International Association of
Business Analytics Certification

Data Science Model Curriculum

EDSF MC-DS - Release 2

www.iabac.org

International Association of Business Analytics Certification

Data Science Model Curriculum (MC-DS)

EDSF MC-DS - Release 2

EDISON DATA SCIENCE FRAMEWORK
EUROPEAN COMMISION GRANT AGREEMENT NO: 675419

IABAC is a registered B.V (equivalent of UK English Private Limited) company in Netherlands.
RSIN : 859414206
Vestigingsnummer : 000041293304
Statutaire naam : IABAC B.V.
Statutaire zetel : Eindhoven, The Netherlands

.

Document Editors: Yuri Demchenko		
Contributors:		
Author Initials	Name of Author	Institution
YD	Yuri Demchenko	University of Amsterdam
AB	Adam Belloum	University of Amsterdam
TW	Tomasz Wiktorski	University of Stavanger

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 lacks a commonly accepted, harmonized instructional model that reflects by design the whole lifecycle of data-handling in modern, data driven research and the digital economy.

The presented Data Science Model Curriculum is a part of the 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), and Data Science Professional Profiles definition (DSPP).

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 a form of Tier 1, Tier 2, or Elective courses. Further MC-DS refinement will be done based on consultation with the universities 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 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 academic community and professional training community. The proposed version is intended to initiate community discussion and solicit contribution from the subject matter expects and practitioners.

TABLE OF CONTENTS

1 Introduction

Data Science is an emerging field of science, which requires a multi-disciplinary approach and should be built with the strong link to Big Data and data driven technologies that created 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 present time most of the existing university curricula and training programs are built based on available courses and cover limited set of academic subjects related to a full Data Science Body of Knowledge covering only 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 in the real working environment (both in industry and academia).

The presented in this document the Data Science Model Curriculum is the part of the EDISON Data Science Framework that includes 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), and Data Science Professional Profiles (DSPP) definition.

The proposed Data Science Model Curriculum reuses the best practices in curriculum design and new educational model to facilitate the students learning as well as existing staff professional training and re-skilling for data related technologies. Building on insights gathered through thorough analyses of existing Data Science programmes (performed in the EDISON project, see Deliverable D2.2) 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 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 matching the supply-side and demand-side requirements for Data Science education. The formal definition of the Data Science Model Curriculum will create a basis for 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 academic community and professional training community. The proposed MC-DS version will facilitate Data Science curriculum harmonisation and contribution from the subject matter expects and practitioners. The MC-DS has been presented to the EDISON Liaison Groups of experts and will undergo further community discussion via EDISON community forum and by presentation 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 competence based learning model. 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 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 example how the proposed MC-DS can be used in practice for Data Science programmes and courses assessment. Section 7 provides 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 enabling the definition of the other components related to Data Science education, training, organisational roles definition and skills management, as well as professional certification.

Figure 1 below illustrates the main components of the EDISON Data Science Framework (EDSF) and their inter-relations that provides conceptual basis for the development of the Data Science profession:

- CF-DS – Data Science Competence Framework [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 basis for other components of the Data Science professional ecosystem 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

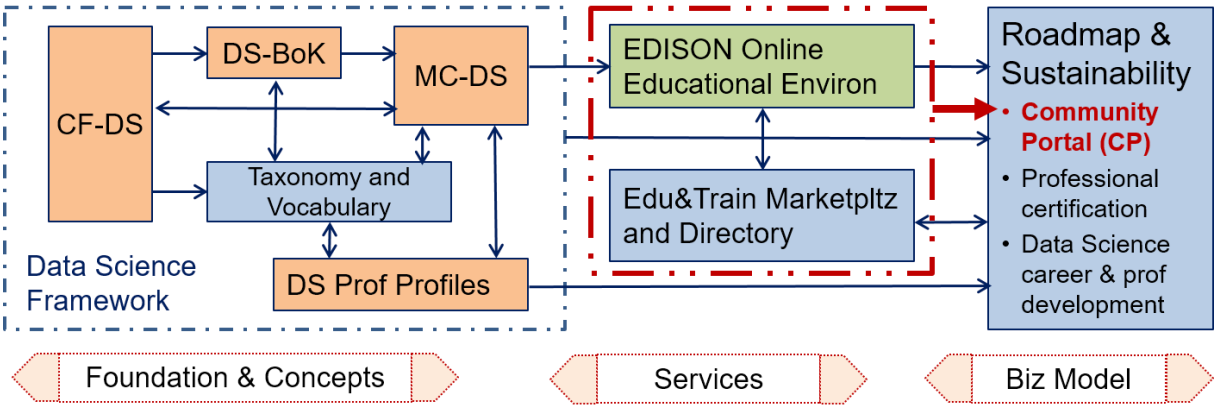


Figure 1 EDISON Data Science Framework components.

The CF-DS provides the overall basis for the whole framework, it first version has been published in November 2015 and was used as a foundation all following EDSF components developments. The CF-DS has been widely discussed at the numerous workshops, conferences and meetings, organised by the EDISON project and where the project partners contributed. The core CF-DS competences has been reviewed

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

The DS-BoK defines the Knowledge Areas (KA) for building Data Science curricula that are required to support identified Data Science competences. DS-BoK is organised by Knowledge Area Groups (KAG) that correspond to the CF-DS competence groups. DS-BoK follows the same approach to collect community feedback and contribution: Open Access CC-BY community discussion document is published on the project website. DS-BoK incorporates best practices in Computer Science and domain specific BoK's and includes KAs defined based on the Classification Computer Science (CCS2012), components taken from other BoKs and proposed new KA to incorporate new technologies used in Data Science and their recent developments.

The MC-DS is built based on CF-DS and DS-BoK where Learning Outcomes are defined based on CF-DS competences and Learning Units are mapped to Knowledge Units in DS-BoK. Three mastery (or proficiency) levels are defined for each Learning Outcome to allow for flexible curricula development and profiling for different Data Science professional profiles. The proposed Learning outcomes are enumerated to have direct mapping to the enumerated competences in CF-DS. The preliminary version of MC-DS has been discussed at the first EDISON Champions Conference in June 2016 and collected feedback is incorporated in current version of MC-DS.

The DSPP are defined as an extension to European Skills, Competences, Qualifications and Occupations (ESCO) 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 be also used for building individual career path 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 subjects in CF-DS, knowledge groups, areas and units in DS-BoK, 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 ongoing EDSF dissemination and sustainability activity.

The EDISON Data Science professional ecosystem illustrated in Figure 1 uses core EDSF components to specify the potential services that can be offered for professional Data Science community and provide basis for the sustainable Data Science and related general data skills sustainability. In particular, CF-DS and DS-BoK can be used for individual competences and knowledge benchmarking and play instrumental role in constructing personalised learning paths and professional (up/re-) skilling programs based on MC-DS.

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 Body 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 [5], their components are used in the DS-BoK presented in section 4:

- 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 will require combination of different BoKs and different approaches to curriculum definition, different subject domains and learning models. The final curriculum definition will depend on local conditions defined by demand side, available teaching staff and expertise, and available educational base and infrastructure.

3.1 Learning models and curriculum design approaches

To define consistently the MC-DS, 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, also consider alignment with the 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 a given topics or concept 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 learning (OBE), it is focusing 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 has called for a rethinking of education towards 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, 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 term 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 with particular level of learning. **Error! Reference source not found.** Illustrates (see **Figure 2**). For instance, students start at the *knowledge* level when they can *name* and *identify* relevant technologies. The further move to *comprehension* level when they can *explain* how technologies work. They can then move to *application* level when they can *choose* right technology to *solve* a problem. Further they can progress to *analysis*, *synthesis*, and finally *evaluation* levels.

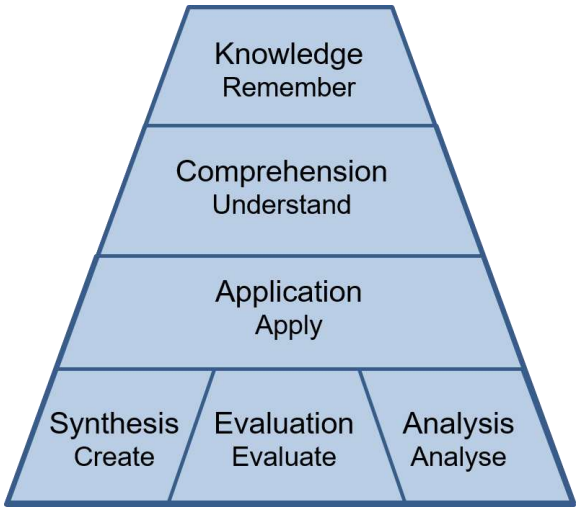


Figure 2 Simple Bloom's taxonomy: Learning levels and action verbs.

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?

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 how these services can be implemented with the selected data analytics platform, including on-premises or outsourced to 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 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 clouds and which will remain at the enterprise premises and run by company's personnel?

Evaluation
Present and defend opinions by making judgments about information, 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 Agile Data Driven Enterprise model creates benefits for enterprises, short term and long term?

Figure 3 provides consolidated presentation of the 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 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.

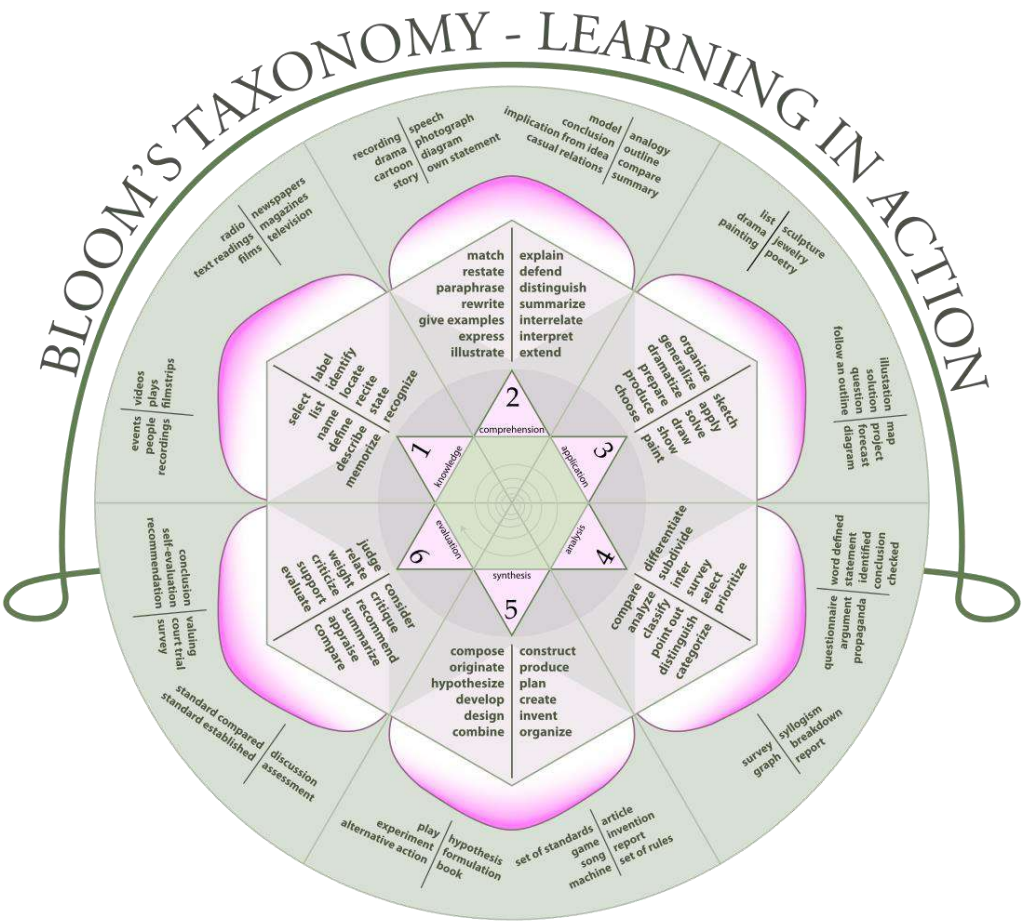


Figure 3 Extended Bloom's taxonomy[1]: consolidated presentation of learning levels, action verbs, and associated learning instruments

[1] 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. Student's 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 a 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 approach on an institutional scale is 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 existing programmes evaluation.

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 require competences through the assessment system or formal testing.

The CBL can also allow the students to learn in their own pace, practicing necessary skills as much as they need to achieve necessary mastery level. It works naturally with both individual self-study and with teacher or instructor supervised/facilitated study, so well suited for online and remote education, and in particular for post-graduate education. CBL is also associated with such educational technologies and models as MOOCs, flipped classrooms, learning analytics, and others targeting growing needs of life-long learning and self-re-skilling dictated by current fast technologies development. 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 universities CBL/CBE model, its practical implementation may create problems in some universities. Paper [27] by 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 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 needs of non-traditional students who cannot devoted their full time to traditional academic study as well as 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 are 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 distinguish between two type 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 work-place skills. 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 work-place skills describe the ability the student/trainee to:

(1) function effectively as a member of a diverse team,
(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 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 rapid changing area
(5) CS2013 should identify the fundamental skills and knowledge that all computer Science graduate should possess while providing the greatest flexibility in selecting topics
(6) CS2013 should provide a great flexibility in organizing topics into courses and curricula.

Through these principles ACM provides graduate 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 graduate 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 follow a simple straight forward 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 subject. Necessary extensions/KAs related to identified Data Science competence groups are provided as CCS2012 extension points (see 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 measureable 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 defines and organises 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 was focused on developing the new curricula for e leadership skills in Europe. (refer to the original report [3] 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 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 which components (courses, and learning units) complement each other and are ordered in such a way that it form a continuous, logical, and progressive learning path. There are several common frameworks used to develop model curricula some are subject or discipline centric while others are organized around concept and skills that are revised as we progress across the curriculum. In practice, model curricula should define either the time students have to spend learning given topics (usually using credit units) or the outcome assessing whether students have mastered the given competences (knowledge, abilities and skills). The latter approach is known as Competence-Based Education (CBE) or Outcomes-Based Learning (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 are organized as core and elective topics, following the ACM definition [8]. Core topics are required to every Data Science program while Elective topics aim to cover in depth the knowledge on a specific area of data science. The last step identifies the Learning outcomes associated to 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) [28], The European e-Competence Framework (e-CFv3.0) [29], 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) [28] 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 amount 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 EDISON. The verb usage is not fully consistent with the original Bloom's taxonomy [17] or 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, requires also full coverage of levels in Bloom's taxonomy. 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, creation as understood by synthesis and evaluation in Bloom's taxonomy. We present the three levels again in this document for reference in **Table 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 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, 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 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 necessary comprehension or mastery level.

Table 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 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 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 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 to deploy appropriate models for analysis and prediction	Choose and execute standard methods from existing statistical libraries to provide overview.	Select most appropriate statistical techniques and model available data to deliver insights.	Assess and optimize organization processes using statistical techniques.
LO1.03	DSDA03 Identify, extract, and pull together available and pertinent heterogeneous	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

	data, including modern data sources such as social media data, open data, governmental data			
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 most appropriate for specific type of data analytics application.	Evaluate and recommend the most appropriate metrics, 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, 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, instrument, machines, experiments, processes, and systems.	Model data analytics application to better develop suitable instruments, machines, experiments, processes, and systems.	Create 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 to a given data analytics platform.
LO2.03	DSENG03 Develop and prototype specialised data analysis applicaions, tools and supporting infrastructures for data driven scientific, business or organisational workflow; use distributed, parallel, batch and streaming processing platforms, including online	Identify a set of potential data analytics tools to fit specification.	Survey various specialized data analytics tools and identify the best option.	Evaluate and recommend optimal data analytics tools to influence decision making.

	and cloud based solutions for on-demand provisioned and scalable services			
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 term 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 the 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 environment to integrate existing DW and database technologies for 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 a form of 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 quality, accessibility and publications of data.**	**Create Data Management Plan aligned with the organizational needs, evaluate IPR and ethical issues.**
LO3.01	DSDM01 - Develop and implement data strategy, in particular, in a form of Data Management Plan (DMP).	Explain and execute data strategy in a form of Data Management Plan.	Develop components of data strategy in a form of Data Management Plan.	Assess various data strategies and create strategy, in a 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 source and provide them for further analysis.	Collect different data sources.	Survey and visualize 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 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 data repository.	Evaluate IPR and ethical issues in 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 scientific or similar method and identify appropriate actions for organizational strategy to create new capabilities.**	**Apply scientific or similar method 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 scientific or similar method to a given problem	Apply scientific method to create a new understandings and capabilities.	Evaluate various methods and predict which method can optimize creation of new understandings and capabilities.
LO4.02	DSRM02 Direct systematic study toward understanding of 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 a 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 hypothesis for specified task or research question.	Create full experiment to test hypothesis for domain specific task or experiment	Analysis domain related models and propose analytics methods, suggest new data or improve 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 guidance of a 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 which 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 Develop and guide data driven projects, including project planning, experiment design, data collection and handling	Identify appropriate actions for a given project plan or experiment.	Develop actions and action plan to translate strategies into actionable plan.	Recommend effective action plans to translate 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 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 context 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 specification of services.	Analyze services to develop data specification.	Assess and improve 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 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 development of organizational objectives.	Recommend strategies that optimize the development of organizational objectives.
LO5.04	DSBA04 Analyse opportunity and suggest use of historical data available at 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 the 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 example defining courses or modules related to KAG1-DSDA and KAG2-DSENG. The individual units or courses are defined in accordance with the existing classification of academic disciplines, in particular, the ACM Classification Computer Science (2012) [12] and in 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 a context information for future Data Science curricula development, modification or enhancement with the linked courses and disciplines.

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

The fully defined MC-DS will be linked to other components of the EDISON Data Science Framework such as educational materials inventory, certification scheme and services, and EDISON Online Educational Environment (EOEE).

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 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 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. It also provides suggestions for ECTS points specification for main professional profiles group: Data Science Professionals DSP04-DSP09 (refer to section 6 or DSPP document [4]). Full MC-DS version is presented in the MC-DS document [3] and can be found on EDSF website. It contains MC-DS definitions for all Knowledge Area Groups, extended Learning Outcomes inventory, and ECTS points specification for all professional profile groups.

5.1 Organization and Application of Model Curriculum

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

5.1.1 Organization of Model Curriculum

MC-DS organisation is based on Data Science Competence Framework, Professional Profiles and Body of Knowledge. For each enumerated competence, MC-DS defines Learning Outcome 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 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 relation between different EDSF components when defining specific academic or professional training programme that can be tailored for specific target Data Science professional group.

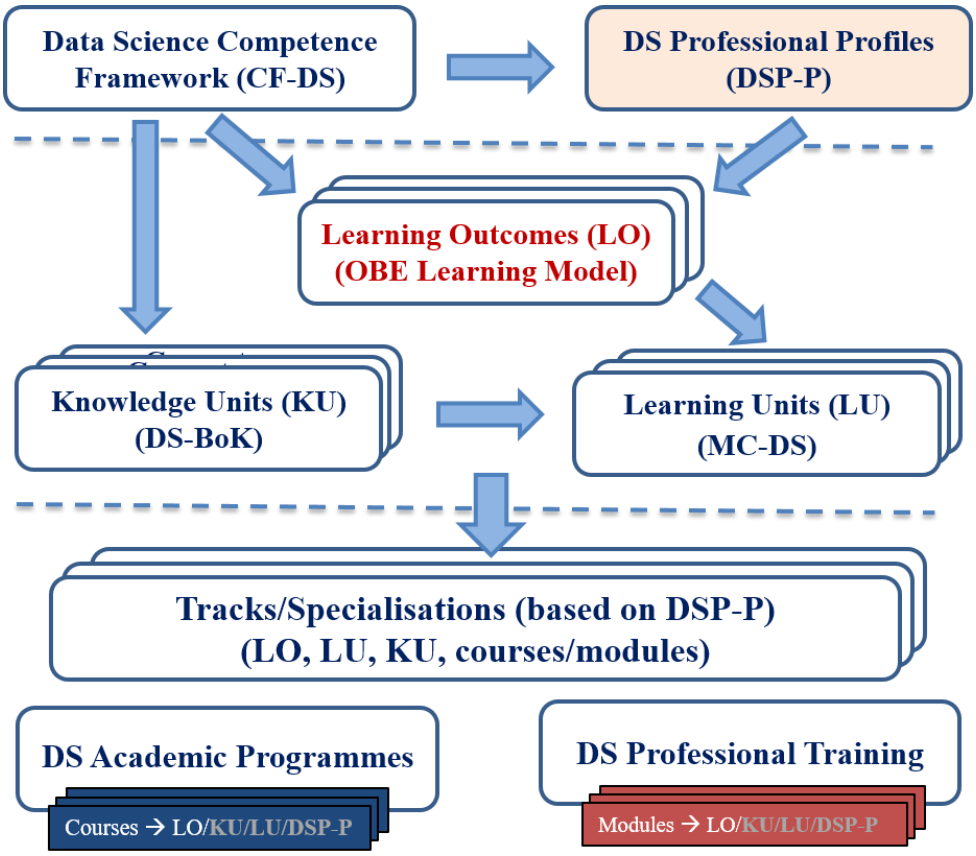


Figure 5.1. Interaction between different components of EDSF when using Model Curriculum for defining academic of professional training programme for target professional group.

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 address 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 point, which are defined for each Professional Profile. ECTS points specifications include a degree of flexibility to adjust to the particular needs. For each Knowledge Area, MC-DS defines a set of topics based on BoK and a set of learning outcomes based on Competence Framework. Topics and learning outcomes become a base for 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 done based on the full MC-DS definition in [3] that defines learning outcomes for all CF-DS competences and for all mastery/proficiency levels. Learning outcomes can repeat between subgroups within the same KAG, however adjusted to a specific course and topics context.

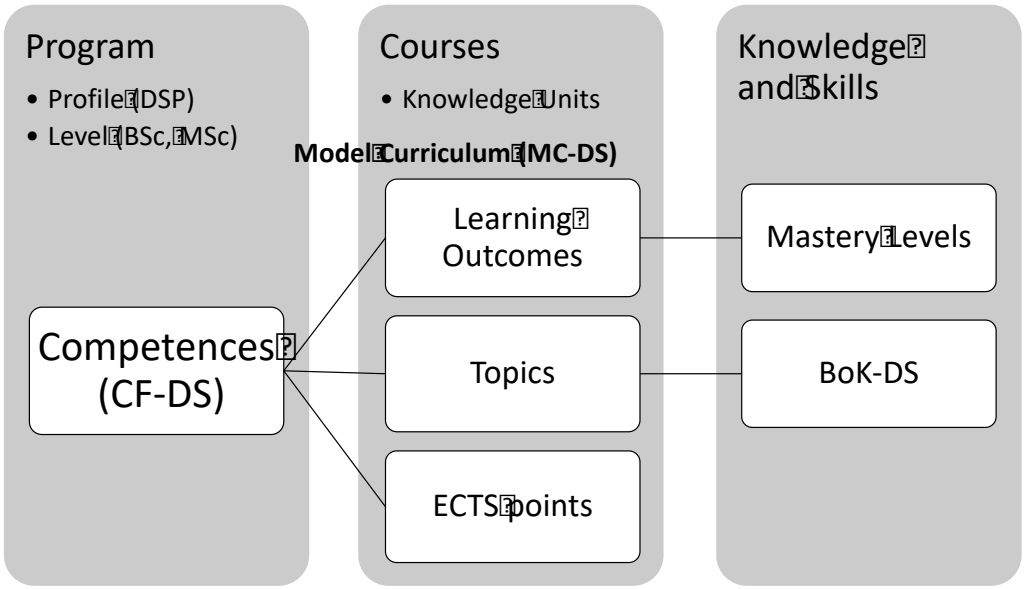


Figure 5.2. Visualization of Model Curriculum application for programs and courses.

5.2 Assignments of ECTS points to Competence Groups and Knowledge Areas

This section presents an example ECTS points specification for main professional profile group: Data Science Professionals. Table 5.1 contains example specification for a program on a Bachelor level, while Table 5.2 provides example specification for a program on 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 basic competence necessary to succeed in Master education, but at the same time it allows students from 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, program wants to accept student with a different profile, e.g. pure Computer Science or pure Statistics, we recommend that 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.

ECTS specification for Data Analytics and Data Science Engineering Knowledge Area groups is presented here. Points for Data Management and Research methods can be found in full specification of MC. They complement ECTS points from two groups presented here to provide 180 ECTS for Bachelor programs and 120 ECTS for Master programs.

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:

- 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 is 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 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 a Data Science competences and knowledge. This module provides an insight into major statistical and data 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 overview (LODA.02 L1)
- Select 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 simple application of algorithms to achieve expected results. New problems they face might require in depth understanding of 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)

5.3.3 DSDA/DM - Data Mining

Mathematical and theoretical aspects of data analytics must be implemented in a computational form appropriate for both problem at hand and data size. This module builds familiarity with 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 overview (LODA.02 L1)
- Select most appropriate statistical techniques and model available data to deliver insights (LODA.02 L2)
- Analyze available data sources and develop tool 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)

5.3.4 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 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 overview (LODA.02 L1)
- Analyze available data sources and develop tool that work with complex datasets (LODA.03 L2)
- Evaluate and recommend data analytics w.r.t. organizational strategy (LODA.05 L3)

5.3.5 DSDA/PA - Predictive Analytics

Predictive analytics are a 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 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)

5.3.6 DSDA/MODSIM - Modelling, simulation and optimization

Modeling and simulation are essential approaches to handle complexity of some systems and event chains. This module provides an introduction in 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 to support Big Data and Data Science applications in and outside the Cloud. The following are commonly defined Data Science Engineering Knowledge Areas:

- 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 approached 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 focus on data warehouse and data centers.

5.4.1 DSENG/BDI - Big Data infrastructure and technologies

Big data infrastructures and technologies drive many of the 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 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 distribute 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 most common platforms, such as Hadoop or Spark, hosted either on private or public Cloud. The application must be also tied to a whole data processing pipeline including ingestion and storage. This module covers these aspects with additional focus on handling 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 application 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 distribute 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 a 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 distribute 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 sometime 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 spliting 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 scientific context than other. This module covers most common languages for data science and big data processing together with 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 application 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 relay on some form of Information Systems to preserve knowledge and drive decision processes. This module focuses on basics of well-established data warehouse, 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

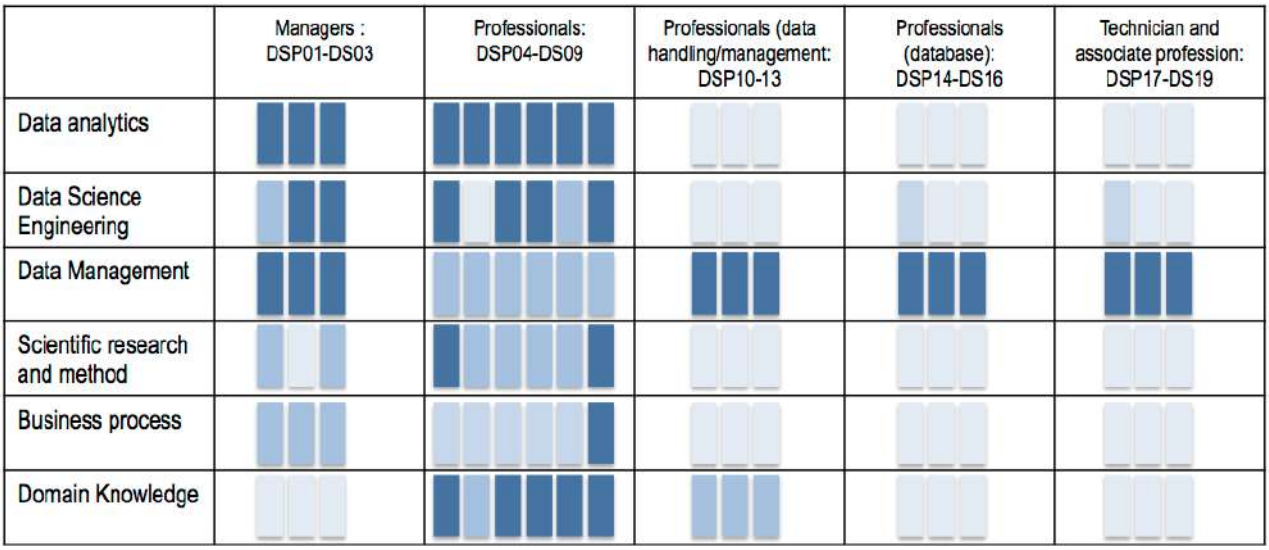
This section provides an example how the proposed EDISON Data Science Framework, in particular its components CF-DS, DS-BoK, MC-DS, and DSP profiles, can be used for designing a new Data Science curriculum or evaluating the existing curriculum for compliance to the selected Data Science professional 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 gives usually 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 institutions depending on the typical semester load and its proportion to 30 ECTS points.

Required proficiency in each competence group for each professional profile is summarized in **Table 3**. Data Science Professional profiles are described in deliverable D2.2 and competence groups in deliverable D2.1. It creates a basis for division of points between Learning Outcomes and related Learning Unit. 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 3 Proficiency/mastery level needed by different Data Science Profile for each of Data Science competence groups



Legend:

1. Bars represent individual DSP profiles
2. color represent mastery level: familiarity –light blue; usage- blue; assessment – dark blue.

The following Table 4 provides example distribution of ECTS point between competence groups for Data Science professional profiles.

Table 4 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 5 presents an exemplary distribution of ECTS points between specific Learning Outcomes and related Learning Units for 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 levels distribution in Table 4. 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 D. Example ECTS points assignment to different Data Science Professional groups.

Table 5 Distribution of ECTS credit points between specific learning outcomes for profiles DSP04-09

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		
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

LO ID	Data Science Competence	ECTS credit points by Knowledge levels.					
		Familiarity		**Usage**		**Creation**	
		BSc	**MSc**	**BSc**	**MSc**	**BSc**	**MSc**
LO1.06	DSDA06 - Visualise complex and variable data.	5			5		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.**			**5**			**5**
LO2.01	DSDM01 - Develop and implement data strategy, in particular, in a 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 source 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, publications (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 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							
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**

LO ID	Data Science Competence	ECTS credit points by Knowledge levels.					
		Familiarity		**Usage**		**Creation**	
		BSc	**MSc**	**BSc**	**MSc**	**BSc**	**MSc**
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 that are defined based on the ESCO Taxonomy [32] (see DSP profiles definition in D2.2 [3]

and discussion document [7]). Such use of the DS-MC will help to close the gap between the offered Data Science education and demand from the job market.

Preliminary study of a few existing Data Science programs from the EDISON Data Science programs inventory list [50] and programmes developed by the EDISON Champion universities allowed us to make few observations. The best existing programmes and those developed by the Champions are primarily covering the required competences profiles DSP04-DSP09 for Data Science Professional and profiles DSP01-DSP03 for Data Science Managers (see [3] and [7] for DSP taxonomy and hierarchy). However competences related to Data Management are not explicitly covered in most of existing Computer Science based programmes which are primarily reviewed in the project[2]. 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 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 working place.

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 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.

[2] This gap is recognized in the project and efforts has been taken to initiate a reference Data Management curriculum and modular course developments at the University of Amsterdam under umbrella of the Research Data Alliance initiative on Research Data Management Literacy initiative that will host the BoF meeting at the next RDA8 Plenary meeting in Denver on 15-17 September 2016.

7 Conclusion and further developments

The presented initial definition of the Data Science Model Curriculum (MC-DS) have been done with wide consultation and engagement of different stakeholders, primarily from research community and Research Infrastructures, but also involving industry via standardisation bodies, professional communities and directly via the project network.

7.1 Summary of findings

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

The approach and first draft of the proposed Model Curriculum has been presented and discussed at the EDISON Champions Conference on (13-14 July 2016, New Forest, UK). The internal Champions demonstrated the application of the Data Science Competence Framework and Body of Knowledge components for developing their own Data Science programmes and other academic offerings, providing valuable insights, comments, and suggestions that have been incorporated into the current MC-DS version that will be further presented to the ELG meeting that is planned for 27 September 2016.

7.2 Further developments to formalize MC-DS and DS-BoK

It is anticipated that the presented here the first versions of the Data Science Body of Knowledge will require further development and validation by experts and communities of practice that will include the following specific tasks and activities:

- Collect feedback on the Data Science Model Curriculum initial version further improvement and extension.
- Engage with the partner and champion universities into pilot implementation of MC-DS and DS-BoK and collecting feedback from practitioners.
- 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 suggestion for new knowledge areas and classifications classes.

Validation is an important part of the products that could be widely accepted by community. Validation of the proposed MC-DS and DS-BoK will be done in two main ways. First is presenting the proposed development to the communities of practice and soliciting feedback and contribution from the academic and professional community, including experts' interviews. The second way suggests involving the champion universities into validation and pilot implementation of the proposed DS-BoK and Model Curriculum.

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

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

8 References

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

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

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

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

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

[6] EDISON Deliverable D2.2 "Existing educational and training resources inventory and analysis", May 2016, http://edison-project.eu/sites/edison-project.eu/files/filefield_paths/edison_d2.2_inventory_and_analysis_v1_final.pdf.

[7] EDISON Deliverable D2.1 "Data Scientist Competences and Skills Framework (CF-DS) and BoK definition (first version)", February 2016, http://edison-project.eu/sites/edison-project.eu/files/filefield_paths/edison_d2.1_cf-ds_and_bok_definition_first_version.pdf.

[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.

Additional literature related to competence based learning and curricula design approaches

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

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

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

[31] EDISON Deliverable D3.2 EDISON Online Education Environment, September 2016.

[32] European Skills, Competences, Qualifications and Occupations (ESCO) [online] https://ec.europa.eu/esco/portal/home

[33] University of Bedfordshire - Data Science - BSc (Hons) programme [online] http://www.beds.ac.uk/howtoapply/courses/undergraduate/next-year/data-science

[34] University of Perugia – Master Certificate in Data Science [online] http://masterds.unipg.it/

[35] Goethe University Frankfurt – Faculty of Compuzter Science and Mathematics [online] http://www.goethe-university-frankfurt.de/58074304/Faculty-of-Computer-Science-and-Mathematics

[36] Goethe University Frankfurt - Frankfrut Big Data Lab [online] http://www.bigdata.uni-frankfurt.de/teaching/

[37] Amsterdam Data Science [online] http://amsterdamdatascience.nl/

[38] Amsterdam Data Science – Education [online] http://amsterdamdatascience.nl/education/

[39] Vrije Universiteit Amsterdam – Computer Science Specializations [online] http://vu.nl/nl/opleidingen/masteropleidingen/opleidingenoverzicht/c-d/computer-science/specializations/index.aspx

[40] University of Amsterdam/SURFsara - HPC and BigData course [online] http://hpc.uva.nl/

[41] University of Stavanger – Course Data-intensive Systems [online] http://www.uis.no/studies/study-courses/?categoryID=10648&parentcat=9835&code=DAT500_1&name=Data-Intensive+systems

[42] University of Lodz – Specialization Biomedical Engineering [online] http://www.programy.p.lodz.pl/kierunekSiatka.jsp?l=en&w=Biomedical Engineering&p=4291&stopien=first-cycle programme&tryb=full-time

[43] Engineering Training Center "Enrico Della Valle" [online] http://www.eng.it/lavora-con-noi/it-academy.dot

[44] Persontyle Data Science Centre of Excellence [online] http://www.persontyle.com/school/

[45] Persontyle Data Science Centre of Excellence – Learning Programs [online] http://www.persontyle.com/courses/

[46] Lucerne University of Applied Sciences and Art – Lucerne School of Information Technology [online] https://www.hslu.ch/en/lucerne-school-of-information-technology/

[47] Lucerne University of Applied Sciences and Art [online] https://www.hslu.ch/en/

[48] Lucerne University of Applied Sciences and Art – Degree Programmes [online] https://www.hslu.ch/en/lucerne-school-of-information-technology/degree-programs/bachelor/majors/

[49] Alistair Cockburn - Use case fundamentals [online] http://alistair.cockburn.us/Use+case+fundamentals

[50] EDISON Project: University Programs list [online] http://edison-project.eu/university-programs-list

[51] EDISON Project Documents Library [online] http://edison-project.eu/library

[52] ENQA - European Association for Quality Assurance in Higher Education [online] http://www.enqa.eu/

[53] EQANIE - the European Quality Assurance Network for Informatics Education [online] http://www.eqanie.eu/

[54] 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

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

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
CS	Computer Science
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
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
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 short overview and compare definition of mastery levels as used in the European Qualifications Framework (EQF) [25], e-CF, ACM/IEEE guidelines for Computer Science curriculum [6] 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 [25] 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 is presented in Table 6. 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 6 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 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,	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

EQF level	EQF level description	e-CF level	e-CF level description
	reviewing and developing performance of self and others.		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 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 straight forward 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 amount 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 [7] define the three levels as: emerging, developed and highly developed. Computer Science guidelines [6] 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 **Error! Reference source not found.**. The verb usage is not fully c onsistent with the original Bloom's taxonomy [16] or 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. Deeper inspection suggests that ACM/IEEE's familiarity level maps to knowledge and comprehension levels in Bloom's taxonomy. Further, usage level in ACM/IEEE maps to analysis level in Bloom's taxonomy; and finally, assessment level in ACM/IEEE maps to 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 7 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 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, 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 8. Action verbs were defined based on the original and revised Bloom's taxonomy with adjustments tailored to Data Science curricula.

Table 8 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) [7] 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 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 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 is 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 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 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 initial analysis of ACM CCS (2012) classification and 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 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 Research Data community and documented in community best practices.

Table 9 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

DS-BoK Knowledge Groups *)	ACM (2012) Classification facets related to Data Science
	DSA Extension point: Big Data systems (e.g. cloud based) Information systems applications DSA Extension point: Big Data applications 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

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

*) All Acronyms for classification groups and DS-BoK Knowledge Area Groups are brought in accordance to CF-DS-competence groups

**) Due to important role of the Internet and networking technologies, basic knowledge about networks are required. However, as a technology domain, Networks knowledge area group should be considered as a domain specific knowledge area in the general Data Science competences and knowledge definition.

Appendix C. Data Science Body of Knowledge (excerption from DS-BoK [2])

The DS-BoK is defined based on the proposed competences model CF-DS defies five competence areas that should be mapped into corresponding knowledge areas and groups (refer to the recent CF-DS version online [1]). The DS-BoK definition requires combination and synthesis of different domain knowledge areas with necessary selection or adaptation of educational and instructional models and practices.

C.1. General Approach and Structure of DS-BoK

The intended DS-BoK can be used as a base for defining Data Science related curricula, courses, instructional methods, educational/course materials, and necessary practices for university post and undergraduate programs and professional training courses. The DS-BoK is also intended to be used for defining certification programs and certification exam questions. While CF-DS (comprising of competences, skills and knowledge) can be used for defining job profiles (and correspondingly content of job advertisements) the DS-BoK can provide a basis for interview questions and evaluation of the candidate's knowledge and related skills.

Following the CF-DS competence group definition the DS-BoK should contain the following Knowledge Area groups (KAG):

- KAG1-DSDA: Data Analytics group including Machine Learning, statistical methods, and Business Analytics
- KAG2-DSENG: Data Science Engineering group including Software and infrastructure engineering
- KAG3-DSDM: *Data Management group including data curation, preservation and data infrastructure*
- KAG4-DSRMP: *Research Methods and Project Management*
- KAG5-DSBA: Business Analytics
- KAG*-DSDK: Placeholder for the Data Science Domain Knowledge groups to include domain specific knowledge

The subject domain related knowledge group (scientific or business) KAG*-DSDK is recognized as essential for practical work of Data Scientist what in fact means not professional work in a specific subject domain but understanding the domain related concepts, models and organisation (as discussed in section 3.8.3) and corresponding data analysis methods and models. These knowledge areas will be a subject for future development in tight cooperation with subject domain specialists.

It is also anticipated that due to complexity of Data Science domain, the DS-BoK will require wide spectrum of background knowledge, first of all in mathematics, statistics, logics and reasoning as well as general computing and cloud computing in particular. Similar to the ACM CS2013 curricula approach, background knowledge can be required as an entry condition or must be studied as elective courses.

The proposed DS-BoK re-uses where possible or provides links to existing BoK's taking necessary KA definitions and combining them into defined above DS-BoK knowledge area groups. The following BoK's can be used or mapped to the selected DS-BoK knowledge groups:
ACM Computer Science CS-BoK [7, 8]
Business Analysis BABOK [10]
Software Engineering SWEBOK [11]
Data Management DMBOK by DAMA [12],
Project Management PM-BoK [13],
Classification Computer Science (CCS2012) [6] for Computer Science related knowledge areas.

C.2. Data Analytics Knowledge Area

Data Analytics includes different methods and algorithms, primarily statistical, to enable data processing, modelling, analysis and inspection with the goal of discovering useful information, providing insight and recommendations, and supporting decision-making. The following are commonly defined the Data Science Analytics Knowledge Areas:

- KA01.01 (DSDA.01/SMA) 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.02/ML) Machine learning and related methods for information search, image recognition, decision support, classification;
- KA01.03 (DSDA.03/DM) *Data mining* is a particular data analysis technique that focuses on modelling and knowledge discovery for predictive rather than purely descriptive purposes;
- KA01.04 (DSDA.04/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.05/PA) Predictive analytics focuses on application of statistical models for predictive forecasting or classification.
- KA01.06 (DSDA.06/BA) Business Analytics and Business Intelligence covers data analysis that relies heavily on aggregation and different data sources and focusing on business information;
- KA01.07 (DSDA.07/MSO) Computational modelling, simulation and optimisation

C.3. DS-BoK Knowledge Area Groups

Presented analysis allows us to propose an initial version of the Data Science Body of Knowledge implementing the proposed DS-BoK structure as explained in previous section. Table C.1 provides consolidated view of the identified Knowledge Areas in the Data Science Body of Knowledge. The table contains detailed definition of the KAG1-DSDA, KAG2-DSENG, KAG3-DSDM groups that are well supported by existing BoK's and academic materials. General suggestions are provided for KAG4-DSRMP, KAG5-DSBA groups that corresponds to newly identified competences and knowledge areas and require additional study of existing practices and contribution from experts in corresponding scientific or business domains.

The KAG2-DSENG group includes selected KAs from ACM CS-BoK and SWEBOK and extends them with new technologies and engineering technologies and paradigm such as cloud based, agile technologies and DevOps that are promoted as continuous deployment and improvement paradigm and allow organisation implement agile business and operational models.

The KAG3-DSDM group includes most of KAs from DM-BoK however extended it with KAs related to RDA recommendations, community data management models (Open Access, Open Data, etc.) and general Data Lifecycle Management that is used as a central concept in many data management related education and training courses.

Knowledge Units (KU) corresponding to suggested KAs are defined from different sources: existing BoK, CCS2012, and from practices in designing academic curricula and corresponding courses by universities and professional training organisations.

For the detailed definition of the KA and KU refer to the DS-BoK document [2]. The DS-BoK document contains detailed definition of the KAG1-DSA, KAG2-DSE, KAG3-DSDM, KAG4-DSRM, KAG5-DSBA groups that corresponds to newly identified competences and knowledge areas and require additional study of existing practices and contribution from experts in corresponding scientific or business domains.

Table C.1. DS-BoK Knowledge Area Groups and corresponding Knowledge Areas

KA Groups	Suggested DS Knowledge Areas (KA)	Knowledge Areas from existing BoK and CCS2012 scientific subject groups
KAG1-DSDA: Data Science Analytics	KA01.01 (DSDA.01/SMDA) Statistical methods for data analysis KA01.02 (DSDA.02/ML) Machine Learning KA01.03 (DSDA.03/DM) Data Mining KA01.04 (DSDA.04/TDM) Text Data Mining KA01.05 (DSDA.05/PA) Predictive Analytics KA01.06 (DSDA.06/MODSIM) Computational modelling, simulation and optimisation	There is no formal BoK defined for Data Analytics. Data Science Analytics related scientific subjects from CCS2012: CCS2012: Computing methodologies CCS2012: Mathematics of computing CCS2012: Computing methodologies
KAG2-DSENG: Data Science Engineering	KA02.01 (DSENG.01/BDI) Big Data Infrastructure and Technologies KA02.02 (DSENG.02/DSIAPP) Infrastructure and platforms for Data Science applications KA02.03 (DSENG.03/CCT) Cloud Computing technologies for Big Data and Data Analytics KA02.04 (DSENG.04/SEC) Data and Applications security KA02.05 (DSENG.05/BDSE) Big Data systems organisation and engineering KA02.06 (DSENG.06/DSAPPD) Data Science (Big Data) applications design KA02.07 (DSENG.07/IS) Information systems (to support data driven decision making)	ACM CS-BoK selected KAs: AL - Algorithms and Complexity AR - Architecture and Organization (including computer architectures and network architectures) CN - Computational Science GV - Graphics and Visualization IM - Information Management PBD - Platform-based Development (new) SE - Software Engineering (can be extended with specific SWEBOK KAs) SWEBOK selected KAs • Software requirements • Software design • Software engineering process • Software engineering models and methods • Software quality Data Science Analytics related scientific subjects from CCS2012: CCS2012: Computer systems organization CCS2012: Information systems CCS2012: Software and its engineering
KAG3-DSDM: Data Management	KA03.01 (DSDM.01/DMORG) General principles and concepts in Data Management and organisation KA03.02 (DSDM.02/DMS) Data management systems KA03.03 (DSDM.03/EDMI) Data Management and Enterprise data infrastructure KA03.04 (DSDM.04/DGOV) Data Governance KA03.05 (DSDM.05/BDST0R) Big Data storage (large scale)	DM-BoK selected KAs (1) Data Governance, (2) Data Architecture, (3) Data Modelling and Design, (4) Data Storage and Operations, (5) Data Security, (6) Data Integration and Interoperability, (7) Documents and Content, (8) Reference and Master Data, (9) Data Warehousing and Business Intelligence, (10) Metadata, and (11) Data Quality.

KA Groups	Suggested DS Knowledge Areas (KA)	Knowledge Areas from existing BoK and CCS2012 scientific subject groups
	KA03.06 (DSDM.05/DLIB) Digital libraries and archives	Data Science Analytics related scientific subjects from CCS2012: CCS2012: Information systems
KAG4-DSRM: Research Methods and Project Management	KA04.01 (DSRMP.01/RM) Research Methods KA04.01 (DSRMP.02/PM) Project Management	There are no formally defined BoK for research methods PMI-BoK selected KAs • Project Integration Management • Project Scope Management • Project Quality • Project Risk Management
KAG5-DSBPM: Business Analytics	KA05.01 (DSBA.01/BAF) Business Analytics Foundation KA05.02 (DSBA.02/BAEM) Business Analytics organisation and enterprise management	BABOK selected KAs *) • Business Analysis Planning and Monitoring: describes the tasks used to organize and coordinate business analysis efforts. • Requirements Analysis and Design Definition. • Requirements Life Cycle Management (from inception to retirement). • Solution Evaluation and improvements recommendation.

*) BABOK KA are more business focused and related to KAG5-DSBA, however its specific topics related to data analysis can be reflected in the KAG1-DSDA

8.1 Data Science Body of Knowledge Areas and Knowledge Units

Presented analysis allows us to propose an initial version of the Data Science Body of Knowledge implementing the proposed DS-BoK structure as explained in previous section. Table C.2 provides consolidated view of the identified Knowledge Areas in the Data Science Body of Knowledge. The table contains detailed definition of the KAG1-DSA, KAG2-DSE, KAG3-DSDM groups that are well supported by existing BoK's and academic materials. General suggestions are provided for KAG4-DSRM, KAG5-DSBP groups that corresponds to newly identified competences and knowledge areas and require additional study of existing practices and contribution from experts in corresponding scientific or business domains.

The KAG2-DSE group includes selected KAs from ACM CS-BoK and SWEBOK and extends them with new technologies and engineering technologies and paradigm such as cloud based, agile technologies and DevOps that are promoted as continuous deployment and improvement paradigm and allow organisation implement agile business and operational models.

The KAG3-DSDM group includes most of KAs from DM-BoK however extended it with KAs related to RDA recommendations, community data management models (Open Access, Open Data, etc) and general Data Lifecycle Management that is used as a central concept in many data management related education and training courses.

The presented DS-BoK high level content is not exhaustive at this stage and will undergo further development based on feedback from MC-DS implementation. The project will present the current version of DS-BoK to ELG to obtain feedback and expert opinion. Numerous experts will be invited to review and contribute to the specific KAs definition.

Table C.2 Detailed definition of the DS-BoK and suggested Knowledge Units (KU)

Knowledge Area Groups (KAG)	Knowledge Areas (KA)	Knowledge Unit (KU)	Suggested Knowledge Units (KU)	Mapping to CCS2012 and existing BoKs (DMBOK, BABOK, PMI-BoK, SWEBOK, ACM BoK)
KAG1-DSDA: Data Science Analytics	KA01.01 DSDA.01/SMDA Statistical methods for data analysis	KU1.01.01	Probability & Statistics	**CCS2012: 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
		KU1.01.02	Statistical paradigms (regression, time series, dimensionality, clusters)	
		KU1.01.03	Probabilistic representations (causal networks, Bayesian analysis, Markov nets)	
		KU1.01.04	Frequentist and Bayesian statistics	
		KU1.01.05	Probabilistic reasoning	
		KU1.01.06	Exploratory and confirmatory data analysis	
		KU1.01.07	Quantitative analytics	
		KU1.01.08	Performance analysis	
		KU1.01.09	Markov models, Markov networks	
		KU1.01.10	Operations research	
		KU1.01.11	Information theory	
		KU1.01.12	Discrete Mathematics and Graph Theory	
		KU1.01.13	Mathematical analysis	
		KU1.01.14	Mathematical software and tools	
KAG1-DSDA: Data Science Analytics	KA01.02 DSDA.02/ML Machine Learning	KU1.02.01	Machine Learning theory and algorithms	**CCS2012: Computing methodologies** • Artificial intelligence ○ Machine learning ○ Learning paradigms ▪ Supervised learning ▪ Unsupervised learning ▪ Reinforcement learning ▪ Multi-task learning • Machine learning approaches ○ Machine learning algorithms
		KU1.02.02	Supervised Machine Learning	
		KU1.02.03	Unsupervised Machine Learning	
		KU1.02.04	Reinforced learning	
		KU1.02.05	Classification methods	
		KU1.02.06	Design and Analysis of Algorithms	
		KU1.02.07	Game Theory & Mechanism design	
		KU1.02.08	Artificial Intelligence	
		KU1.01.02	Statistical paradigms (regression, time series, dimensionality, clusters)	
		KU1.01.03	Probabilistic representations (causal networks, Bayesian analysis, Markov nets)	
		KU1.01.04	Frequentist and Bayesian statistics	**CCS2012: Theory of computation**
		KU1.01.05	Probabilistic reasoning	
		KU1.01.08	Performance analysis	

Knowledge Area Groups (KAG)	Knowledge Areas (KA)	Knowledge Unit (KU)	Suggested Knowledge Units (KU)	Mapping to CCS2012 and existing BoKs (DMBOK, BABOK, PMI-BoK, SWEBOK, ACM BoK)
				• 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 • Semantics and reasoning
KAG1-DSDA: Data Science Analytics	KA01.03 DSDA.03/DM Data Mining	KU1.01.08	Performance analysis	**CCS2012: 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 • Semantics and reasoning
		KU1.02.01	Machine Learning theory and algorithms	
		KU1.02.02	Supervised Machine Learning	
		KU1.02.03	Unsupervised Machine Learning	
		KU1.02.04	Reinforced learning	
		KU1.02.05	Classification methods	
		KU1.03.01	Data mining and knowledge discovery	
		KU1.03.02	Knowledge Representation and Reasoning	
		KU1.03.03	CRISP-DM and data mining stages	
		KU1.03.04	Anomaly Detection	
		KU1.03.05	Time series analysis	
		KU1.03.06	Feature selection, Apriori algorithm	
		KU1.03.07	Graph data analytics	
KAG1-DSDA: Data Science Analytics	KA01.04 DSDA.04/TDM Text Data Mining	KU1.04.01	Text analytics including statistical, linguistic, and structural techniques to analyse structured and unstructured data	**CCS2012: Computing methodologies** • Artificial intelligence ○ Natural language processing ○ Knowledge representation and reasoning ○ Search methodologies
		KU1.04.02	Data mining and text analytics	
		KU1.04.03	Natural Language Processing	
		KU1.04.04	Predictive Models for Text	
		KU1.04.05	Retrieval and Clustering of Documents	
		KU1.04.06	Information Extraction	

Knowledge Area Groups (KAG)	Knowledge Areas (KA)	Knowledge Unit (KU)	Suggested Knowledge Units (KU)	Mapping to CCS2012 and existing BoKs (DMBOK, BABOK, PMI-BoK, SWEBOK, ACM BoK)
		KU1.04.07	Sentiments analysis	
KAG1-DSDA: Data Science Analytics	KA01.05 DSDA.05/PA Predictive Analytics	KU1.05.01	Predictive modeling and analytics	
		KU1.05.02	Inferential and predictive statistics	
		KU1.05.03	Machine Learning for predictive analytics	
		KU1.05.04	Regression and Multi Analysis	
		KU1.05.05	Generalised linear models	
		KU1.05.06	Time series analysis and forecasting	
		KU1.05.07	Deploying and refining predictive models	
KAG1-DSDA: Data Science Analytics	KA01.06 DSDA.06/MODSIM Computational modelling, simulation and optimisation	KU1.06.01	Modelling and simulation theory and techniques (general and domain oriented)	**CCS2012: Computing methodologies** • Modeling and simulation ○ Model development and analysis ○ Simulation theory ○ Simulation types and techniques ○ Simulation support systems
		KU1.06.02	Operations research and optimisation	
		KU1.06.03	Large scale modelling and simulation systems	
		KU1.06.04	Network oprtimisation	
		KU1.06.05	Risk simulation and queueing	
KAG2-DSENG: Data Science Engineering	KA02.01 DSENG.01/BDI Big Data Infrastructure and Technologies	KU2.01.01	Computer systems organisation for Big Data applications, CAP, BASE and ACID theorems	**CCS2012: Computer systems organization** • Architectures ○ Parallel architectures ○ Distributed architectures • Networks *) ○ Network Architectures ○ Network Services ○ Cloud Computing
		KU2.01.02	Parallel and Distributed Computer Architecture	
		KU2.01.03	High Performance and Cloud Computing	
		KU2.01.04	Clouds and scalable computing	
		KU2.01.05	Cloud based Big Data platforms and services	
		KU2.01.06	Big Data (large scale) storage and filesystems (HDFS, Ceph, etc)	
		KU2.01.07	NoSQL databases	
		KU2.01.08	Computer networks for high-performance computing and Big Data infrastructure	
		KU2.01.09	Computer networks: architectures and protocols	

Knowledge Area Groups (KAG)	Knowledge Areas (KA)	Knowledge Unit (KU)	Suggested Knowledge Units (KU)	Mapping to CCS2012 and existing BoKs (DMBOK, BABOK, PMI-BoK, SWEBOK, ACM BoK)
		KU2.01.10	Big Data Infrastructure management and operation	
KAG2-DSENG: Data Science Engineering	KA02.02 DSENG.02/DSIAPP Infrastructure and platforms for Data Science applications	KU2.02.01	Big Data Infrastructure: services and components, including data storage infrastructure	• Proposed new KA for DS-BoK • Infrastructure and platforms for Data Science applications group: • CCENG - Cloud Computing Engineering (infrastructure and services design, management and operation) • CCAS - Cloud based applications and services development and deployment • BDA – Big Data Analytics platforms (including cloud based) • BDI - Big Data Infrastructure services and platforms, including data storage infrastructure
		KU2.02.02	Big Data analytics platforms and tools (including Hadoop, Spark, and cloud based Big Data services)	
		KU2.02.03	Large scale cloud based storage and data management	
		KU2.02.04	Cloud based applications and services operation and management	
		KU2.02.05	Big Data and cloud based systems design and development	
		KU2.02.06	Data processing models (batch, steaming, parallel)	
		KU2.02.07	Enterprise information systems	**CCS2012: Information systems** • Information storage systems • Information systems applications
		KU2.02.08	Data security and protection	
KAG2-DSENG: Data Science Engineering	KA02.03 DSENG.03/CCT Cloud Computing technologies for Big Data and Data Analytics	KU2.03.01	Cloud Computing architecture and services	**DSDA Extension group for CCS201** **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**
		KU2.03.02	Cloud Computing Engineering (infrastructure and services design, management and operation)	
		KU2.03.03	Cloud based applications and services operation and management	
KAG2-DSENG: Data Science Engineering	KA02.04 DSENG.04/SEC Data and Applications security	KU2.04.01	Infrastructure, applications and data security	
		KU2.04.02	Data encryption and key management, bockchain based technologies	

Knowledge Area Groups (KAG)	Knowledge Areas (KA)	Knowledge Unit (KU)	Suggested Knowledge Units (KU)	Mapping to CCS2012 and existing BoKs (DMBOK, BABOK, PMI-BoK, SWEBOK, ACM BoK)
		KU2.04.03	Access Control and Identiy Management	• 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 DSA Extension point: Doman specific Data applications
		KU2.04.04	Security services management, including compliance and certification	
		KU2.04.05	Data anonymisation	
		KU2.04.06	Data privacy	
KAG2-DSENG: Data Science Engineering	KA02.05 DSENG.05/BDSE Big Data systems organisation and engineering	KU2.05.01	Big Data systems organisation and design	**CCS2012: Software and its engineering** • Software organization and properties ○ Software system structures • Software architectures ○ Software system models ○ Distributed systems organizing principles ▪ Cloud computing ▪ Grid computing • Software notations and tools ○ General programming languages ○ Software creation and management
		KU2.05.02	Big Data algorithms for large scale data processing	
		KU2.05.03	Big Data Analytics	
		KU2.05.04	Big Data analytics platforms and tools (including Hadoop, Spark, and cloud based Big Data services)	
		KU2.05.05	Big Data algorithms for data ingest, pre-processing, and visualisation	
		KU2.05.06	Big Data systems for application domains	
		KU2.05.07	Big Data software (systems) architectures	
		KU2.05.08	Requirements engineering and software systems development	
		KU2.05.09	Large and ultra-large scale software systems organisation	
		KU2.05.10	DevOps and cloud enabled applications development	
		KU2.05.11	Big Data Infrastructure management and operation	
KAG2-DSENG: Data Science Engineering	KA02.06 DSENG.06/DSAPPD Data Science (Big Data) applications design	KU2.06.01	Data analytics, data handling software requirements and design	**SWEBOK selected KAs** • Software requirements • Software design • Software construction • Software testing
		KU2.06.02	Applications engineering management	
		KU2.06.03	Software engineering models and methods	
		KU2.06.04	Software quality assurance	

Knowledge Area Groups (KAG)	Knowledge Areas (KA)	Knowledge Unit (KU)	Suggested Knowledge Units (KU)	Mapping to CCS2012 and existing BoKs (DMBOK, BABOK, PMI-BoK, SWEBOK, ACM BoK)
		KU2.06.05	Programming languages for Big Data analytics: R, python, Pig, Hive, others	• Software maintenance • Software configuration management • Software engineering management • Software engineering process • Software engineering models and methods • Software quality • Agile development technologies • Methods, platforms and tools • DevOps and continuous deployment and improvement paradigm
		KU2.06.06	Models and languages for complex interlinked data presentation and visualisation	
		KU2.06.07	Agile development methods, platforms and tools	
		KU2.06.08	DevOps and continuous deployment and improvement paradigm	
KAG2-DSENG: Data Science Engineering	KA02.07 DSENG.07/IS Information systems (to support data driven decision making)	KU2.07.01	Decision Analysis and Decision Support Systems	**CCS2012: Information systems** • Information systems applications ○ Decision support systems ▪ Data warehouses ▪ Expert systems ▪ Data analytics ▪ Online analytical processing ○ Multimedia information systems ○ Data mining
		KU2.07.02	Predictive analytics and predictive forecasting	
		KU2.07.03	Data Analysis and statistics	
		KU2.07.04	Data warehousing and Data Mining	
		KU2.07.05	Data Mining	
		KU2.07.06	Multimedia information systems	
		KU2.07.07	Enterprise information systems	
		KU2.07.08	Collaborative and social computing systems and tools	
				CCS2012: Information systems • Information systems applications ○ Enterprise information systems ○ Collaborative and social computing systems and tools
	KA03.01	KU3.01.01	Data type registries, PID, metadata	

Knowledge Area Groups (KAG)	Knowledge Areas (KA)	Knowledge Unit (KU)	Suggested Knowledge Units (KU)	Mapping to CCS2012 and existing BoKs (DMBOK, BABOK, PMI-BoK, SWEBOK, ACM BoK)
KAG3-DSDM: Data Management	DSDM.01/DMORG General principles and concepts in Data Management and organisation	KU3.01.02	Data Lifecycle Management	**Proposed new KA for DS-BoK** General Data Management KA's • Data Lifecycle Management • Data archives/storage compliance and certification
		KU3.01.03	Data infrastructure and Data Factories	
		KU3.01.04	Research data infrastructure, Open Science, Open Data, Open Access, ORCID	
		KU3.01.05	Data infrastructure compliance and certification	New KAs to support RDA recommendations and community data management models (Open Access, Open Data, etc) • Data type registries, PIDs • Data infrastructure and Data Factories • New KAs to follow RDA and ERA community developments
		KU3.01.06	Ethical principle and data privacy	
		KU3.01.07	FAIR (Findable, Accessible, Interoperable,) principles in Data Management	
KAG3-DSDM: Data Management	KA03.02 DSDM.02/DMS Data management systems	KU3.02.01	Data architectures (OLAP, OLTP, ETL)	**CCS2012: Information systems** • Data management systems ○ Database design and models ○ Data structures ○ Database management system engines ○ Query languages ○ Database administration ○ Middleware for databases ○ Information integration • CCS2012: Theory of computation ○ Database theory
		KU3.02.02	Data Modelling, Databases and Database Management Systems	
		KU3.02.03	Data structures	
		KU3.02.04	Data Models and Query Languages	
		KU3.02.05	Database design and models	
		KU3.02.06	Database administration	
		KU3.02.07	Data warehouses	
		KU3.02.08	Middleware for databases	
KAG3-DSDM: Data Management	KA03.03 DSDM.03/EDMI Data Management and Enterprise data infrastructure	KU3.03.01	Data management, including Reference and Master Data	**DM-BoK selected KAs** (1) Data Governance, (2) Data Architecture, (3) Data Modelling and Design,
		KU3.03.02	Data Warehousing and Business Intelligence	
		KU3.03.03	Data storage and operations	

Knowledge Area Groups (KAG)	Knowledge Areas (KA)	Knowledge Unit (KU)	Suggested Knowledge Units (KU)	Mapping to CCS2012 and existing BoKs (DMBOK, BABOK, PMI-BoK, SWEBOK, ACM BoK)
		KU3.03.04	Data archives/storage compliance and certification	(4) Data Storage and Operations, (5) Data Security, (6) Data Integration and Interoperability, (7) Documents and Content, (8) Reference and Master Data, (9) Data Warehousing and Business Intelligence, (10) Metadata, and (11) Data Quality.
		KU3.03.05	Metadata, linked data, provenance	
		KU3.03.06	Data infrastructure, data registries and data factories	
		KU3.03.07	Data security and protection	
		KU3.03.08	Data backup	
		KU3.03.09	Data anonymisation	
		KU3.03.10	Data Privacy	
KAG3-DSDM: Data Management	KA03.04 DSDM.04/DGOV Data Governance	KU3.04.01	Data governance, data quality, data Integration and Interoperability	DM-BoK (as above)
		KU3.04.02	Data Management Planning	
		KU3.04.03	Data Management Policy	
		KU3.04.04	Data interoperability	
		KU3.04.05	Data curation	
		KU3.04.06	Data provenance	
		KU3.04.07	Responsible data use, data privacy, ethical principles, IPR, legal issues	
KAG3-DSDM: Data Management	KA03.05 DSDM.05/BDST0R Big Data storage (large scale)	KU3.05.01	Big Data storage infrastructure and operations	New DSENG Knowledge area: Big Data Storage • Distributed file systems • Data Lakes • Data Factories
		KU3.05.02	Storage architectures, distributed files systems (HDFS, Ceph, Lustre, Gluster, etc)	
		KU3.05.03	Data storage redundancy and backup	
		KU3.05.04	Data factories, data pipelines	
		KU3.05.05	Cloud based storage, Data Lakes	
KAG3-DSDM: Data Management	KA03.06 DSDM.05/DLIB Digital libraries and archives	KU3.06.01	Digital libraries and archives organisation	**CCS2012: Information systems** • Information systems applications ○ Digital libraries and archives
		KU3.06.02	Information Retrieval	
		KU3.06.03	Data curation and provenance	
		KU3.06.04	Search Engines technologies	
KAG4-DSRMP: Research Methods and Project Management	KA04.01 DSRMP.01/RM Research Methods	KU4.01.01	Research methodology, paradigms and research cycle	**Proposed new KA for DS-BoK for DSRM related competences:** • Research methodology, research cycle (e.g. 4 steps model
		KU4.01.02	Modelling and experiment planning	
		KU4.01.03	Data selection and quality evaluation	
		KU4.01.04	Data lifecycle	

Knowledge Area Groups (KAG)	Knowledge Areas (KA)	Knowledge Unit (KU)	Suggested Knowledge Units (KU)	Mapping to CCS2012 and existing BoKs (DMBOK, BABOK, PMI-BoK, SWEBOK, ACM BoK)
		KU4.01.05	Use cases analysis: research infrastructures and projects	Hypothesis – Research Methods – Artefact – Validation) • Modelling and experiment planning • Data selection and quality evaluation • Use cases analysis: research infrastructures and projects
		KU4.01.06	Research data management plan and ethical issues	
KAG4-DSRMPM: Research Methods and Project Management	KA04.01 DSRMP.02/PM Project Management	KU4.02.01	Project Integration Management	**PMI-BoK selected KAs** • Project Integration Management • Project Scope Management • Project Quality • Project Risk Management
		KU4.02.02	Project Scope Management	
		KU4.02.03	Project Quality	
		KU4.02.04	Project Risk Management	
KAG5-DSBPM: Business Analytics	KA05.01 DSBA.01/BAF Business Analytics Foundation	KU5.01.01	Business Analytics and Business Intelligence: Data, Models (statistical) and Decisions	**BABOK selected KAs** • Business Analysis Planning and Monitoring: describes the tasks used to organize and coordinate business analysis efforts. • Requirements Analysis and Design Definition. • Requirements Life Cycle Management (from inception to retirement). • Solution Evaluation and improvements recommendation.
		KU5.01.02	Data driven Customer Relations Management (CRP), User Experience (UX) requirements and design	
		KU5.01.03	Operations Analytics	
		KU5.01.04	Business Process Optimization	
		KU5.01.05	Data Warehouses technologies, data integration and analytics	
		KU5.01.06	Data driven marketing technologies	
		KU5.01.07	Business Analytics Capstone	
		KU5.01.08	Econometrics methods and application for Business Analytics	
		KU5.01.09	Cognitive technologies for Business Analytics	
KAG6-DSBA: Business Analytics	KA05.02 DSBA.02/BAEM Business Analytics organisation and enterprise management	KU5.02.01	Business processes and operations	**Proposed new KA/KU for DS-BoK** • General Business processes and operations KAs • Business processes and operations
		KU5.02.02	Project scope and risk management	
		KU5.02.03	Business Analysis Planning and Monitoring	
		KU5.02.04	Requirements Analysis and Design Definition	

Knowledge Area Groups (KAG)	Knowledge Areas (KA)	Knowledge Unit (KU)	Suggested Knowledge Units (KU)	Mapping to CCS2012 and existing BoKs (DMBOK, BABOK, PMI-BoK, SWEBOK, ACM BoK)
		KU5.02.05	Requirements Life Cycle Management (from inception to retirement)	• Agile Data Driven methodologies, processes and enterprises • Use cases analysis: business and industry
		KU5.02.06	Solution Evaluation and improvements recommendation	
		KU5.02.07	Agile Data Driven methodologies, processes and enterprises	
		KU5.02.08	Use cases analysis: business and industry	

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

Table D.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 a 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 source 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, publications (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 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 D.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 a 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 source 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, publications (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 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 D.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 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 a 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 source 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, publications (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 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 D.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 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 a 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 source 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, publications (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 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.						

EDISON DATA SCIENCE FRAMEWORK
EUROPEAN COMMISION GRANT AGREEMENT NO: 675419

IABAC is a registered B.V (equivalent of UK English Private Limited) company in Netherlands.
RSIN (Registration number) : 859414206
Registered Name : IABAC B.V ,
Place: Eindhoven, The Netherlands