



The EDISON Data Science Framework (EDSF)

A general introduction for accelerating and focusing the education of Data Science Professionals

The EDISON Data Science Framework (EDSF) is a comprehensive collection of interrelated documents that can be used by a range of stakeholders to construct their own structured solutions for educating, training, certifying, recruiting, managing, and otherwise supporting data scientists and other data-dependent professionals

The EDISON Data Science Framework (EDSF) has been developed to **support, guide and ultimately accelerate the education process of fit-for-purpose Data Science Professionals (DSP)**.

- ✓ **Educators and Trainers** will be able to design, promote and implement courses and programmes aimed to educate the professionals needed by industry and research organisations to succeed in the digital transformation of their activities;
- ✓ **Education decision makers** will be able to compare and evaluate competing educational offerings, to better allocate scarce educational resources;
- ✓ **Employers** will be able to better understand the competences they should look for and thus improve the design, execution and results of the processes of recruitment, utilisation and progression of Data Science Professionals in their organisations;
- ✓ **Policy makers** will be able to develop faster and more effective policy;
- ✓ **Future Data Science Professionals** will understand and select the educational offering better suited to their personal ambitions, capabilities and expectations.



Why is the EDSF needed?

Data Science is an emerging field that requires multi-disciplinary knowledge and is built with strong links to Big Data and other data-driven technologies. This complexity consequently requires rethinking and redesigning both in terms of traditional educational models and also existing courses. In the jobs market, there is an urgent demand for fit-for-purpose data science skills and competences. In the years ahead, Europe will face a large and growing shortage of Data Scientists with the right skills and competencies to efficiently adopt the big data approach in their daily work. Unfortunately, up to now, there are no standards or guidelines for developing Data Science curricula; trainers and educators are still relying on their own understanding of Data Science which has resulted in a diverse range of Data Science curricula with different content and topics and with a restricted focus on only a few competence areas.

The EDSF components

The EDSF components describe methodologies and tools to support and assist educators and trainers in developing their curricula and other training tracks that will help deliver DSP with the skills and competences required by the job market.

Competence framework for Data Science (CF-DS)¹ is an extended analysis of existing frameworks for Data Science and ICT competences and skills. CF-DS identifies competence groups, knowledge areas, and skills based on an extensive analysis of the actual Data Scientist job market profession in industry and research.

The Competence framework for Data Science (CF-DS) can be used to:

- ✓ Define Data Science competences and skills and associated proficiency levels
- ✓ Develop an Interactive/web based tool for assessing individual competence profiles
- ✓ Build customised Data Science training programmes
- ✓ Produce the associated Data Science body of knowledge (BoK)
- ✓ Develop related data-oriented educational programs
- ✓ Deliver solutions aligned to Bloom's taxonomy

¹ <http://edison-project.eu/data-science-competence-framework-cf-ds>



A general introduction to EDISON Data Science Framework (EDSF) (V1.3)

Data Science Body of Knowledge (DS-BoK)² is built around knowledge area groups that are linked to the CF-DS competence groups. It includes relevant knowledge areas from existing bodies of knowledge and new knowledge areas that are required to fulfil the required Data Science competences.

The Data Science Body of Knowledge (DS-BoK) defines the **foundations for:**

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

EDISON Data Science Model Curriculum (MC-DS)³ is a competence-based curriculum design approach grounded in the Data Science competences defined in CF-DS and the corresponding Learning Outcomes. MC-DS defines Learning Units that help structure academic curricula and courses. Learning Units are based on the ACM Classification of Computer Science (CCS2012) and reflect typical course titles as used by universities in their current programs.

The Data Science Model Curriculum (MC-DS) can be used as:

- ✓ A template for developing Data Science programs, courses and trainings, with a balanced selection according to requirements elicited from research and industry domains.
- ✓ An assessment reference for existing Data Science programs with respect to the knowledge areas and competence groups that are associated with specific professional profiles
- ✓ An improvement reference for existing Data Science programs
- ✓ A tool to personalize and customize curricula to fit a specific Data Science professional

² <http://edison-project.eu/data-science-body-knowledge-ds-bok>

³ <http://edison-project.eu/data-science-model-curriculum-mc-ds>



Applying EDSF for designing Data Science curricula

This section provides an example on how the EDSF, and in particular its components **CF-DS**, **DS-BoK**, **MC-DS**, and **DS-profiles**, can facilitate the design of a new Data Science curriculum or the evaluation of an existing curriculum for compliance with the selected Data Science professional profiles.

Designing a new programme

The EDSF is currently being used by EDISON Champion universities⁴ to design new DS curricula that meet identified industry demands. This is done by targeting different Data Science professional profiles as shown in **Table 1**. The EDSF follows a competence-based approach to assist in defining the focus of educational programs, and adapt them to a specific professional profile. Different levels of proficiency are achievable for each Learning Outcome depending on the targeted professional profile. Three proficiency levels are considered in EDSF: **(1)** general understanding of the Data Science concepts, **(2)** the ability to apply these concepts to solve concrete problems, **(3)** the ability to further assess and develop these concepts to create new knowledge.

Data Science Professional Profiles: competence groups

		Data Science Professional Profiles				
		Managers : DSP01-DS03	Professionals: DSP04-DS09	Professionals (data handling/management: DSP10-13	Professionals (database): DSP14-DS16	Technician and associate profession: DSP17-DS19
Data Science Competence groups	Data analytics					
	Data Science Engineering					
	Data Management					
	Scientific research & method					
	Business process					
	Domain Knowledge					

Legend: 1) Bars represent individual DSP profiles
 2) Colours represent mastery level: familiarity–light blue; usage–blue; assessment–dark blue.
 3) Red rectangle represents proficiencies needed by Data Science Profile 01 (DSP 01) for each of the 6 Data Science competence group

Table 1: The Proficiency Table maps the competency groups needed by each identified Data Science Profile. Rows represent the Data Science competence groups and columns represent the ESCO⁵ occupations family with the respective Data Science specific profiles.

⁴ See website for further details; <http://edison-project.eu>

⁵ European Skills, Competencies, Qualification and Occupation



Assessment of existing educational programmes and identification of potential gaps

EDSF has been used to evaluate the focus of a selection of representative Data Science programs from the EDISON Data Science programs inventory,⁶ including programs developed by the EDISON Champion universities. A clear gap has been identified related to Data Management competences, which are crucial for many Data Science professional profiles. As consequence, an initiative to create a reference Data Management curriculum has been undertaken by one of these universities in collaboration with the Research Data Alliance (RDA)⁷. A series of training modules will be offered on Research Data Management Literacy collocated with future RDA plenary meetings while the universities are finalising the formal adoption of the topic in their current offering.

As Data Science is an emerging and rapidly changing area, it will be necessary to periodically revise and adapt the Data Science curricula in order to remain relevant for the Data Science job market. The EDISON methodology takes into account Data Science professional profiles that are based on the ESCO⁸ occupation family profiles. This implies that the EDSF can be easily adapted, and provides a suitable reference for monitoring and identifying gaps in Data Science curricula and courses.

Contributors:

Adam Belloum (University of Amsterdam),
Yuri Demchenko (University of Amsterdam),
Wouter Los (University of Amsterdam),
Tomasz Wiktorski (University of Stavanger),
Steve Brewer (University of Southampton),
Andrea Manieri (Ingegneria Informatica),
Holger Brocks (Research Institute for Telecommunication and Cooperation),
Ruben Riestra (Inmark Europa),
Themis Athanassiadou (egi.eu)

For more information Contact:

- General email: info@edison-project.eu
- Project director: Yuri Demchenko - y.demchenko@uva.nl
- Project dissemination and engagement: Steve Brewer - s.brewer@soton.ac.uk

⁶ <http://edison-project.eu/university-programs-list>

⁷ <https://www.rd-alliance.org>

⁸ European Skills, Competencies, Qualification and Occupation