EDISON Data Science Framework (EDSF) and Education and training for Data Science and data related competences

EDISON
building the data science profession

Yuri Demchenko, EDISON Project
University of Amsterdam

DI4R Conference
30 November 2017, Amsterdam

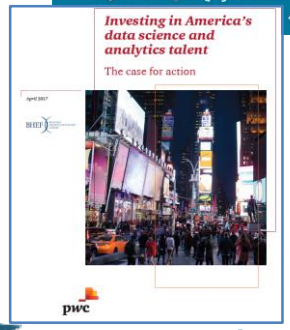
EDISON – **E**ducation for **D**ata **I**ntensive **S**cience to **O**pen **N**ew science frontiers

Grant 675419 (INFRASUPP-4-2015: CSA)

Industry reports on Data Science Analytics and Data enabled skills demand

- Final Report on European Data Market Study by IDC (Feb 2017)
  - Number of data workers 6.1 mln (2016) - increase 2.6% from 2015
    - Estimated EUR 10.4 million in 2020
  - Average number of data workers per company 9.5 - increase 4.4%
  - Gap between demand and supply estimated 769,000 (2020) or 9.8%

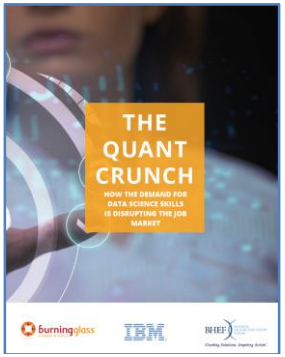
- HLEG report on European Open Science Cloud (October 2016)
  - Demand for 80 K core data experts and data stewards



- PwC and BHEF report “Investing in America’s data science and analytics talent: The case for action” (April 2017)
  - 2.35 mln postings, 23% Data Scientist, 67% DSA enabled jobs
  - DSA enabled jobs growing at higher rate than main Data Science jobs

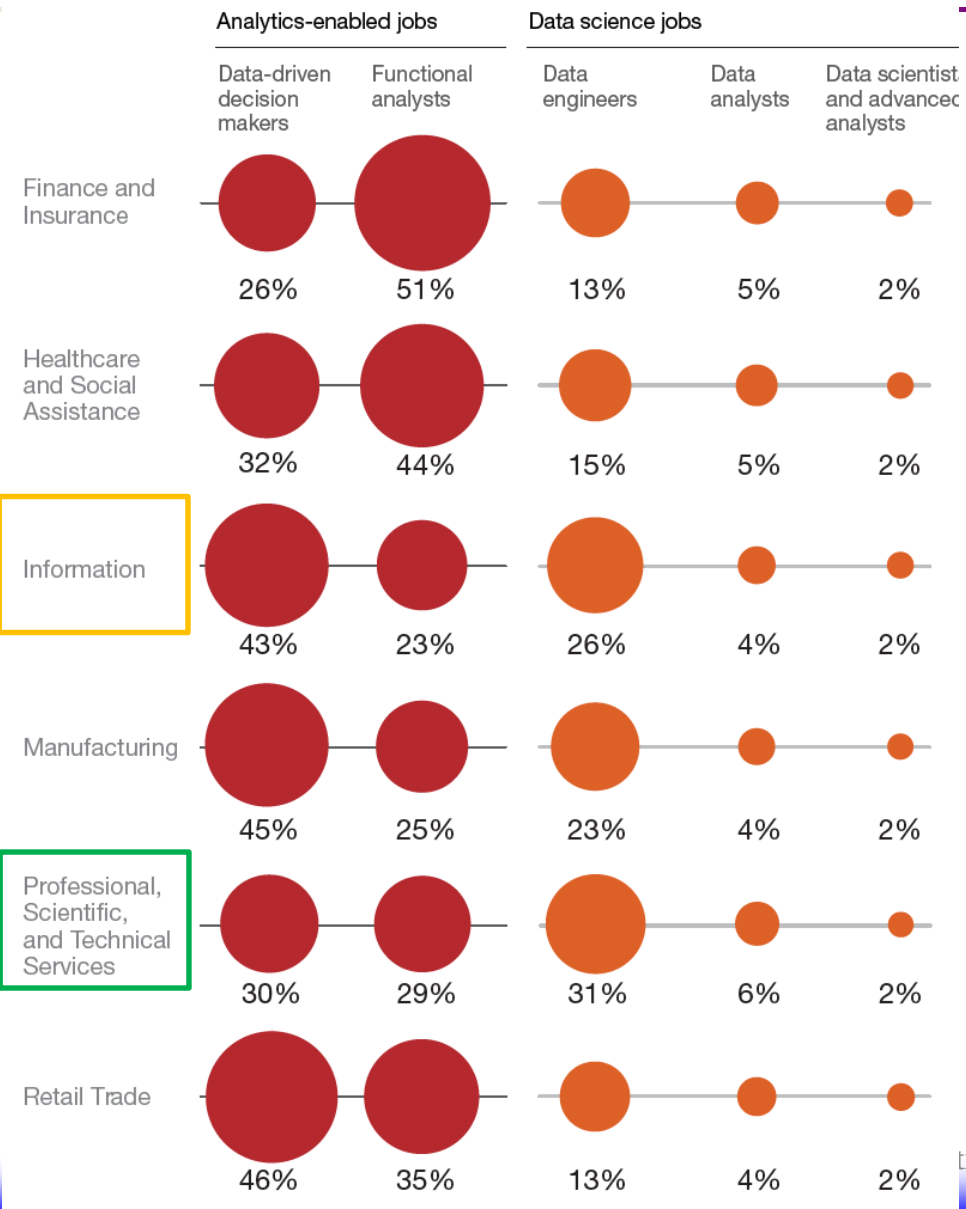
Citing EDISON and EDSF

- Burning Glass Technology, IBM, and BHEF report “The Quant Crunch: How the demand for Data Science Skills is disrupting the job Market” (April 2017)
  - DSA enabled jobs takes 45-58 days to fill: 5 days longer than average
  - Commonly required work experience 3-5 yrs



Influenced by EDISON

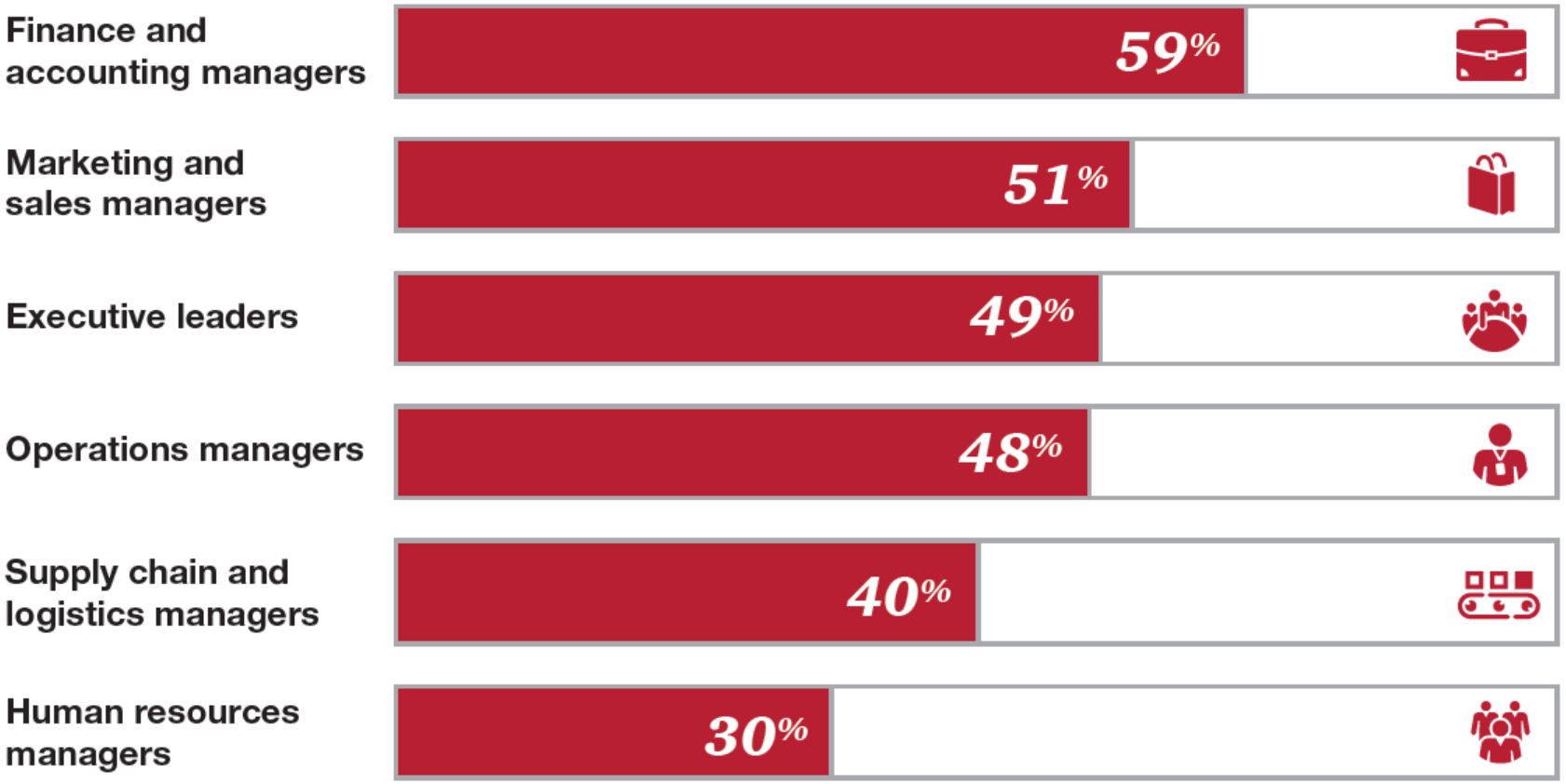
PwC&BHEF: Demand for DSA enabled jobs



Demand for business people with analytics skills, not just data scientists

- Of 2.35 million job postings in the US
  - 23% Data Scientist
  - 67% DSA enabled jobs
- Strong demand for managers and decision makers with Data Science (data analytics) skills/understanding
  - Challenge to deliver actionable knowledge and competences to CEO level managers

PwC&BHEF: Data Science and Data Analytics Competences for Managers and Decision Makers

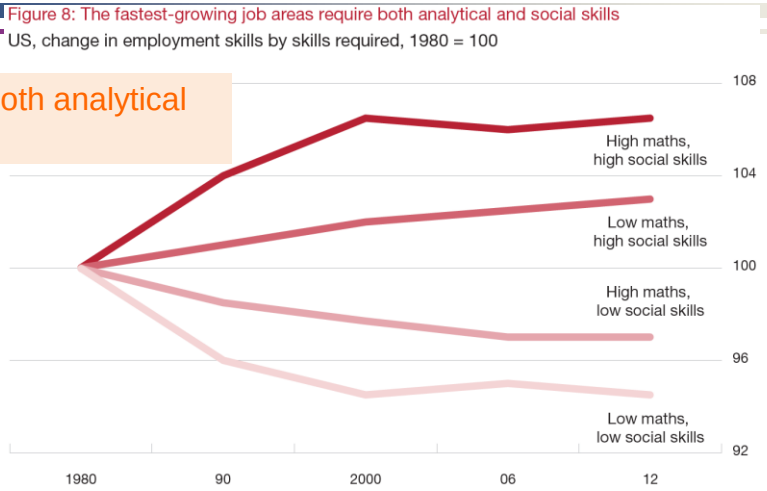


Percent of employers who say data science and analytics skills will be 'required of all managers' by 2020

- Source: BHEF and Gallup, *Data Science and Analytics Business Survey* (December 2016).

PwC&BHEF: Skills that are tough to find

Faster growing jobs require both analytical and social skills



97% Cybersecurity

95% Data science and analytics

83% Critical thinking and problem solving

79% Design/ systems thinking

79% Innovation and creativity

78% Global perspective

78% Cognitive flexibility

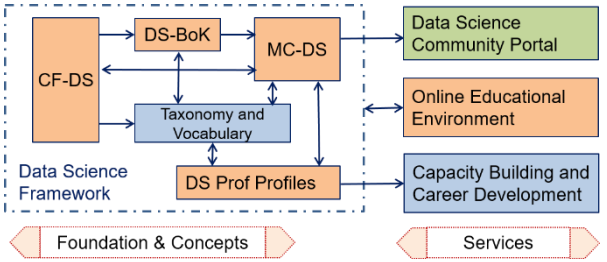
74% Cross-disciplinary ability

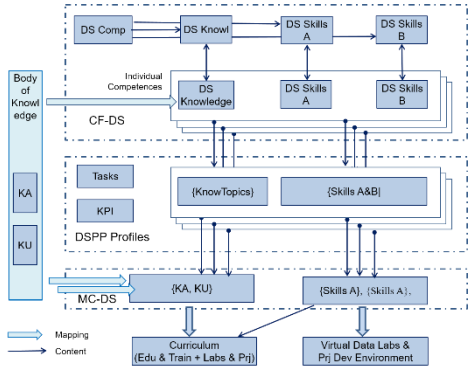
To be mapped to Competences, Knowledge, Skills and Personal (soft) Skills

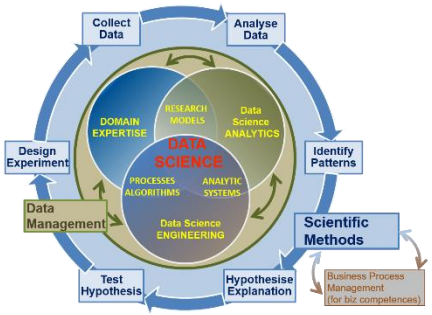
Source: Business Roundtable (2017).

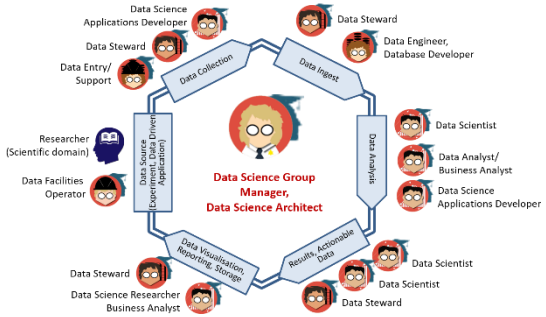
EDISON Products for Data Science Skills Management and Curriculum Design

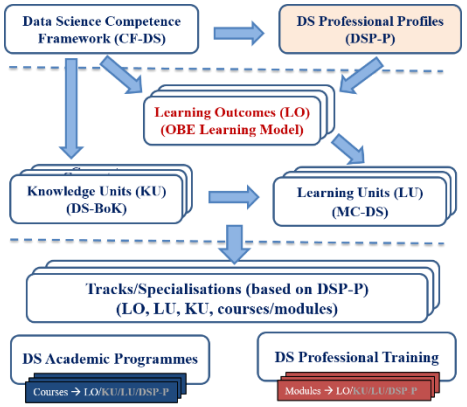
- EDISON Data Science Framework (EDSF)
  - Compliant with EU standards on competences and professional occupations e-CFv3.0, ESCO
  - Customisable courses design for targeted education and training
- Skills development and career management for Core Data Experts and related data handling professions
- Capacity building and Data Science team design
- Academic programmes and professional training courses (self) assessment and design
- Cooperation with International professional organisations IEEE, ACM, BHEF, APEC (AP Economic Cooperation)



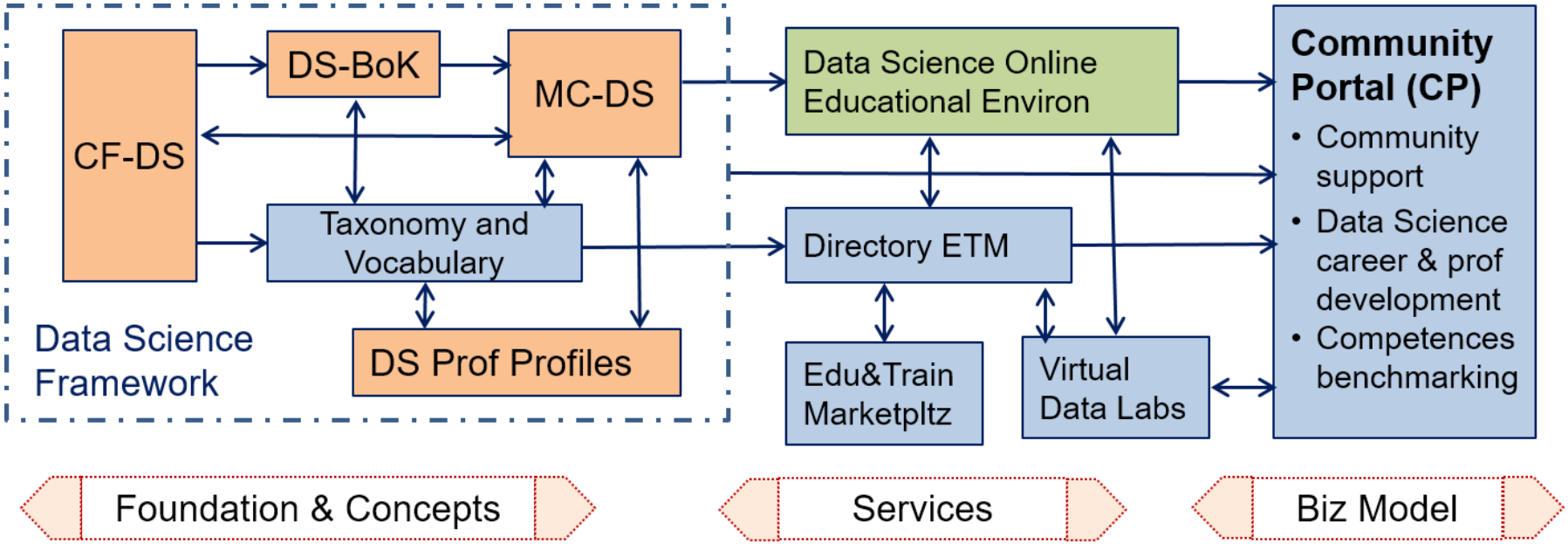








EDISON Data Science Framework (EDSF)



EDISON Framework components

- CF-DS – Data Science Competence Framework
- DS-BoK – Data Science Body of Knowledge
- MC-DS – Data Science Model Curriculum
- DSP – Data Science Professional profiles
- Data Science Taxonomies and Scientific Disciplines Classification
- EOEE - EDISON Online Education Environment

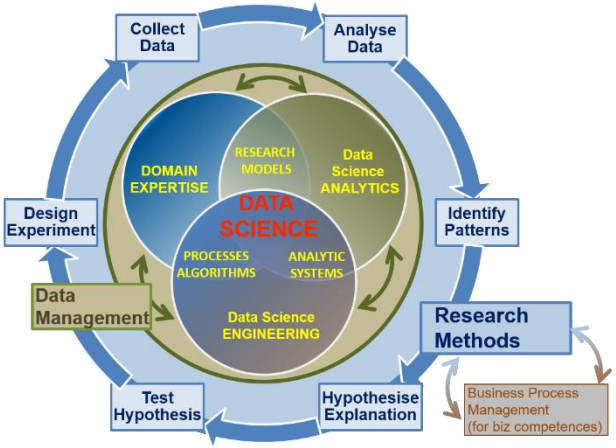
Methodology

- ESDF development based on job market study, existing practices in academic, research and industry.
- Review and feedback from the ELG, expert community, domain experts.
- Input from the champion universities and community of practice.

Data Scientist definition

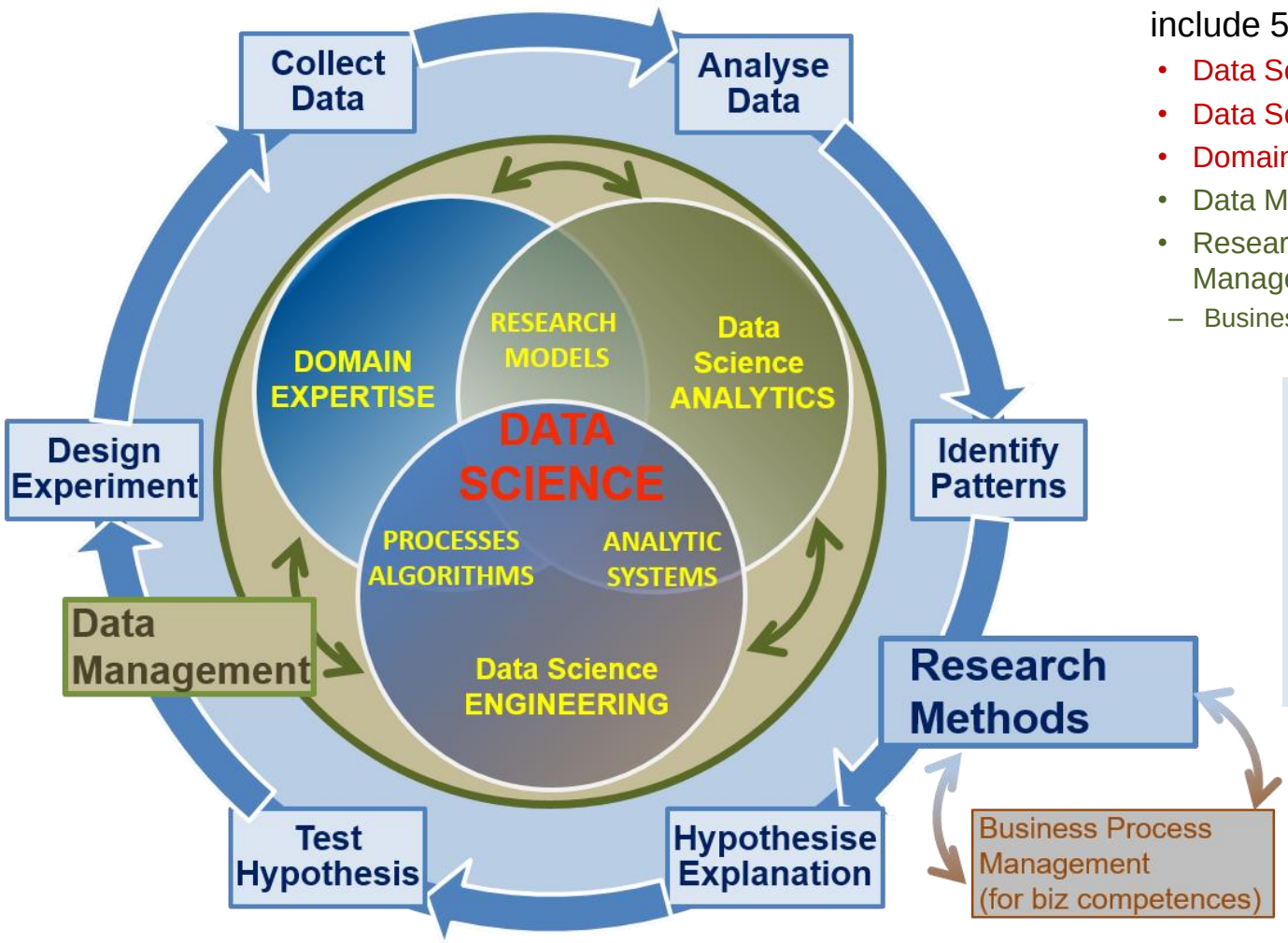
Based on the definitions by NIST SP1500 – 2015, extended by EDISON

- *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 an **expected scientific and business value** to organisation or project.*



- Core Data Science competences and skills groups
  - **Data Science Analytics** (including Statistical Analysis, Machine Learning, Business Analytics)
  - **Data Science Engineering** (including Software and Applications Engineering, Data Warehousing, Big Data Infrastructure and Tools)
  - **Domain Knowledge and Expertise** (Subject/Scientific domain related)
- EDISON identified 2 additional competence groups demanded by organisations
  - **Data Management, Data Governance, Stewardship, Curation, Preservation**
  - **Research Methods and/vs Business Processes/Operations**
- **Data Science professional skills**: Thinking and acting like Data Scientist – required to successfully develop as a Data Scientist and work in Data Science teams

Data Science Competence Groups - Research



Data Science Competences include 5 groups

- Data Science Analytics
- Data Science Engineering
- Domain Knowledge and Expertise
- Data Management
- Research Methods and Project Management
  - Business Process Management (biz)

Scientific Methods

- Design Experiment
- Collect Data
- Analyse Data
- Identify Patterns
- Hypothesis Explanation
- Test Hypothesis

Business Operations

- Operations Strategy
- Plan
- Design & Deploy
- Monitor & Control
- Improve & Re-design

Identified Data Science *Skills/Experience* Groups

Skills Type A – Based on knowledge acquired

- **Group 1: Skills/experience related to competences**
  - Data Analytics and Machine Learning
  - Data Management/Curation (including both general data management and scientific data management)
  - Data Science Engineering (hardware and software) skills
  - Scientific/Research Methods or Business Process Management
  - Application/subject domain related (research or business)
- **Group 2: Mathematics and statistics**
  - Mathematics and Statistics and others

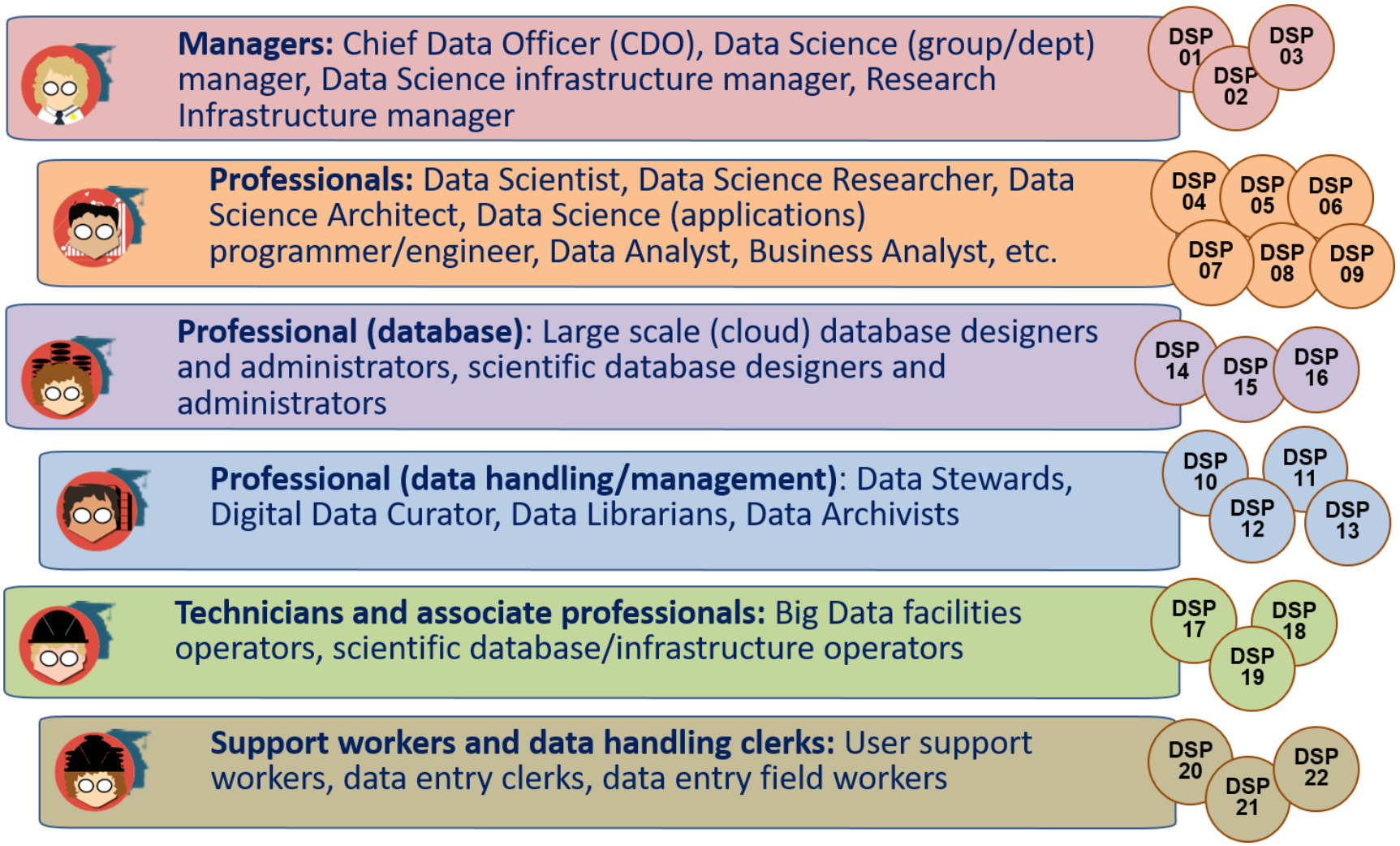
Skills Type B – Base on practical or workplace experience

- **Group 3: Big Data (Data Science) tools and platforms**
  - Big Data Analytics platforms
  - Mathematics & Statistics applications & tools
  - Databases (SQL and NoSQL)
  - Data Management and Curation platform
  - Data and applications visualisation
  - ***Cloud based platforms and tools***
- **Group 4: Data analytics programming languages and IDE**
  - General and specialized development platforms for data analysis and statistics
- **Group 5: Soft skills and Workplace skills**
  - Data Science professional skills: Thinking and Acting like Data Scientist
  - 21st Century Skills: Personal, inter-personal communication, team work, professional network

Data Science Competence Framework (CF-DS): Practical Application

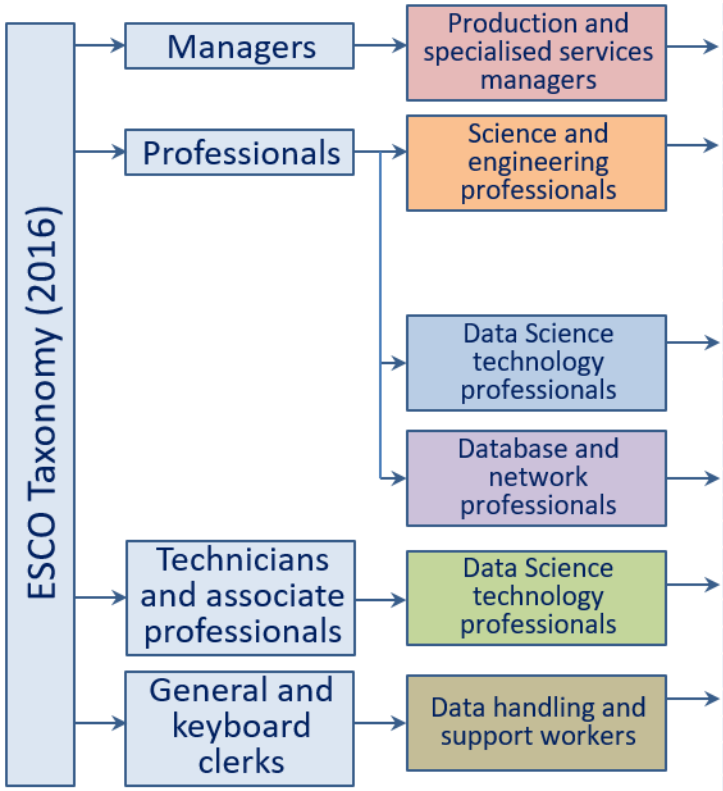
- Basis for the definition of the Data Science Body of Knowledge (DS-BoK) and Data Science Model Curriculum (MC-DS)
  - CF-DS => Learning Outcomes (MC-DS) => Knowledge Areas (DS-BoK)
  - CF-DS => Data Science taxonomy of scientific subjects and vocabulary
- Data Science professional profiles definition
  - Extend existing EU standards and occupations taxonomies: e-CFv3.0, ESCO, others
- Professional competence benchmarking
  - For customizable training and career development
  - Including CV or organisational profiles matching
- Professional certification
  - In combination with DS-BoK professional competences benchmarking
- Vacancy construction tool for job advertisement (for HR)
  - Using controlled vocabulary and Data Science Taxonomy

Data Science Professions Family (DSPP)



Icons used: Credit to [ref] https://www.datacamp.com/community/tutorials/data-science-industry-infographic

DSP Profiles mapping to ESCO Taxonomy High Level Groups



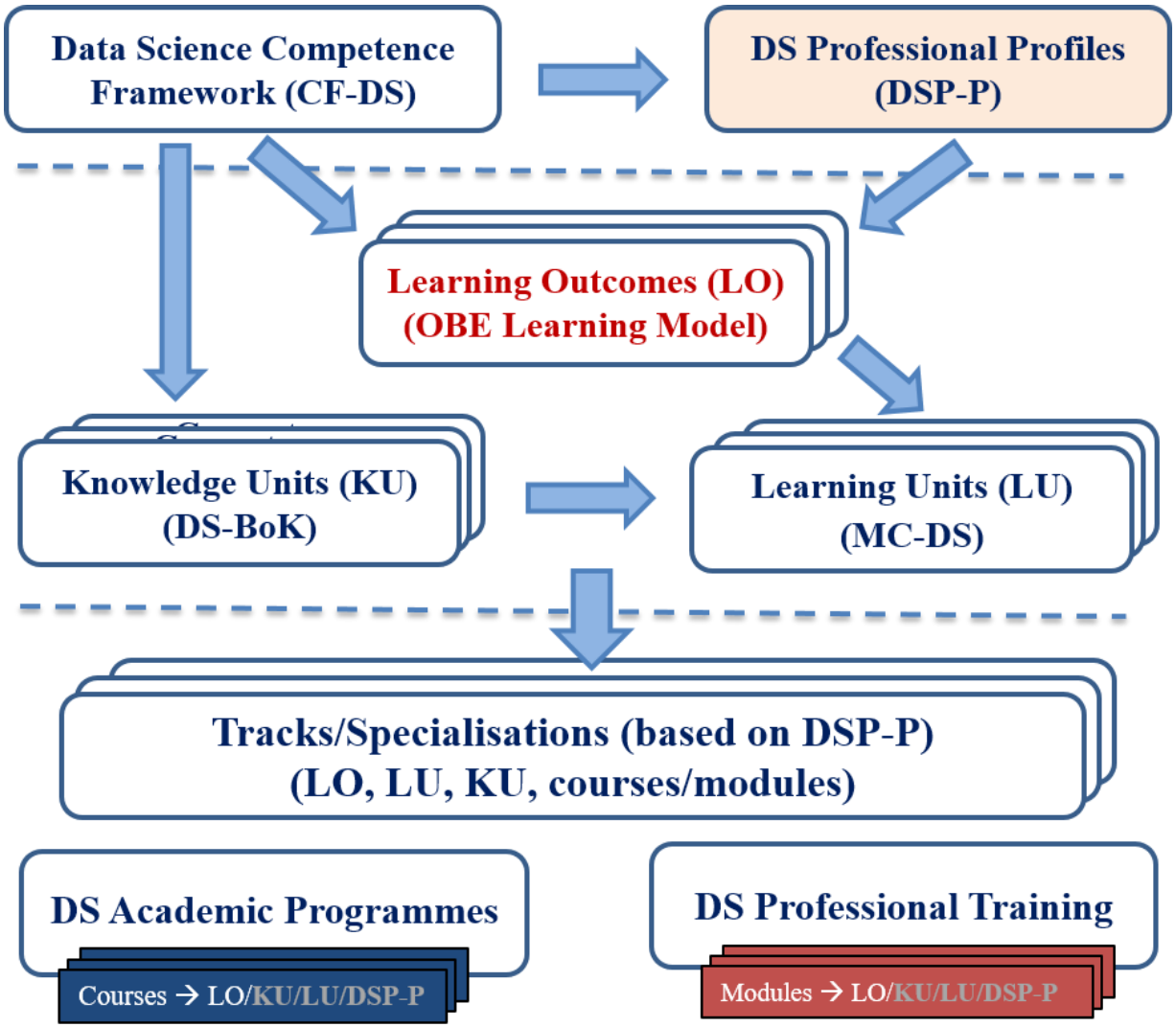
Profile ID	Data Science Profile title	DSDA	DSDM	DSENG	DSRM	DSDK
Data Science Services/Infrastructure Managers						
DSP01	Data Science (group) Manager	3	4	3	3	2
DSP02	Data Science Infrastr Manager	2	4	4	2	2
DSP03	Research Infrastructure Manager	2	4	4	3	2
Data Scince Professionals						
DSP04	Data Scientist	5	3	4	5	3
DSP05	Data Science Researcher	4	3	2	5	4
DSP06	Data Science Architect	4	3	5	3	3
DSP07	Data Science Applic Programmer	4	2	5	3	4
DSP08	Data Analyst	5	3	3	3	4
DSP09	Business Analyst	5	3	3	4	5
Data handling professionals not elsewhere classified						
DSP10	Data Stewards	3	5	3	3	3
DSP11	Digital data curator	1	5	2	2	3
DSP12	Digital Librarians	2	5	2	2	3
DSP13	Data Archivists	1	5	1	1	3
Database and network professionals not elsewhere classified						
DSP14	Large scale database designer	2	4	4	3	3
DSP15	Large scale database admin	2	4	3	2	3
DSP16	Scientific database administrator	2	4	3	2	3
Data Infrastructure engineers and technicians						
DSP17	Big Data facilities Operator	1	4	4	2	3
DSP18	Large scale data storage operator	1	4	3	1	1
DSP19	Scientific database operator	1	4	3	2	3
Data and information entry and access						
DSP20	Data entry/access worker		2	1		2
DSP21	Data entry field workers		2	1		2
DSP22	User support data services		3	2		2

- DSP Profiles mapping to corresponding CF-DS Competence Groups
  - Competences relevance level from 5 – maximum to 1 – minimum

EDSF for Education and Training

- Foundation and methodological base
  - Data Science Body of Knowledge (DS-BoK)
    - Taxonomy and classification of Data Science related scientific subjects
  - Data Science Model Curriculum (MC-DS)
    - Set Learning Units mapped to CF-DS Learning and DS-BoK Knowledge Areas/Units
  - Instructional methodologies and teaching models
- Platforms and environment
  - Virtual labs, datasets, developments platforms
  - Online education environment and courses management
- Services
  - Individual benchmarking and profiling tools (competence assessment)
  - Knowledge evaluation tools
  - Certifications and training for self-made Data Scientists practitioners
  - Education and training marketplace: Courses catalog and repository

Outcome Based Educations and Training Model: Customised curriculum design approach



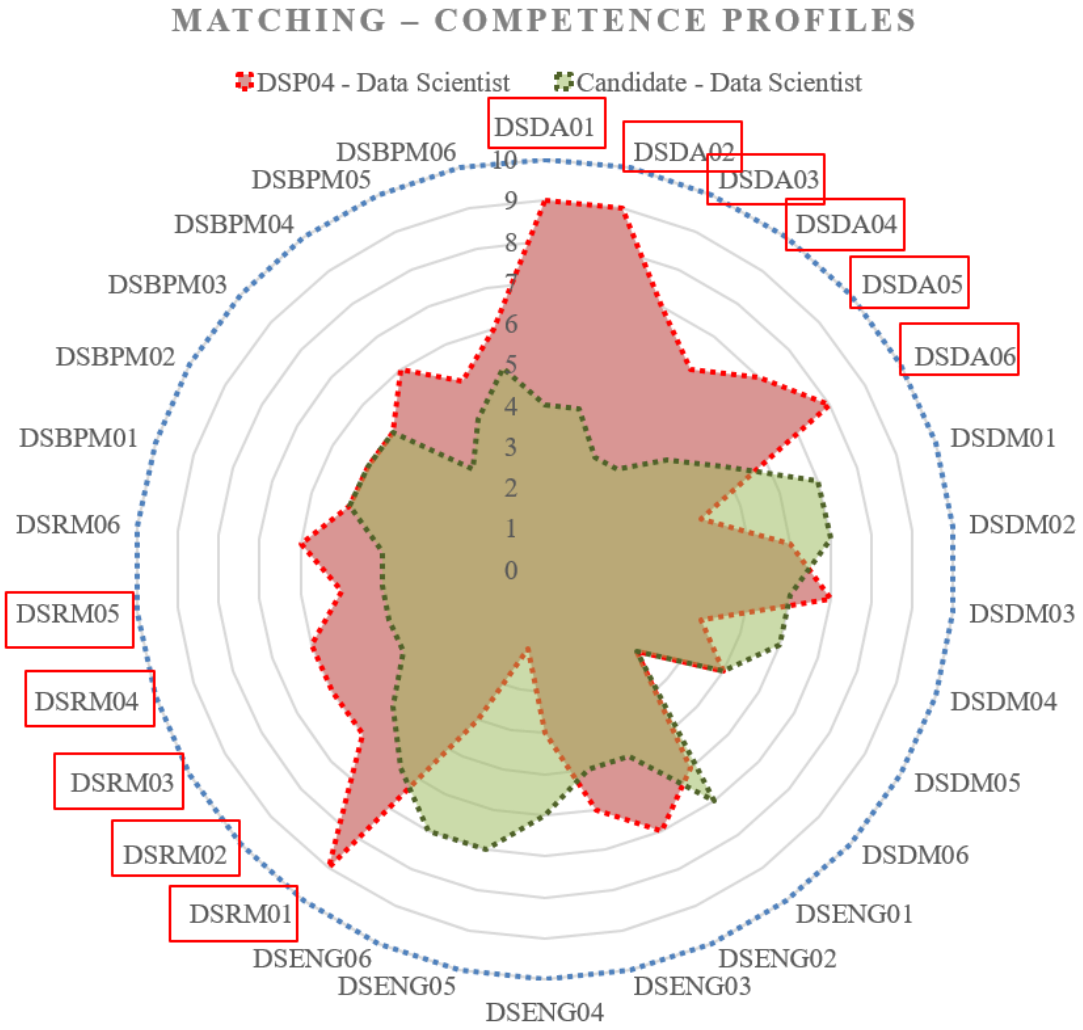
From Competences and DSP Profiles

to Learning Outcomes (LO) and

to Knowledge Unites (KU) and Learning Units (LU)

- EDSF allow for customized educational courses and training modules design

Individual Competences Benchmarking

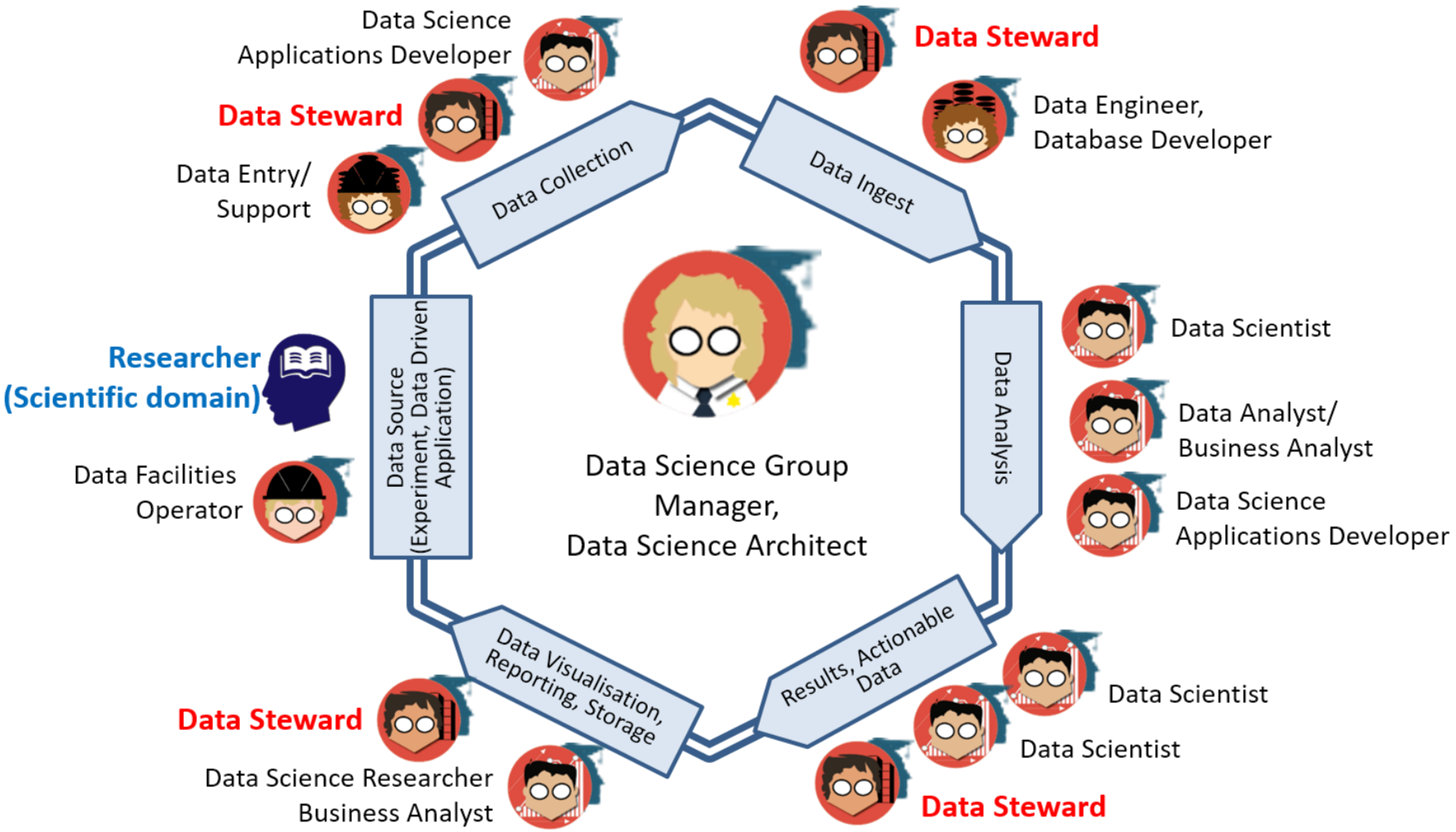


Individual Education/Training Path based on Competence benchmarking

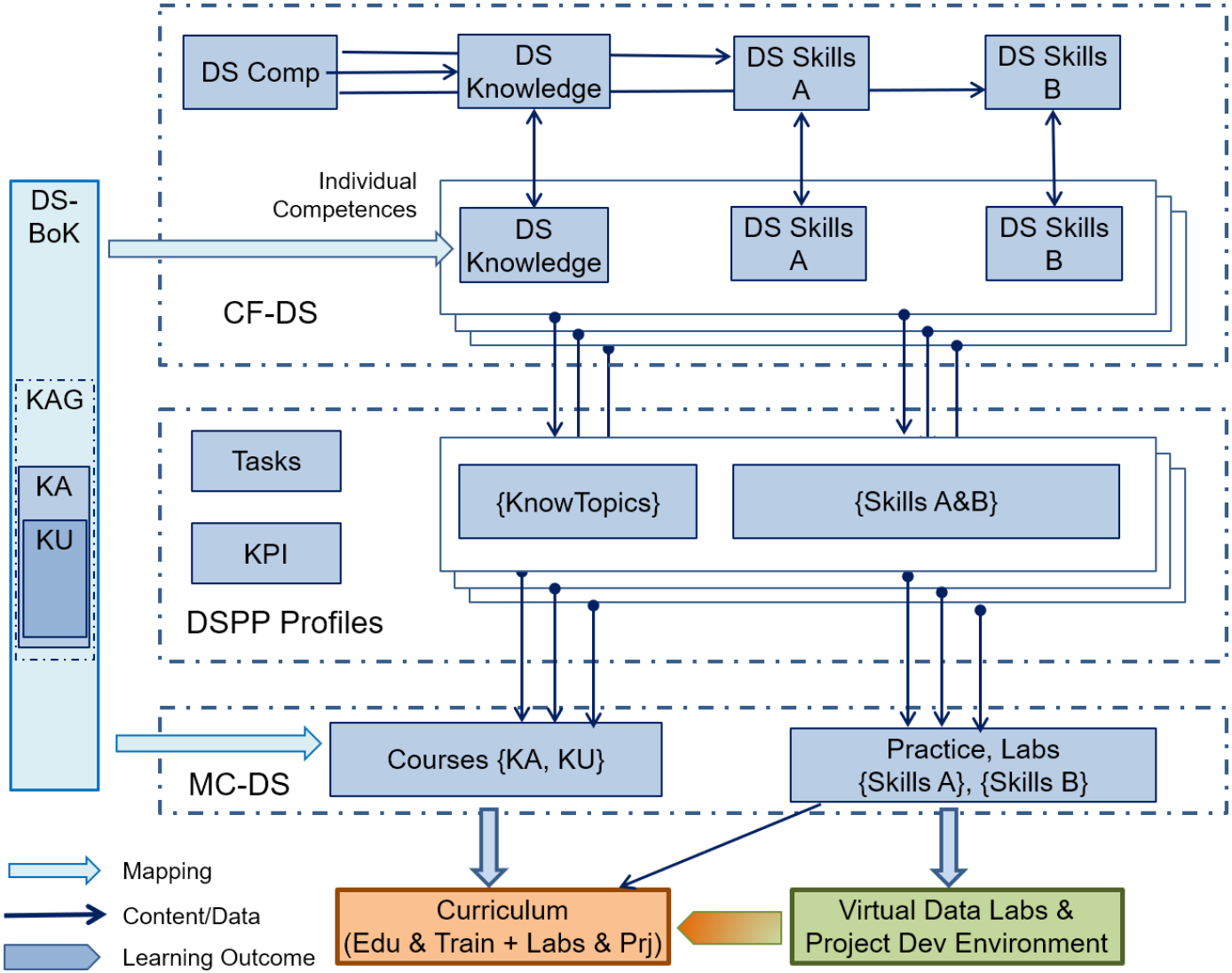
- Red polygon indicates the chosen professional profile: Data Scientist (general)
- Green polygon indicates the candidate or practitioner competences/skills profile
- Insufficient competences (gaps) are highlighted in *red*
  - *DSDA01 – DSDA06 Data Science Analytics*
  - *DSRM01 – DSRM05 Data Science 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.

Building a Data Science Team

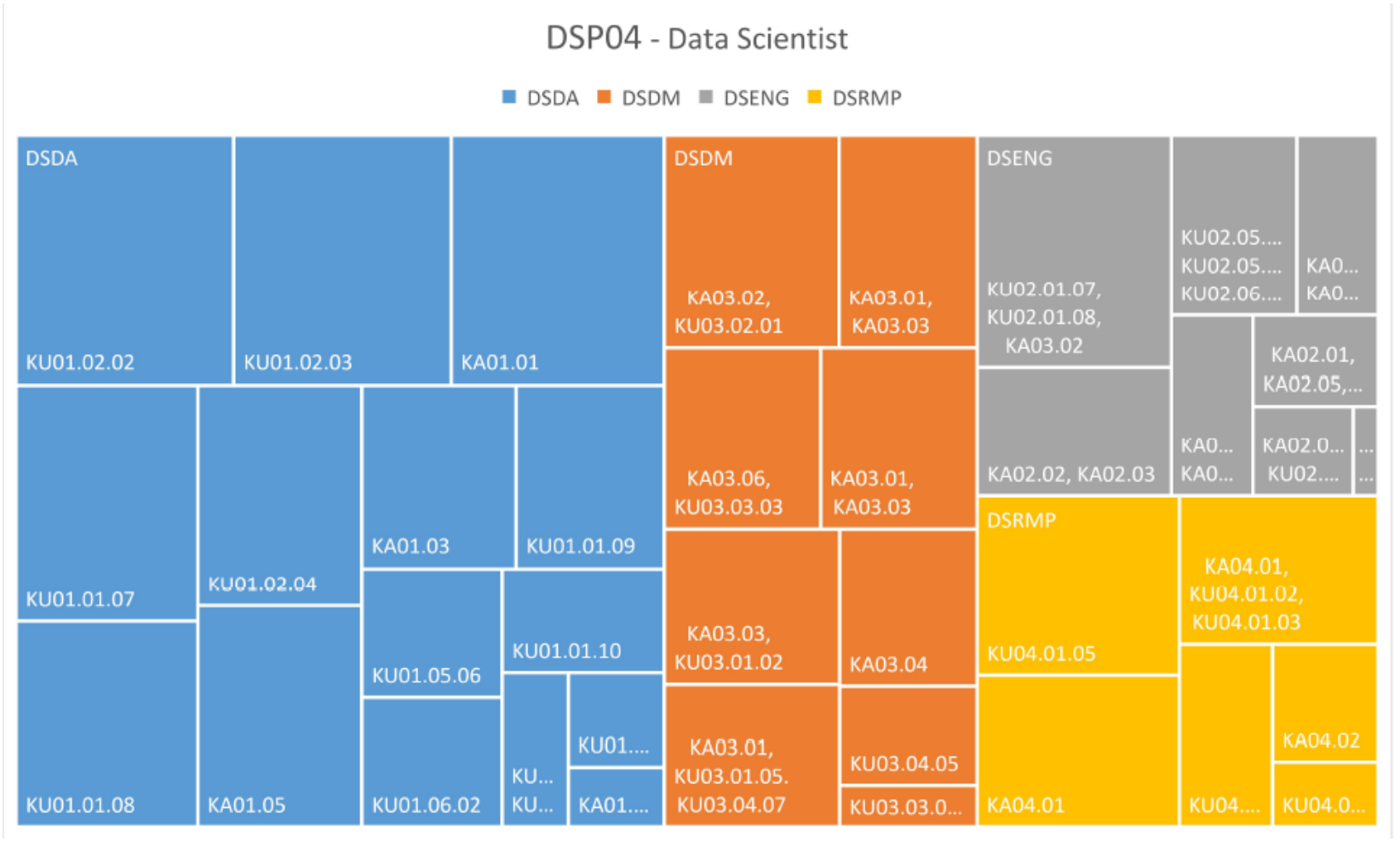


EDSF Data Model and API

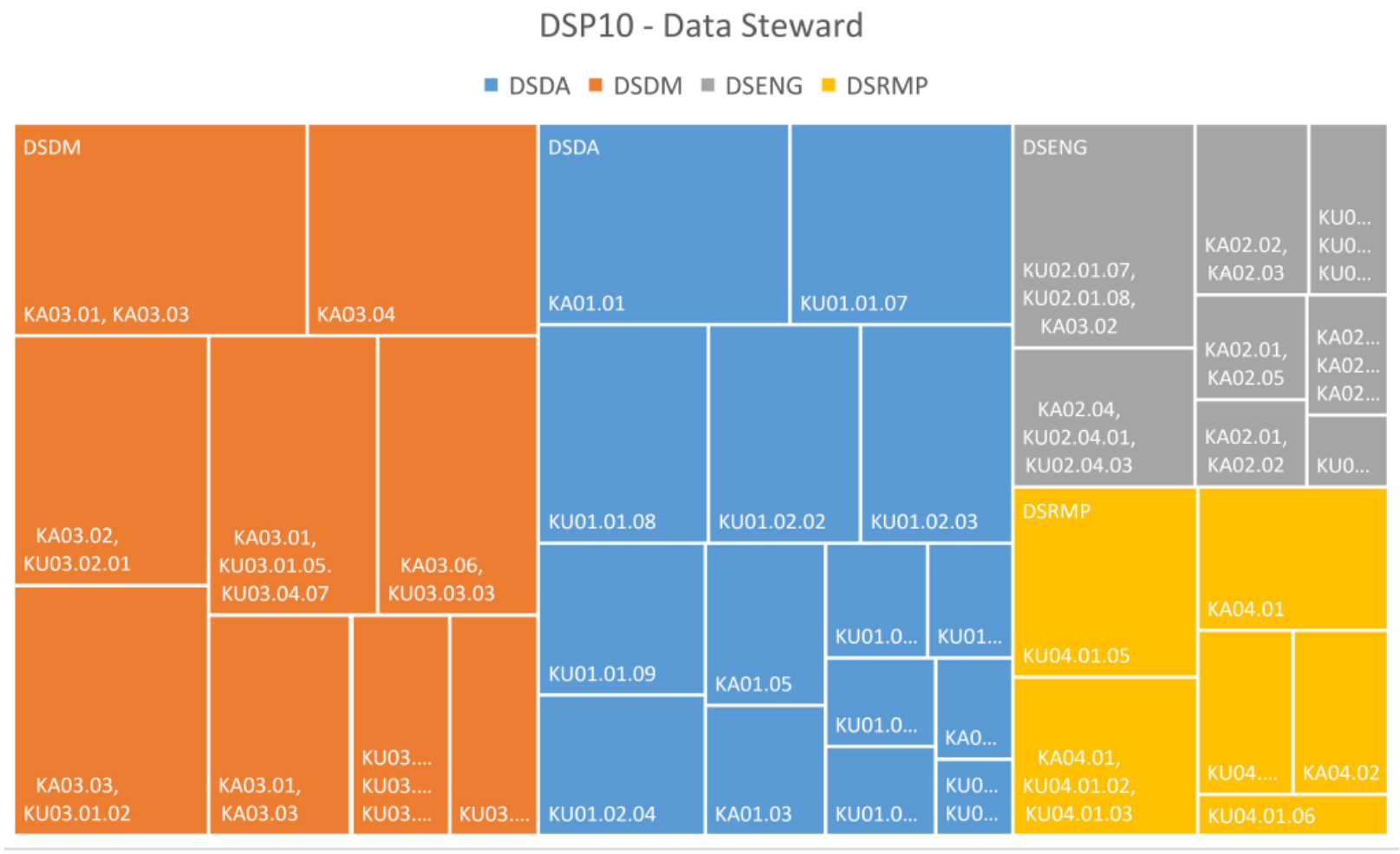


- EDSF API provides access to all EDSF functionality
- Ontology and controlled vocabulary
- Course/curriculum design
- Virtual Labs recipes/config

Example (1): DSP04 – Data Scientist MC structure



Example (2): DSP10 – Data Steward MC structure



Data Science Professional Skills:
Thinking and Acting like Data Scientist

1. **Recognise value of data**, work with raw data, exercise good data intuition, use SN and open data
2. Accept (be ready for) **iterative development**, know when to stop, comfortable with failure, accept the symmetry of outcome (both positive and negative results are valuable)
3. Good **sense of metrics**, understand importance of the results validation, never stop looking at individual examples
4. **Ask the right questions**
5. **Respect domain/subject matter knowledge** in the area of data science
6. **Data driven problem solver** and **impact-driven mindset**
7. **Be aware about power and limitations** of the main machine learning and data analytics algorithms and tools
8. Understand that most of **data analytics algorithms are statistics and probability based**, so any answer or solution has some degree of probability and represent an optimal solution for a number variables and factors
9. Recognise what things are **important** and what things are **not important** (in data modeling)
10. Working in **agile environment** and coordinate with other roles and team members
11. Work in **multi-disciplinary team**, ability to communicate with the domain and subject matter experts
12. Embrace **online learning**, continuously improve your knowledge, use **professional netw**orks and communities
13. **Story Telling**: Deliver actionable result of your analysis
14. **Attitude**: Creativity, curiosity (willingness to challenge status quo), commitment in finding new knowledge and progress to completion
15. **Ethics and responsible use** of data and insight delivered, awareness of dependability (data scientist is a feedback loop in data driven companies)

21st Century Skills (DARE & BHEF & EDISON)

1. **Critical Thinking:** Demonstrating the ability to apply critical thinking skills to solve problems and make effective decisions
2. **Communication:** Understanding and communicating ideas
3. **Collaboration:** Working with other, appreciation of multicultural difference
4. **Creativity and Attitude:** Deliver high quality work and focus on final result, intitiative, intellectual risk
5. **Planning & Organizing:** Planning and prioritizing work to manage time effectively and accomplish assigned tasks
6. **Business Fundamentals:** Having fundamental knowledge of the organization and the industry
7. **Customer Focus:** Actively look for ways to identify market demands and meet customer or client needs
8. **Working with Tools & Technology:** Selecting, using, and maintaining tools and technology to facilitate work activity
9. **Dynamic (self-) re-skilling:** Continuously monitor individual knowledge and skills as shared responsibility between employer and employee, ability to adopt to changes
10. **Professional networking:** Involvement and contribution to professional network activities
11. **Ethics:** Adhere to high ethical and professional norms, responsible use of power data driven technologies, avoid and disregard un-ethical use of technologies and biased data collection and presentation

Questions and discussion

Links to EDISON Resources

- EDISON project website http://edison-project.eu/
- EDISON slides deck
  http://www.uazone.org/demch/presentations/edson2017-10-slides-deck-v03-extended.pdf
- EDISON Data Science Framework Release 1 (EDSF)
  http://edison-project.eu/edison-data-science-framework-edsf
  - Data Science Competence Framework
    http://edison-project.eu/data-science-competence-framework-cf-ds
  - Data Science Body of Knowledge
    http://edison-project.eu/data-science-body-knowledge-ds-bok
  - Data Science Model Curriculum
    http://edison-project.eu/data-science-model-curriculum-mc-ds
  - Data Science Professional Profiles
    http://edison-project.eu/data-science-professional-profiles-definition-dsp

Other related links

- Amsterdam School of Data Science
  - https://www.schoolofdatascience.amsterdam/
  - https://www.schoolofdatascience.amsterdam/education/
- Research Data Alliance interest Group on Education and Training on Handling of Research Data (IG-ETHRD)
  - https://www.rd-alliance.org/groups/education-and-training-handling-research-data.html
- Final Report on European Data Market Study by IDC (Feb 2017)
  - https://ec.europa.eu/digital-single-market/en/news/final-results-european-data-market-study-measuring-size-and-trends-eu-data-economy
- PwC and BHEF report “Investing in America’s data science and analytics talent: The case for action” (April 2017)
  - http://www.bhef.com/publications/investing-americas-data-science-and-analytics-talent
- Burning Glass Technology, IBM, and BHEF report “The Quant Crunch: How the demand for Data Science Skills is disrupting the job Market” (April 2017)
  - http://www.bhef.com/publications/quant-crunch-how-demand-data-science-skills-disrupting-job-market
  - https://public.dhe.ibm.com/common/ssi/ecm/im/en/iml14576usen/IML14576USEN.PDF
- Millennials at work: Reshaping the workspace (2016)
  - https://www.pwc.com/m1/en/services/consulting/documents/millennials-at-work.pdf