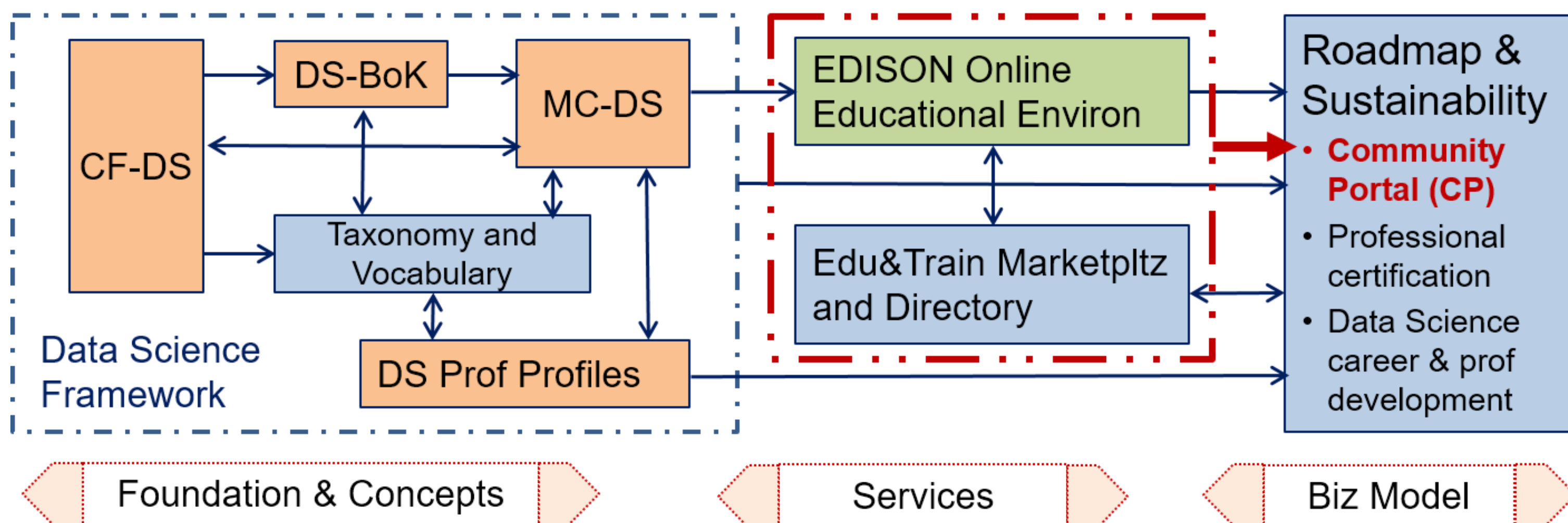




# EDISON Data Science Framework (EDSF): Customising Education and Training for better Career Management



## EDISON Data Science Framework (EDSF)

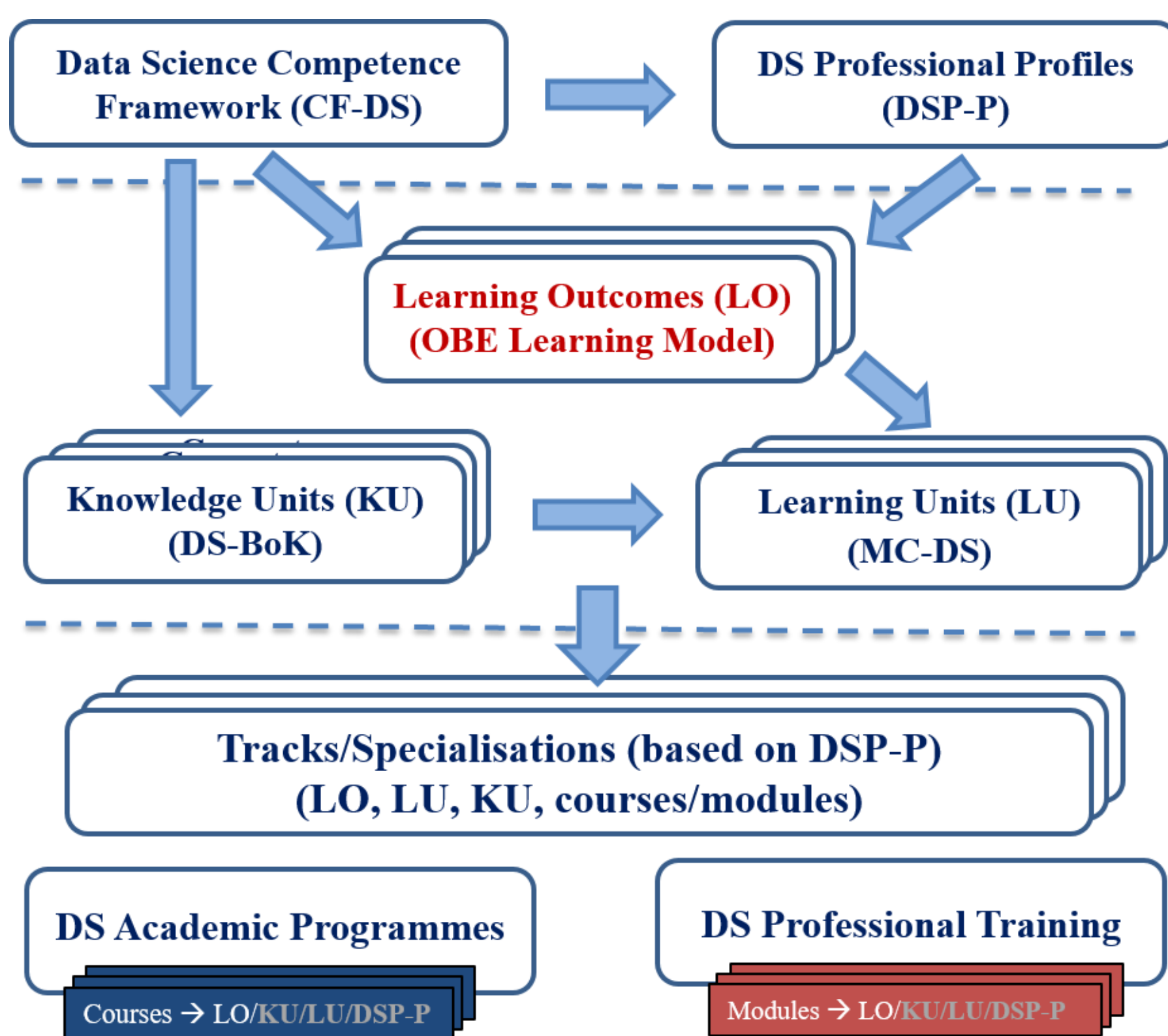


### EDISON Data Science Framework (EDSF) provides conceptual basis for the Data Science profession

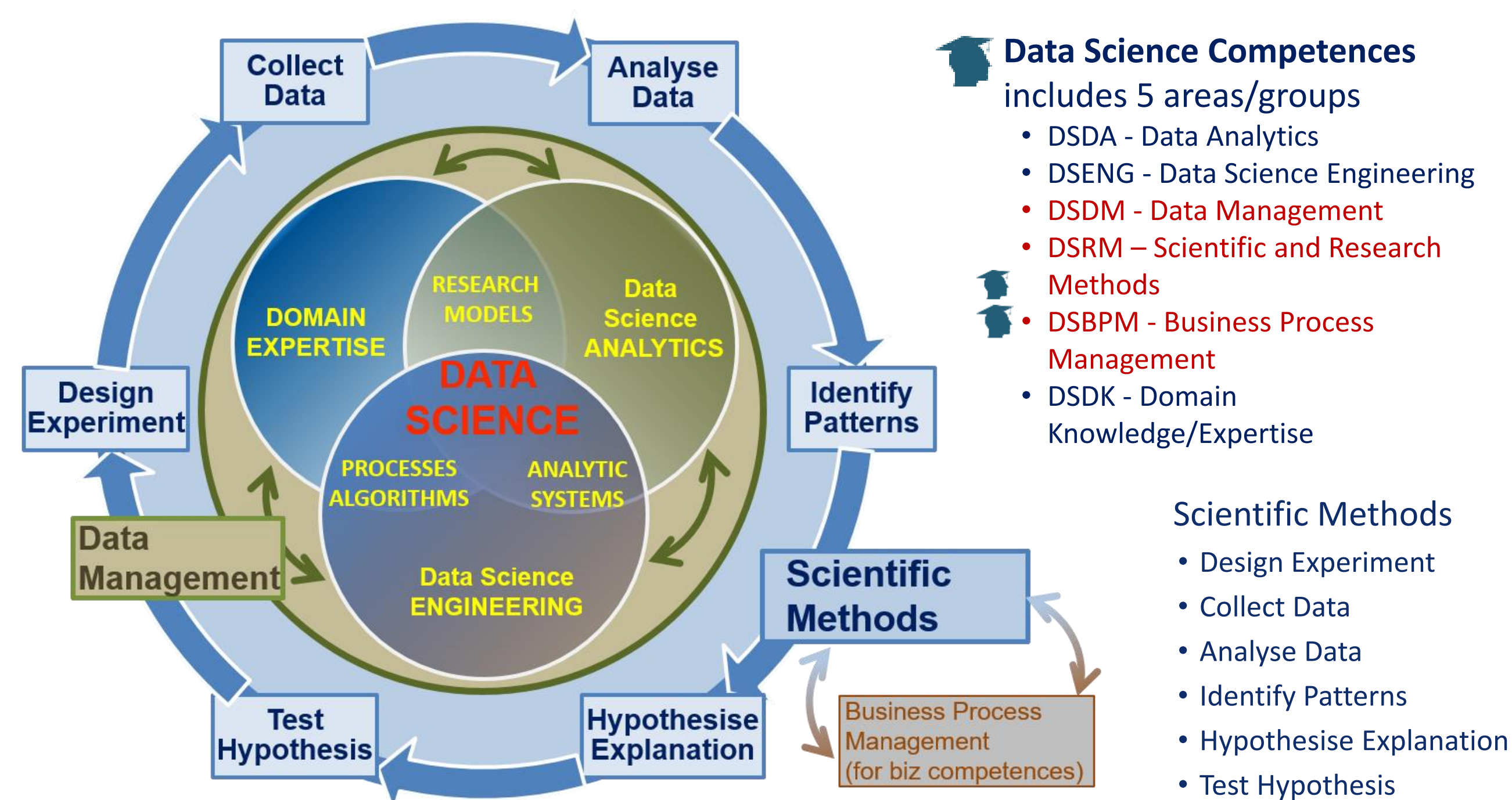
- CF-DS – Data Science Competence Framework
  - DS-BoK – Data Science Body of Knowledge
  - MC-DS – Data Science Model Curriculum
  - DSPP - Data Science Professional Profiles
  - Data Science Taxonomy and Scientific Disciplines Classification
- Service and operational components of the Data Science professional ecosystem
- Community Portal (CP) including
    - EOEE - EDISON Online Education Environment
    - Education and Training Marketplace and Directory
  - Professional certification and career development support

### Outcome Based Educations and Training Model:

From Competences and DSP Profiles to Learning Outcomes (LO) and to Knowledge Unites (KU) and Learning Units (LU)



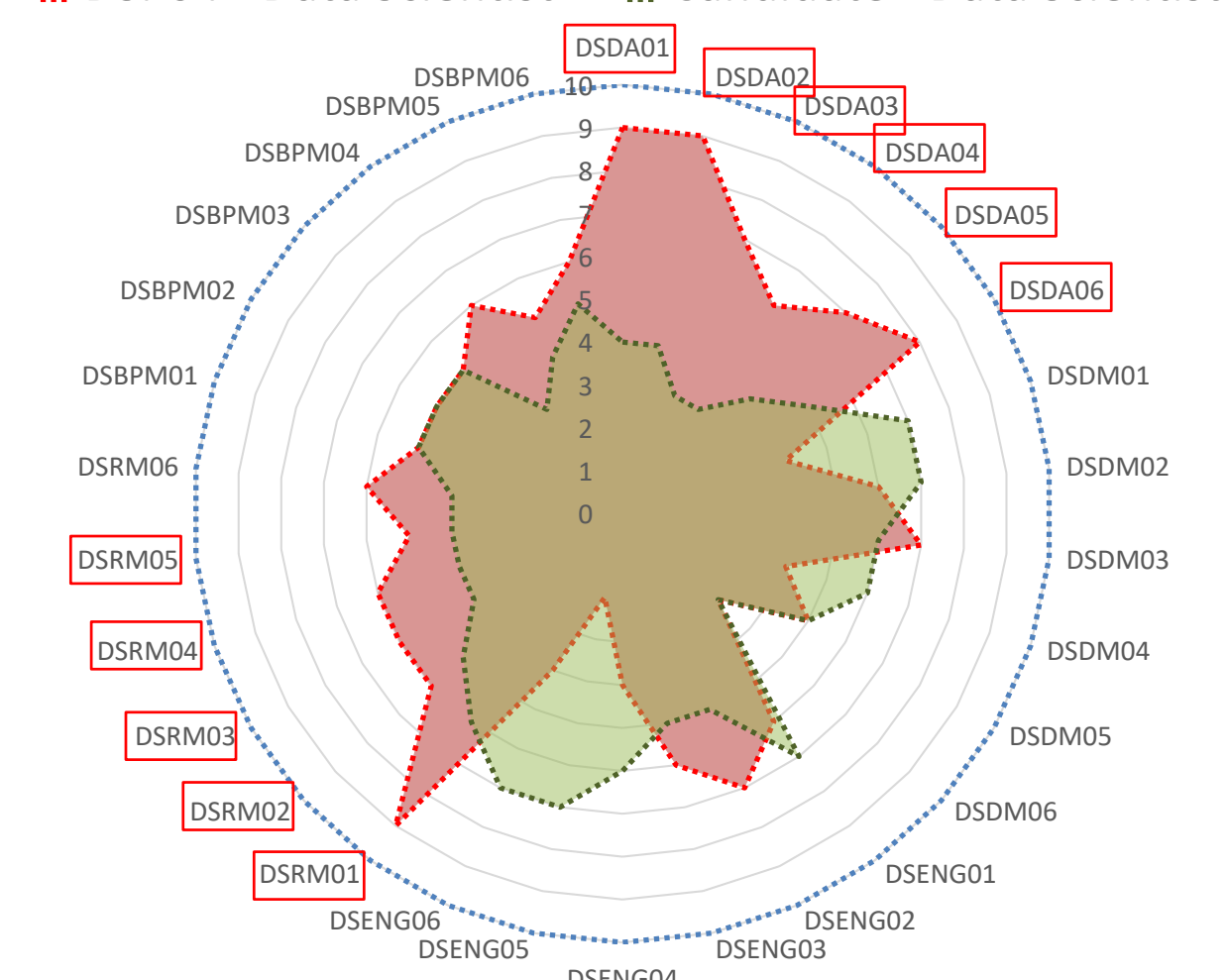
### Data Science Competence Framework (CF-DS)



CF-DS is defined in compliance with the European e-Competence framework for ICT (e-CFv3.0 - 2015)

### MATCHING – COMPETENCE PROFILES

DSP04 - Data Scientist Candidate - Data Scientist



### Individual Education/Training Path based on Competence benchmarking

Red polygon indicates the chosen professional profile: Data Scientist (general)  
Green polygon indicates the candidate or practitioner competences/skills profile

Insufficient competences (gaps) are highlighted in red

- DSDA01 – DSDA06 Data Science Analytics
- DSRM01 – DSRM05 Data Science Research Methods

Can be use for team skills match marking and organisational skills management

[ref] For DSP Profiles definition and for enumerated competences refer to EDSF documents CF-DS and DSP Profiles.

### Data Science Body Of Knowledge (DS-BoK)

#### Data Science Body of Knowledge Area Groups

DS-BoK Knowledge Area Groups (KAG) are defined in compliance with the CF-DS

- 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, Research Data Management, Open Data, Open Access
- KAG4-DSRM: Scientific and Research Methods, use cases/practices
- KAG5-DSBPM: Business Process Management (data driven)

Includes selected KAG and KU defined in ACM CCS (2012), PMI-BoK, BABOK, SWEBOK, DM-BoK,

#### Acknowledgement

EDISON project is supported by the Horizon 2020 Programme Grant number 675419 (INFRA-SUPP-4-2015: CSA).

#### For more information refer to EDISON documents

EDSF - <http://edison-project.eu/edison-data-science-framework-edsf>

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

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

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

DSP Profiles: <http://www.edison-project.eu/data-science-professional-profiles-dsp>



Contact: Yuri Demchenko <y.demchenko@uva.nl>  
Adam Belloum <A.S.Z.Belloum@uva.nl>







# EDISON Project: Coordination and cooperation to establish new profession of Data Scientist for European Research and Industry

## FREQUENTLY ASKED QUESTIONS

### What is the EDISON project?

EDISON is a 2-year EU-funded project that has been set up, starting from September 2015, to pursue the establishment of the data scientist as a new profession in support of e-Infrastructure needs and beyond. Demand for data scientists exists beyond the research infrastructure domain in all areas from research to commerce as well as policy-making and other matters of governance. Understanding the needs of the broader market will guide us in helping define the learning needs for data scientists as well as the additional soft-skills that will equip them for long and fruitful careers.

### What will EDISON deliver?

EDISON has been structured to deliver a collection of specific products that will have a positive and marked influence on the establishment of the data science profession and provide a basis for coordinated work between the demand and supply sides. The core of this collection is three packages that will have captured the key elements of the platform that connects the suppliers of trained data scientists with the organisations that have need of such professionals. These publically available documents will formally define the Data Scientist profession through an **EDISON Data Science Framework (EDSF)** that includes: Data Science Competence Framework (CF-DS) that provides a methodological basis for designing the Data Science curriculum for universities and professional training institutions, Data Science Body of Knowledge (DS-BoK), Data Science Model Curriculum (MC-DS), and Data Science Professional Profiles (DSPP).

### For more information refer to EDISON documents

EDSF - <http://edison-project.eu/edison-data-science-framework-edsf>

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

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

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

DSP Profiles: <http://www.edison-project.eu/data-science-professional-profiles-dsp>

### What is a Data Scientist?

There is no clear definition so far of what a Data Scientist is, but there is a lively debate on the topic and more importantly a lively market of recruiters offering to pay good salaries for what they perceive is a key role for the future. The consensus appears to be that a data scientist role is something between a data analyst, a statistician and a computer scientist but having sufficient domain specific knowledge and possessing of other pioneering scientific research qualities too. Another perspective on this complex role is sometimes referred to as a “T” shaped skill set. By this, it is implied that individuals possess a breadth of skills such as academic curiosity, story telling, product sense, engineering etc. but also deep data analytics and machine learning competences for example. Furthermore, we believe that tomorrow’s Chief Executive Officers will be pooled from these data scientists of today. Such individuals will possess the insight, experience and wisdom to lead the major enterprises of tomorrow in an increasingly data-centric world.

EDISON proposed more formal definition of the Data Scientist based on the NIST Big Data WG (NIST SP1500 - 2015) extended with the essential role of the Data Scientist to bring the added value to science or business:

*A **Data Scientist** is a practitioner who has sufficient knowledge in the overlapping regimes of expertise in business needs, domain knowledge, analytical skills, and programming and systems engineering expertise to manage the end-to-end scientific method process through each stage in the **big data lifecycle**, till the delivery of expected **scientific and business value** to science or industry.*

### Why do we need EDISON if there is such a demand for Data Scientists?

EDISON has been created because we feel there is a clear mismatch between the supply side and the demand side of the data scientist pipeline. Moreover, as the demand for what is understood to be data scientist grows more rapidly, this gap becomes ever wider. We believe that this model is more complex but yet addressable. The two-sides of the marketplace – education and training suppliers – and research and commercial employers on the demand side – can be much better understood and supported, and in doing so, EDISON can play a role in supporting a network or platform that helps orchestrate the players in order to maximise the matching of data scientists to jobs.

### What do you mean by building the Data Scientist profession?

For all of the above efforts and activities to really hit home we believe that EDISON has also to support the process of endorsing the courses and certificates that individuals need and accrue in their quest to become the key professionals of tomorrow. This will also involve identifying and collaborating with the bodies and institutions that will endorse and ratify the professionals as they build their status.

### What other services will EDISON provide to assist with the process of furthering the Data Science profession?

The project will establish the Community Portal that will include an Education and Training platform and an Education and Training Marketplace or Directory of the educational and training resources. The purpose of these is to support Universities in setting up new courses and to this end virtual labs with specially provided datasets to be made available.

All EDISON services will be available via the **Data Science Community Portal** (<http://datasciencepro.eu>)

### How can I get involved?

We are very keen to gather input, feedback and ultimately take-up and endorsement from those across Europe and beyond, with interests in the establishment of the data scientist profession. These will range from workshops and talks, to questionnaires and surveys, to Liaison groups and Champion Universities Conferences – See EDISON Community events.

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

## EDISON Community Engagement Activities

EDISON is hosting a programme of events across Europe that will appeal to different segments of the data science community. Many are co-located with events in the Research Infrastructure calendar:

- EDISON Workshop on Data Infrastructure Competences and Skills Framework: a European and Global Challenge, 9 February 2016, Brussels
- **EDISON Champions Conference** for Universities and Training organisations: first took place on 13-14 July in Southampton, UK; forthcoming 15-16 March 2017 in Madrid; planned for Summer 2017 in Warsaw.
- Regular **EDISON Liaison Group (ELG)** meeting
- RDA Interest Group on Education and Training on Handling Research Data Handling (IG-ETRD) is the EDISON RDA community liaison venue

## Data Science Community Portal (<http://datasciencepro.eu>)

- One-stop one-step shop for becoming Data Science Professional and engaging with the Data Science community
- Data Science Benchmarking and Learning Path Recommendation System
- Clearinghouse of educational and training materials and Virtual Labs powered by EGI Fed

