management group including data curation, preservation and data infrastructure

Two other groups, KAG4-DSRM: Scientific or Research Methods group and KAG5-DSBP: Business process management group cannot be mapped to ACM CCS (2012), and their taxonomy is not provided in this version. It is important to notice that ACM CCS (2012) provides a top level classification entry “Applied computing” that can be used as an extension point domain related knowledge area group KAG6-DSDK (see section 4.3 Knowledge Area groups definition).

The following approach was used when constructing the proposed taxonomy:

- ACM CCS (2012) provides almost full coverage of Data Science related knowledge areas or disciplines related to KAG1, KAG2, and KAG3. The following top level classification groups are used:

- Theory of computation
- Mathematics of computing
- Computing methodologies
- Information systems
- Computer systems organization
- Software and its engineering
- Each of KAGs includes subsets from a few ACM CCS (2012) classification groups to cover theoretical, technology, engineering and technical management aspects.
- Extension points are suggested for possible future extensions of related KAGs together with their hierarchies.
- KAG3-DSDM: Data Management group is currently extended with new concepts and technologies developed by the Research Data community and documented in community best practices.

Table B.1 Data Science classification based on ACM Classification (2012)

DS-BoK Knowledge Groups *)	ACM (2012) Classification facets related to Data Science
Data Science Analytics (DSDA)	Theory of computation Design and analysis of algorithms Data structures design and analysis Theory and algorithms for application domains Machine learning theory Algorithmic game theory and mechanism design Database theory Semantics and reasoning
Data Science Analytics (DSDA)	Mathematics of computing Discrete mathematics Graph theory Probability and statistics Probabilistic representations Probabilistic inference problems Probabilistic reasoning algorithms Probabilistic algorithms Statistical paradigms Mathematical software Information theory Mathematical analysis
Data Science Analytics (DSDA)	Computing methodologies Artificial intelligence Natural language processing Knowledge representation and reasoning Search methodologies Machine learning Learning paradigms Supervised learning Unsupervised learning Reinforcement learning Multi-task learning Machine learning approaches Machine learning algorithms
Data Science Analytics (DSDA)	Information systems Information systems applications Decision support systems Data warehouses Expert systems Data analytics Online analytical processing Multimedia information systems Data mining
Data Science Analytics (DSDA) EXTENSION POINT	Theory of computation DSA Extension point: Algorithms for Big Data computation Mathematics of computing DSA Extension point: Mathematical software for Big Data computation Computing methodologies DSA Extension point: New DSA computing Information systems DSA Extension point: Big Data systems (e.g. cloud based) Information systems applications DSA Extension point: Big Data applications

DS-BoK Knowledge Groups *)	ACM (2012) Classification facets related to Data Science
	DSA Extension point: Doman specific Data applications
Data Science Data Management (DSDM)	Information systems Data management systems Database design and models Data structures Database management system engines Query languages Database administration Middleware for databases Information integration
Data Science Data Management (DSDM)	Information systems Information systems applications Digital libraries and archives Information retrieval Document representation Retrieval models and ranking Search engine architectures and scalability Specialized information retrieval
Data Science Data Management (DSDM) EXTENSION POINT	Information systems Data management systems Data types and structures description Metadata standards Persistent identifiers (PID) Data types registries
Data Science Engineering (DSE)	Computer systems organization Architectures Parallel architectures Distributed architectures
Data Science Engineering (DSENG)	Networks **) Network Architectures Network Services Cloud Computing
Data Science Engineering (DSENG)	Software and its engineering Software organization and properties Software system structures Software architectures Software system models Ultra-large-scale systems Distributed systems organizing principles Cloud computing Grid computing Abstraction, modeling and modularity Real-time systems software Software notations and tools General programming languages Software creation and management
Data Science Engineering (DSENG)	Computing methodologies Modeling and simulation Model development and analysis Simulation theory Simulation types and techniques Simulation support systems
Data Science Engineering (DSENG)	Information systems Information storage systems Information systems applications Enterprise information systems Collaborative and social computing systems and tools
Data Science Engineering (DSENG) EXTENSION POINT	Software and its engineering Software organization and properties DSE Extension point: Big Data applications design Data Analytics programming languages Information systems DSE Extension point: Big Data and cloud based systems design Information systems applications DSA Extension point: Big Data applications DSA Extension point: Doman specific Data applications
DS Domain Knowledge (DSDK) EXTENSION POINT	Applied computing Physical sciences and engineering Life and medical sciences Law, social and behavioral sciences Computer forensics Arts and humanities Computers in other domains Operations research Education

DS-BoK Knowledge Groups *)	ACM (2012) Classification facets related to Data Science
	Document management and text processing

*) All Acronyms for classification groups and DS-BoK Knowledge Area Groups are brought in accordance with CF-DS-competence groups
**) Due to the important role of the Internet and networking technologies, basic knowledge about networks is required. However, as a technology domain, Networks knowledge area group should be considered a domain specific knowledge area in the general Data Science competences and knowledge definition.

Appendix C. Example ECTS points assignment to different Data Science Professional groups

Table C.1. Distribution of ECTS credit points between specific learning outcomes for profiles DSP01-03

LO ID	Data Science Competence	ECTS credit points by Knowledge levels.					
		Familiarity		**Usage**		**Creation**	
		BSc	**MSc**	**BSc**	**MSc**	**BSc**	**MSc**
Data Science Data Analytics							
LO1-DA	**DSDA-DA - Use appropriate statistical techniques and predictive analytics on available data to deliver insights and discover new relations.**				**5**		**25**
LO1.01	DSDA01 - Use predictive analytics to analyze big data and discover new relations.						
LO1.02	DSDA02 - Use appropriate statistical techniques on available data to deliver insights.				5		10
LO1.03	DSDA03 - Develop specialized analytics to enable agile decision making.						5
LO1.04	DSDA04 - Research and analyze complex data sets, combine different sources and types of data to improve analysis.						
LO1.05	DSDA05 - Use different data analytics platforms to process complex data.						5
LO1.06	DSDA06 - Visualise complex and variable data.						5
Data Science Data Management							
LO2-DM	**DSDM-DM - Develop and implement data management strategy for data collection, storage, preservation, and availability for further processing.**				**15**		**15**
LO2.01	DSDM01 - Develop and implement data strategy, in particular, in the form of Data Management Plan (DMP).				10		10
LO2.02	DSDM02 - Develop and implement relevant data models, including metadata.						
LO2.03	DSDM03 - Collect and integrate different data sources and provide them for further analysis.						
LO2.04	DSDM04 - Develop and maintain a historical data repository of analysis results (data provenance).						
LO2.05	DSDM05 - Ensure data quality, accessibility, publication (data curation).						
LO2.06	DSDM06 - Manage IPR and ethical issues in data management.				5		5
Data Science Engineering							

LO ID	Data Science Competence	ECTS credit points by Knowledge levels.					
		Familiarity		**Usage**		**Creation**	
		BSc	**MSc**	**BSc**	**MSc**	**BSc**	**MSc**
LO3-ENG	**DSENG-ENG - Use engineering principles to research, design, develop and implement new instruments and applications for data collection, analysis and management.**				**5**		**15**
LO3.01	DSENG01 - Use engineering principles to research, design, prototype data analytics applications, or develop structures, instruments, machines, experiments, processes, systems.						
LO3.02	DSENG02 - Develop and apply computational solutions to domain related problems using a wide range of data analytics platforms.						
LO3.03	DSENG03 - Develops specialized data analysis tools to support executive decision making.						5
LO3.04	DSENG04 - Design, build, operate database technologies.				5		
LO3.05	DSENG05 - Develop solutions for secure and reliable data access.						10
LO3.06	DSENG06 - Prototype new data analytics applications.						
Data Science Research Methods							
LO4-RM	**DSRM-RM - Create new understandings and capabilities by using the scientific method (hypothesis, test/artefact, evaluation) or similar engineering methods to discover new approaches to create new knowledge and achieve research or organizational goals.**				**2**		**8**
LO4.01	DSRM01 - Create new understandings and capabilities by using the scientific method (hypothesis, test, and evaluation) or similar engineering research and development methods.						
LO4.02	DSRM02 - Direct systematic study toward a fuller knowledge or understanding of the observable facts, and discovers new approaches to achieve research or organizational goals.						
LO4.03	DSRM03 - Undertakes creative work, making systematic use of investigation or experimentation, to discover or revise knowledge of reality, and uses this knowledge to devise new applications						2
LO4.04	DSRM04 - Ability to translate strategies into action plans and follow through to completion.				2		2
LO4.05	DSRM05 - Contribute to and influence the development of organizational objectives.						2

LO ID	Data Science Competence	ECTS credit points by Knowledge levels.					
		Familiarity		**Usage**		**Creation**	
		BSc	**MSc**	**BSc**	**MSc**	**BSc**	**MSc**
LO4.06	DSRM06 - Apply ingenuity to complex problems, develop innovative ideas						2
Business Process Management							
LO5-BPM	**DSBPM-BPM - Use domain knowledge (scientific or business) to develop relevant data analytics applications, and adopt general Data Science methods to domain specific data types and presentations, data and process models, organisational roles and relations.**				**4**		**6**
LO5.01	DSBPM01 - Understand business and provide insight, translate unstructured business problems into an abstract mathematical framework.						
LO5.02	DSBPM02 - Use data to improve existing services or develop new services.						
LO5.03	DSBPM03 - Participate strategically and tactically in financial decisions that impact management and organizations.				2		2
LO5.04	DSBPM04 - Provides scientific, technical, and analytic support services to other organizational roles.				2		2
LO5.05	DSBPM05 - Analyse customer data to identify/optimise customer relations actions.						2
LO5.06	DSBPM06 - Analyse multiple data sources for marketing purposes.						

Table C.2. Distribution of ECTS credit points between specific learning outcomes for profiles DSP10-13

LO ID	Data Science Competence	ECTS credit points by Knowledge levels.					
		Familiarity		**Usage**		**Creation**	
		BSc	**MSc**	**BSc**	**MSc**	**BSc**	**MSc**
Data Science Data Analytics							
LO1-DA	**DSDA-DA - Use appropriate statistical techniques and predictive analytics on available data to deliver insights and discover new relations.**	**25**		**5**	**20**		
LO1.01	DSDA01 - Use predictive analytics to analyze big data and discover new relations.	5					
LO1.02	DSDA02 - Use appropriate statistical techniques on available data to deliver insights.	5					
LO1.03	DSDA03 - Develop specialized analytics to enable agile decision making.	5					
LO1.04	DSDA04 - Research and analyze complex data sets, combine different sources and types of data to improve analysis.	5		5	10		

LO ID	Data Science Competence	ECTS credit points by Knowledge levels.					
		Familiarity		**Usage**		**Creation**	
		BSc	**MSc**	**BSc**	**MSc**	**BSc**	**MSc**
LO1.05	DSDA05 - Use different data analytics platforms to process complex data.						
LO1.06	DSDA06 - Visualise complex and variable data.	5			10		
Data Science Data Management							
LO2-DM	**DSDM-DM - Develop and implement data management strategy for data collection, storage, preservation, and availability for further processing.**			**10**			**10**
LO2.01	DSDM01 - Develop and implement data strategy, in particular, in the form of Data Management Plan (DMP).			2			2
LO2.02	DSDM02 - Develop and implement relevant data models, including metadata.			2			2
LO2.03	DSDM03 - Collect and integrate different data sources and provide them for further analysis.			2			2
LO2.04	DSDM04 - Develop and maintain a historical data repository of analysis results (data provenance).						
LO2.05	DSDM05 - Ensure data quality, accessibility, publication (data curation).			2			2
LO2.06	DSDM06 - Manage IPR and ethical issues in data management.			2			2
Data Science Engineering							
LO3-ENG	**DSENG-ENG - Use engineering principles to research, design, develop and implement new instruments and applications for data collection, analysis and management.**	**25**		**25**	**20**		**10**
LO3.01	DSENG01 - Use engineering principles to research, design, prototype data analytics applications, or develop structures, instruments, machines, experiments, processes, systems.	5		5	5		
LO3.02	DSENG02 - Develop and apply computational solutions to domain related problems using a wide range of data analytics platforms.						
LO3.03	DSENG03 - Develops specialized data analysis tools to support executive decision making.						
LO3.04	DSENG04 - Design, build, operate database technologies.	10		10	5		
LO3.05	DSENG05 - Develop solutions for secure and reliable data access.	10		10	10		10
LO3.06	DSENG06 - Prototype new data analytics applications.						
Data Science Research Methods							

LO ID	Data Science Competence	ECTS credit points by Knowledge levels.					
		Familiarity		**Usage**		**Creation**	
		BSc	**MSc**	**BSc**	**MSc**	**BSc**	**MSc**
LO4-RM	**DSRM-RM - Create new understandings and capabilities by using the scientific method (hypothesis, test/artefact, evaluation) or similar engineering methods to discover new approaches to create new knowledge and achieve research or organizational goals.**	**10**	**10**				
LO4.01	DSRM01 - Create new understandings and capabilities by using the scientific method (hypothesis, test, and evaluation) or similar engineering research and development methods.	2	2				
LO4.02	DSRM02 - Direct systematic study toward a fuller knowledge or understanding of the observable facts, and discovers new approaches to achieve research or organizational goals.	2	2				
LO4.03	DSRM03 - Undertakes creative work, making systematic use of investigation or experimentation, to discover or revise knowledge of reality, and uses this knowledge to devise new applications	2	2				
LO4.04	DSRM04 - Ability to translate strategies into action plans and follow through to completion.	2	2				
LO4.05	DSRM05 - Contribute to and influence the development of organizational objectives.	2	2				
LO4.06	DSRM06 - Apply ingenuity to complex problems, develop innovative ideas						
Business Process Management							
LO5-BPM	**DSBPM-BPM - Use domain knowledge (scientific or business) to develop relevant data analytics applications, and adopt general Data Science methods to domain specific data types and presentations, data and process models, organisational roles and relations.**	**10**	**10**				
LO5.01	DSBPM01 - Understand business and provide insight, translate unstructured business problems into an abstract mathematical framework.	2	2				
LO5.02	DSBPM02 - Use data to improve existing services or develop new services.	2	2				
LO5.03	DSBPM03 - Participate strategically and tactically in financial decisions that impact management and organizations.						

LO ID	Data Science Competence	ECTS credit points by Knowledge levels.					
		Familiarity		**Usage**		**Creation**	
		BSc	**MSc**	**BSc**	**MSc**	**BSc**	**MSc**
LO5.04	DSBPM04 - Provides scientific, technical, and analytic support services to other organizational roles.	4	4				
LO5.05	DSBPM05 - Analyse customer data to identify/optimise customer relations actions.	2	2				
LO5.06	DSBPM06 - Analyse multiple data sources for marketing purposes.						

Table C.3. Distribution of ECTS credit points between specific learning outcomes for profiles DSP14-16

LO ID	Data Science Competence	ECTS credit points by Knowledge levels.					
		Familiarity		**Usage**		**Creation**	
		BSc	**MSc**	**BSc**	**MSc**	**BSc**	**MSc**
Data Science Data Analytics							
LO1-DA	**DSDA-DA - Use appropriate statistical techniques and predictive analytics on available data to deliver insights and discover new relations.**	**20**		**5**	**15**		
LO1.01	DSDA01 - Use predictive analytics to analyze big data and discover new relations.	5					
LO1.02	DSDA02 - Use appropriate statistical techniques on available data to deliver insights.	5					
LO1.03	DSDA03 - Develop specialized analytics to enable agile decision making.						
LO1.04	DSDA04 - Research and analyze complex data sets, combine different sources and types of data to improve analysis.	5			10		
LO1.05	DSDA05 - Use different data analytics platforms to process complex data.	5		5	5		
LO1.06	DSDA06 - Visualise complex and variable data.						
Data Science Data Management							
LO2-DM	**DSDM-DM - Develop and implement a data management strategy for data collection, storage, preservation, and availability for further processing.**			**10**			**10**
LO2.01	DSDM01 - Develop and implement data strategy, in particular, in the form of a Data Management Plan (DMP).						
LO2.02	DSDM02 - Develop and implement relevant data models, including metadata.			2			2
LO2.03	DSDM03 - Collect and integrate different data sources and provide them for further analysis.			2			2
LO2.04	DSDM04 - Develop and maintain a historical data repository of analysis results (data provenance).			4			4

LO ID	Data Science Competence	ECTS credit points by Knowledge levels.					
		Familiarity		**Usage**		**Creation**	
		BSc	**MSc**	**BSc**	**MSc**	**BSc**	**MSc**
LO2.05	DSDM05 - Ensure data quality, accessibility, publication (data curation).			2			2
LO2.06	DSDM06 - Manage IPR and ethical issues in data management.						
Data Science Engineering							
LO3-ENG	**DSENG-ENG - Use engineering principles to research, design, develop and implement new instruments and applications for data collection, analysis and management.**	**70**		**45**	**75**		
LO3.01	DSENG01 - Use engineering principles to research, design, prototype data analytics applications, or develop structures, instruments, machines, experiments, processes, systems.	10		5	10		
LO3.02	DSENG02 - Develop and apply computational solutions to domain related problems using a wide range of data analytics platforms.	10		10	10		
LO3.03	DSENG03 - Develops specialized data analysis tools to support executive decision making.	10		5	10		
LO3.04	DSENG04 - Design, build, operate database technologies.	30		10	30		
LO3.05	DSENG05 - Develop solutions for secure and reliable data access.	5		5	5		
LO3.06	DSENG06 - Prototype new data analytics applications.	5		10	10		
Data Science Research Methods							
LO4-RM	**DSRM-RM - Create new understandings and capabilities by using the scientific method (hypothesis, test/artefact, evaluation) or similar engineering methods to discover new approaches to create new knowledge and achieve research or organizational goals.**	**5**			**5**		
LO4.01	DSRM01 - Create new understandings and capabilities by using the scientific method (hypothesis, test, and evaluation) or similar engineering research and development methods.	2			2		
LO4.02	DSRM02 - Direct systematic study toward a fuller knowledge or understanding of the observable facts, and discovers new approaches to achieve research or organizational goals.	2			2		
LO4.03	DSRM03 - Undertakes creative work, making systematic use of investigation or experimentation, to discover or revise knowledge of reality, and uses this knowledge to devise new applications						

LO ID	Data Science Competence	ECTS credit points by Knowledge levels.					
		Familiarity		**Usage**		**Creation**	
		BSc	**MSc**	**BSc**	**MSc**	**BSc**	**MSc**
LO4.04	DSRM04 - Ability to translate strategies into action plans and follow through to completion.	1			1		
LO4.05	DSRM05 - Contribute to and influence the development of organizational objectives.						
LO4.06	DSRM06 - Apply ingenuity to complex problems, develop innovative ideas						
Business Process Management							
LO5-BPM	**DSBPM-BPM - Use domain knowledge (scientific or business) to develop relevant data analytics applications, and adopt general Data Science methods to domain specific data types and presentations, data and process models, organisational roles and relations.**	**5**			**5**		
LO5.01	DSBPM01 - Understand business and provide insight, translate unstructured business problems into an abstract mathematical framework.	2			2		
LO5.02	DSBPM02 - Use data to improve existing services or develop new services.	1			1		
LO5.03	DSBPM03 - Participate strategically and tactically in financial decisions that impact management and organizations.						
LO5.04	DSBPM04 - Provides scientific, technical, and analytic support services to other organizational roles.	2			2		
LO5.05	DSBPM05 - Analyse customer data to identify/optimise customer relations actions.						
LO5.06	DSBPM06 - Analyse multiple data sources for marketing purposes.						

Table C.4. Distribution of ECTS credit points between specific learning outcomes for profiles DSP17-19

LO ID	Data Science Competence	ECTS credit points by Knowledge levels.					
		Familiarity		**Usage**		**Creation**	
		BSc	**MSc**	**BSc**	**MSc**	**BSc**	**MSc**
Data Science Data Analytics							
LO1-DA	**DSDA-DA - Use appropriate statistical techniques and predictive analytics on available data to deliver insights and discover new relations.**	**15**					
LO1.01	DSDA01 - Use predictive analytics to analyze big data and discover new relations.	5					
LO1.02	DSDA02 - Use appropriate statistical techniques on available data to deliver insights.	2					
LO1.03	DSDA03 - Develop specialized analytics to enable agile decision making.						

LO ID	Data Science Competence	ECTS credit points by Knowledge levels.					
		Familiarity		**Usage**		**Creation**	
		BSc	**MSc**	**BSc**	**MSc**	**BSc**	**MSc**
LO1.04	DSDA04 - Research and analyze complex data sets, combine different sources and types of data to improve analysis.						
LO1.05	DSDA05 - Use different data analytics platforms to process complex data.	5					
LO1.06	DSDA06 - Visualise complex and variable data.	3					
Data Science Data Management							
LO2-DM	**DSDM-DM - Develop and implement a data management strategy for data collection, storage, preservation, and availability for further processing.**			**10**			
LO2.01	DSDM01 - Develop and implement data strategy, in particular, in the form of Data Management Plan (DMP).						
LO2.02	DSDM02 - Develop and implement relevant data models, including metadata.			5			
LO2.03	DSDM03 - Collect and integrate different data sources and provide them for further analysis.			5			
LO2.04	DSDM04 - Develop and maintain a historical data repository of analysis results (data provenance).						
LO2.05	DSDM05 - Ensure data quality, accessibility, publication (data curation).						
LO2.06	DSDM06 - Manage IPR and ethical issues in data management.						
Data Science Engineering							
LO3-ENG	**DSENG-ENG - Use engineering principles to research, design, develop and implement new instruments and applications for data collection, analysis and management.**	**85**		**50**			
LO3.01	DSENG01 - Use engineering principles to research, design, prototype data analytics applications, or develop structures, instruments, machines, experiments, processes, systems.	10		5			
LO3.02	DSENG02 - Develop and apply computational solutions to domain related problems using a wide range of data analytics platforms.	10		10			
LO3.03	DSENG03 - Develops specialized data analysis tools to support executive decision making.	10		5			
LO3.04	DSENG04 - Design, build, operate database technologies.	40		15			
LO3.05	DSENG05 - Develop solutions for secure and reliable data access.	5		5			
LO3.06	DSENG06 - Prototype new data analytics applications.	5		10			
Data Science Research Methods							

LO ID	Data Science Competence	ECTS credit points by Knowledge levels.					
		Familiarity		**Usage**		**Creation**	
		BSc	**MSc**	**BSc**	**MSc**	**BSc**	**MSc**
LO4-RM	**DSRM-RM - Create new understandings and capabilities by using the scientific method (hypothesis, test/artefact, evaluation) or similar engineering methods to discover new approaches to create new knowledge and achieve research or organizational goals.**	**5**					
LO4.01	DSRM01 - Create new understandings and capabilities by using the scientific method (hypothesis, test, and evaluation) or similar engineering research and development methods.	2					
LO4.02	DSRM02 - Direct systematic study toward a fuller knowledge or understanding of the observable facts, and discovers new approaches to achieve research or organizational goals.	2					
LO4.03	DSRM03 - Undertakes creative work, making systematic use of investigation or experimentation, to discover or revise knowledge of reality, and uses this knowledge to devise new applications						
LO4.04	DSRM04 - Ability to translate strategies into action plans and follow through to completion.	1					
LO4.05	DSRM05 - Contribute to and influence the development of organizational objectives.						
LO4.06	DSRM06 - Apply ingenuity to complex problems, develop innovative ideas						
Business Process Management							
LO5-BPM	**DSBPM-BPM - Use domain knowledge (scientific or business) to develop relevant data analytics applications, and adopt general Data Science methods to domain specific data types and presentations, data and process models, organisational roles and relations.**	**5**					
LO5.01	DSBPM01 - Understand business and provide insight, translate unstructured business problems into an abstract mathematical framework.	2					
LO5.02	DSBPM02 - Use data to improve existing services or develop new services.	1					
LO5.03	DSBPM03 - Participate strategically and tactically in financial decisions that impact management and organizations.						

LO ID	Data Science Competence	ECTS credit points by Knowledge levels.					
		Familiarity		**Usage**		**Creation**	
		BSc	**MSc**	**BSc**	**MSc**	**BSc**	**MSc**
LO5.04	DSBPM04 - Provides scientific, technical, and analytic support services to other organizational roles.	2					
LO5.05	DSBPM05 - Analyse customer data to identify/optimise customer relations actions.						
LO5.06	DSBPM06 - Analyse multiple data sources for marketing purposes.						