EDISON Data Science Framework (EDSF)

Data Science Professional Education and Training

EDISON Project value proposition and legacy

Yuri Demchenko, EDISON Project
University of Amsterdam

EDISON Initiative, January 2023

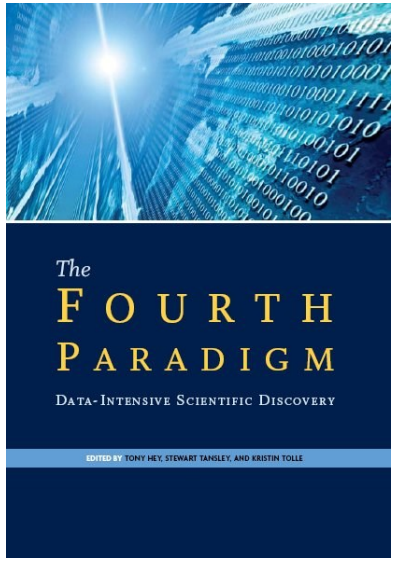
EDISON
building the data science profession

EDISON Project (2015-2017)
Grant 675419 (INFRASUPP-4-2015: CSA)

Outline

- Background: Data driven research and demand for new skills
  - Foundation, recent reports, studies and facts
- EDISON Data Science Framework (EDSF)
  - Data Science competences and skills
  - Essential Data Scientist professional skills: Thinking and doing like Data Scientist
- Data Science Professional Profiles
  - Managing Data Science Teams
- Data Science Body of Knowledge and Model Curriculum
- Use of EDSF and Example curricula
  - Competences assessment
  - Building Data Science team
- References and additional materials

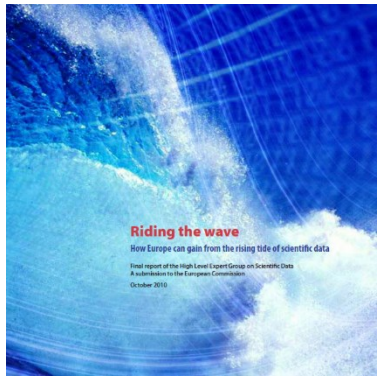
Historical: Visionaries and Drivers: Seminal works, High level reports, Activities



The Fourth Paradigm: Data-Intensive Scientific Discovery.

By Jim Gray, Microsoft, 2009. Edited by Tony Hey, Kristin Tolle, et al.

http://research.microsoft.com/en-us/collaboration/fourthparadigm/



Riding the wave: How Europe can gain from the rising tide of scientific data.

Final report of the High Level Expert Group on Scientific Data. October 2010.

http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

The Data Harvest: How sharing research data can yield knowledge, jobs and growth.

An RDA Europe Report. December 2014

https://rd-alliance.org/data-harvest-report-sharing-data-knowledge-jobs-and-growth.html

RDA RESEARCH DATA ALLIANCE Research Data Sharing without barriers

https://www.rd-alliance.org/

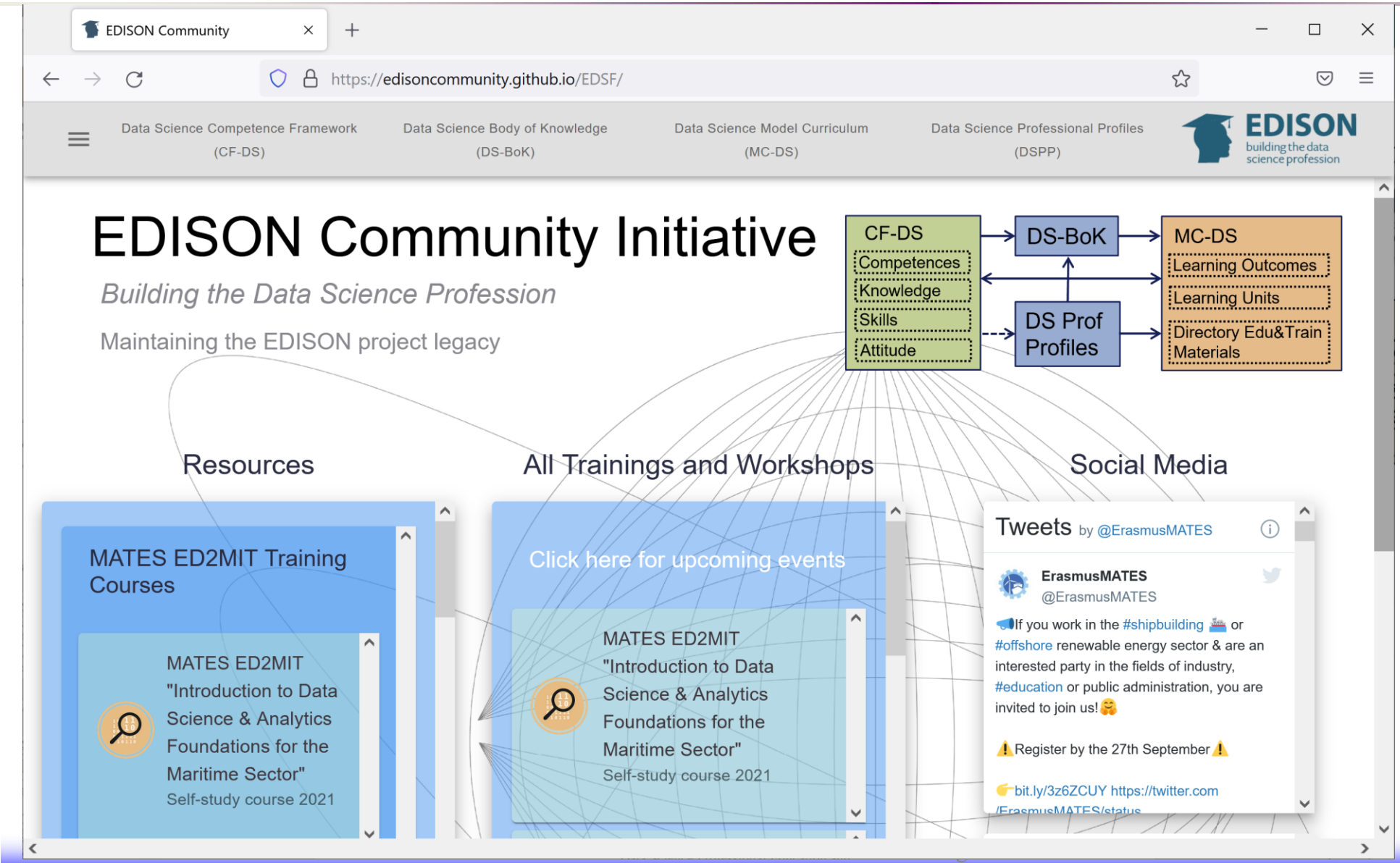
HLEG report on European Open Science Cloud

(October 2016)

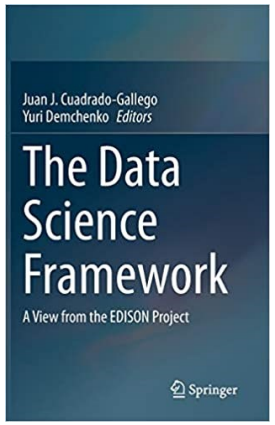
https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf

Data Spaces as a New Data Governance Layer
International Data Spaces Alliance, RDA, EOSC

https://edisoncommunity.github.io/EDSF/
http://www.edison-project.net



- EDISON Project, EU funded Horizon 2020 (2015-2017)
- Community Initiative supported by University of Amsterdam since 2017
- Further development in EU projects MATES, FAIRsFAIR
- Book on EDSF published

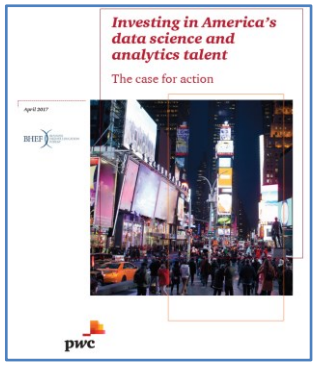


The Data Science Framework: A View from the EDISON Project, Oct 2020.
Editors J.J.Cuadrado-Gallego, Y.Demchenko

Industry reports on Data Science Analytics and Data enabled skills demand

- Final Report on European Data Market Study by IDC (Feb 2017)
  - The EU data market in 2016 estimated EUR 60 Bln (growth 9.5% from EUR 54.3 Bln in 2015)
    - Estimated EUR 106 Bln in 2020
  - 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%

- 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
  - 2.35 mln postings, 23% Data Scientist, 67% DSA enabled jobs
  - DSA enabled jobs growing at higher rate than main Data Science jobs



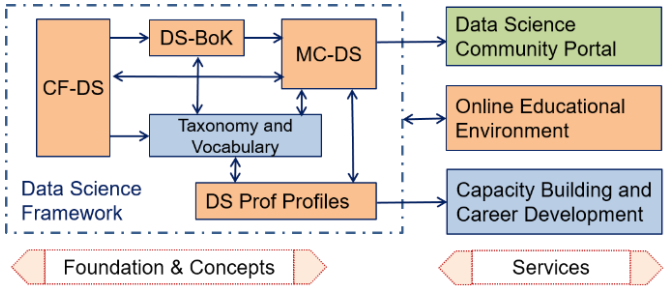
Citing EDISON and EDSF

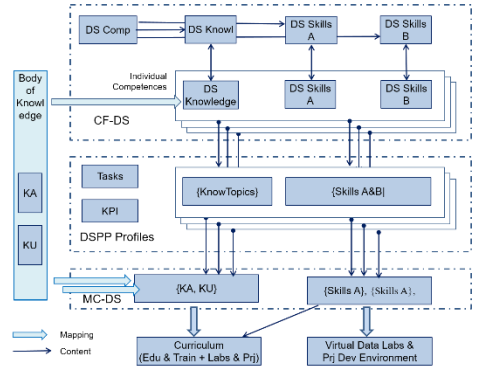
- Burning Glass Technology, IBM, and BHEF report “The Quant Crunch: How the demand for Data Science Skills is disrupting the job Market” (April 2017) - Edited
  - https://public.dhe.ibm.com/common/ssi/ecm/im/en/iml14576usen/IML14576USEN.PDF
  - DSA enabled jobs takes 45-58 days to fill: 5 days longer than average
  - Commonly required work experience 3-5 yrs

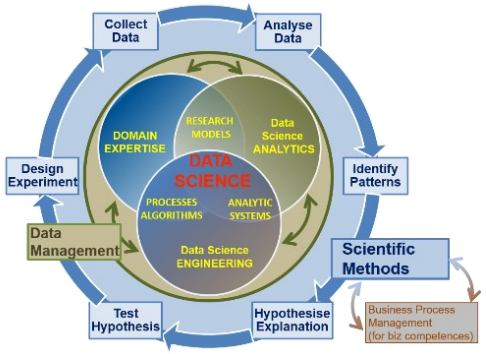
Influenced by EDISON

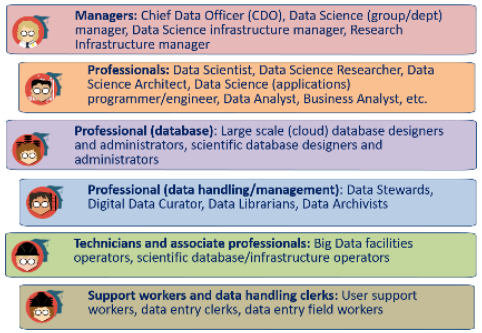
EDISON Products for Data Science Skills Management and Curriculum Design

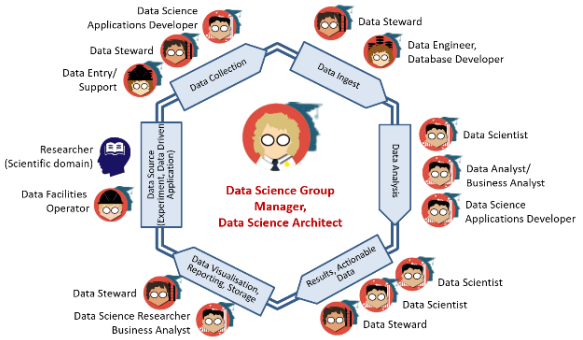
- EDISON Data Science Framework (EDSF)
  - Release 3 components CF-DS, DS-BoK, MC-DS, DSPP
  - Compliant with EU standards on competences and professional occupations e-CFv3.0, ESCO
- Skills development and career management for Core Data Experts and related data handling professions
- Academic programmes and professional training courses (self) assessment and design
- Individual competences benchmarking and Data Science team design
- Cooperation with International professional organisations IEEE, ACM, BHEF, APEC (AP Economic Cooperation)

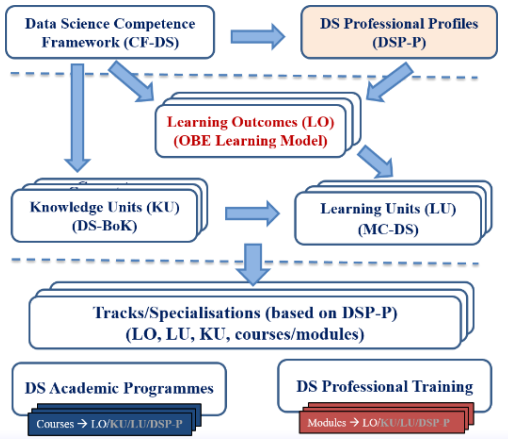




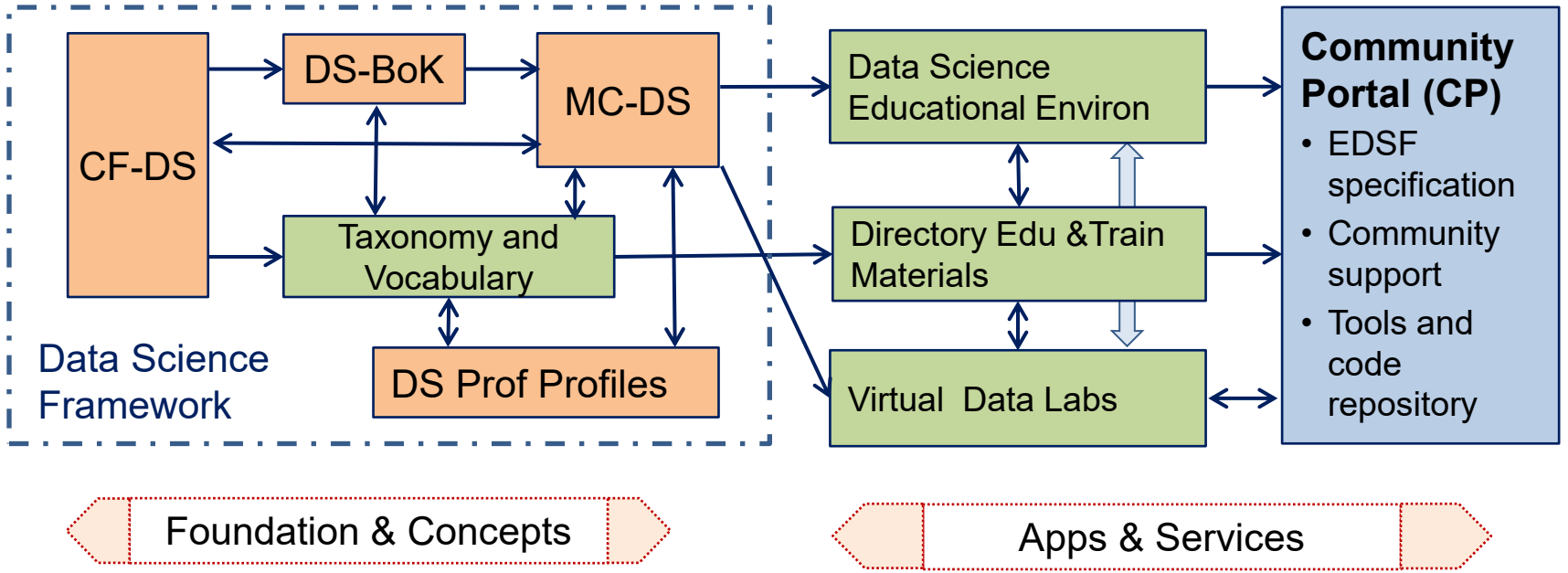








EDISON Data Science Framework (EDSF release 4) – Core components and community maintained services



EDISON Framework core components and documents

- CF-DS – Data Science Competence Framework (Part 1)
- DS-BoK – Data Science Body of Knowledge (Part 2)
- MC-DS – Data Science Model Curriculum (Part 3)
- DSPP – Data Science Professional profiles (Part 4)
- EDSF-UCA – EDSF Use Cases and Applications
- Data Science Taxonomies and Scientific Disciplines Classification

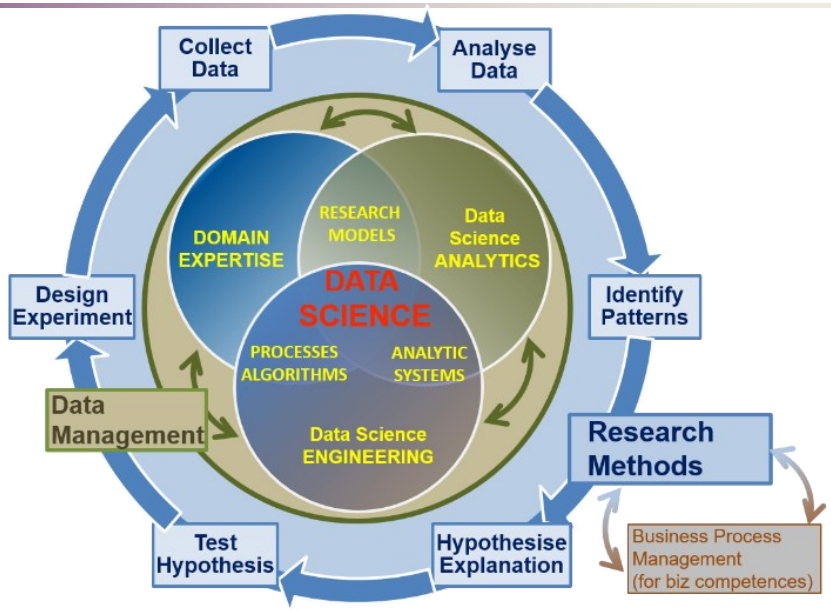
Applications and Services

- Virtual Data Science Labs
- Data Science Educational Environment
- Directory of edu & train resources
- Community Portal – currently github

Data Scientist definition

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

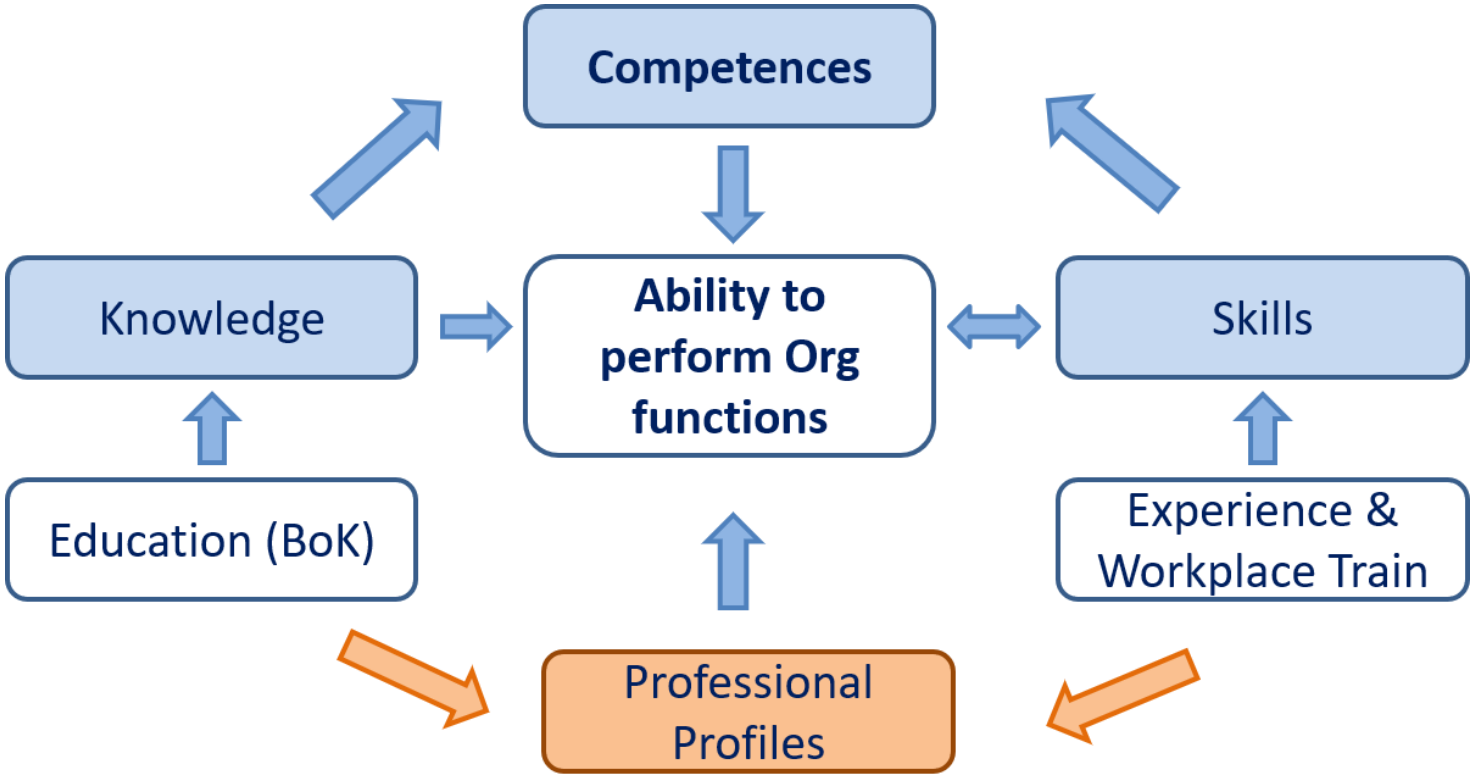
- *A **Data Scientist** is a practitioner who has sufficient knowledge in the overlapping regimes of expertise in **business needs, domain knowledge, analytical skills, and programming and systems engineering expertise** to manage the end-to-end scientific method process through each stage in the **big 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

Competences Map to Knowledge and Skills

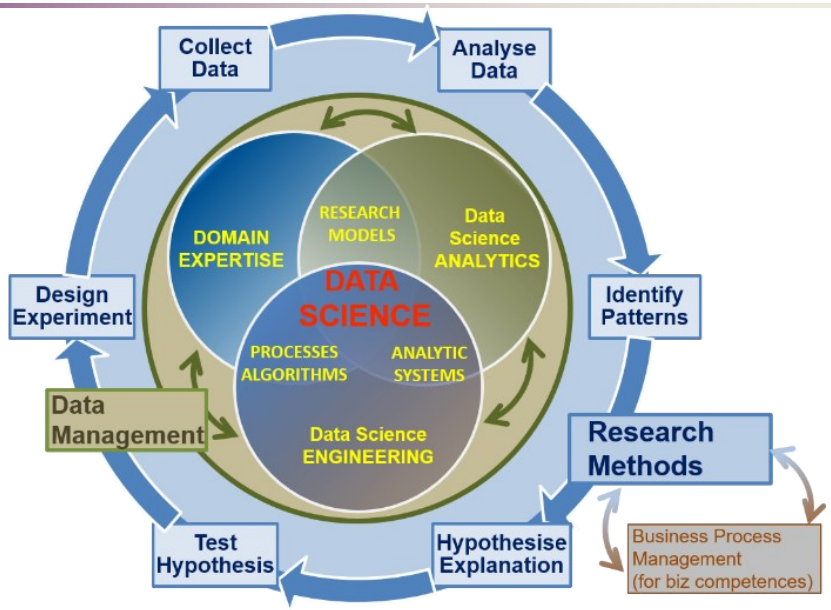
- **Competence** is a demonstrated ability to apply knowledge, skills and attitudes for achieving observable results



Data Scientist definition

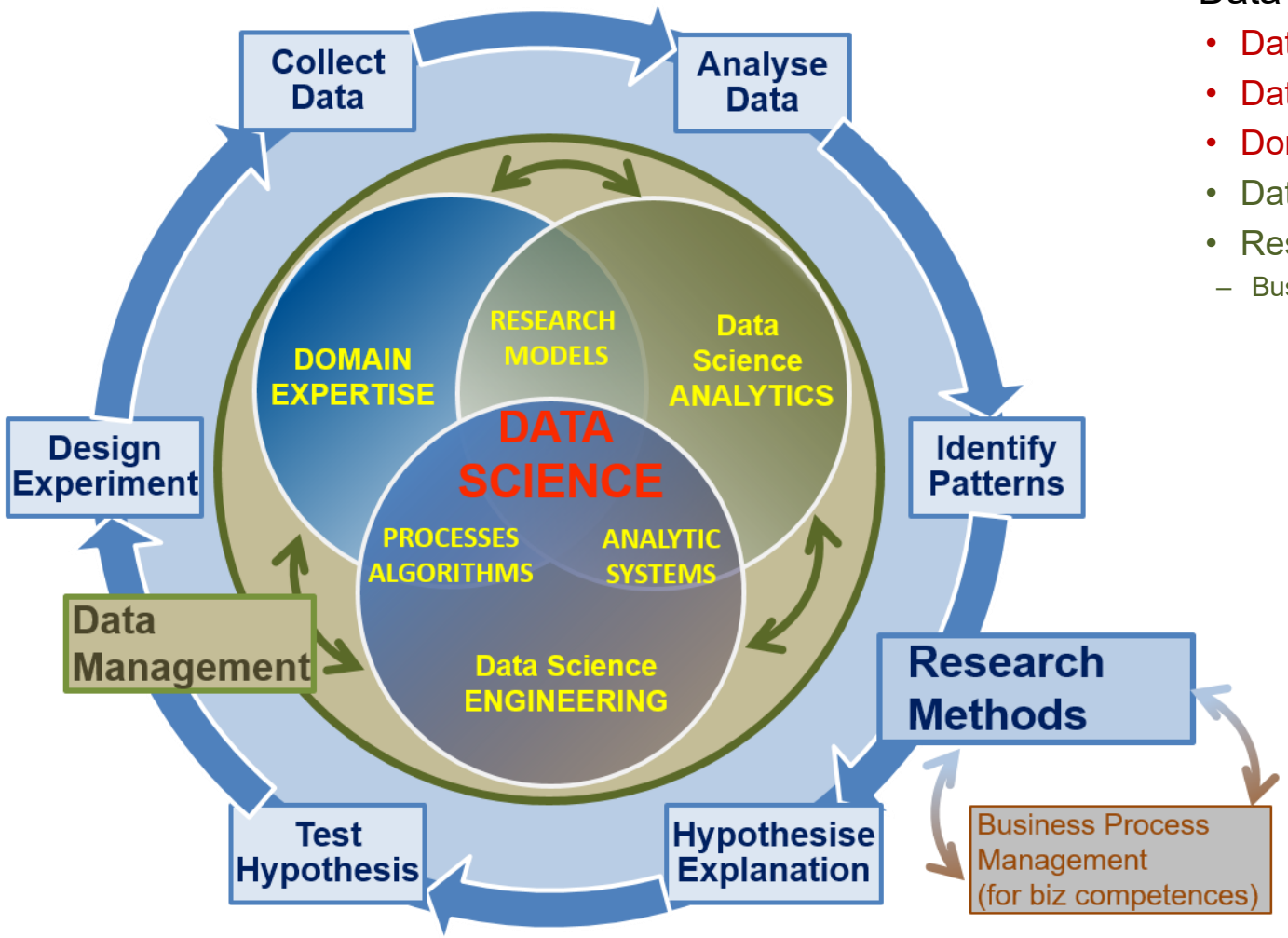
Based on the definitions by NIST SP1500 – 2015, extended by EDISON

- *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)

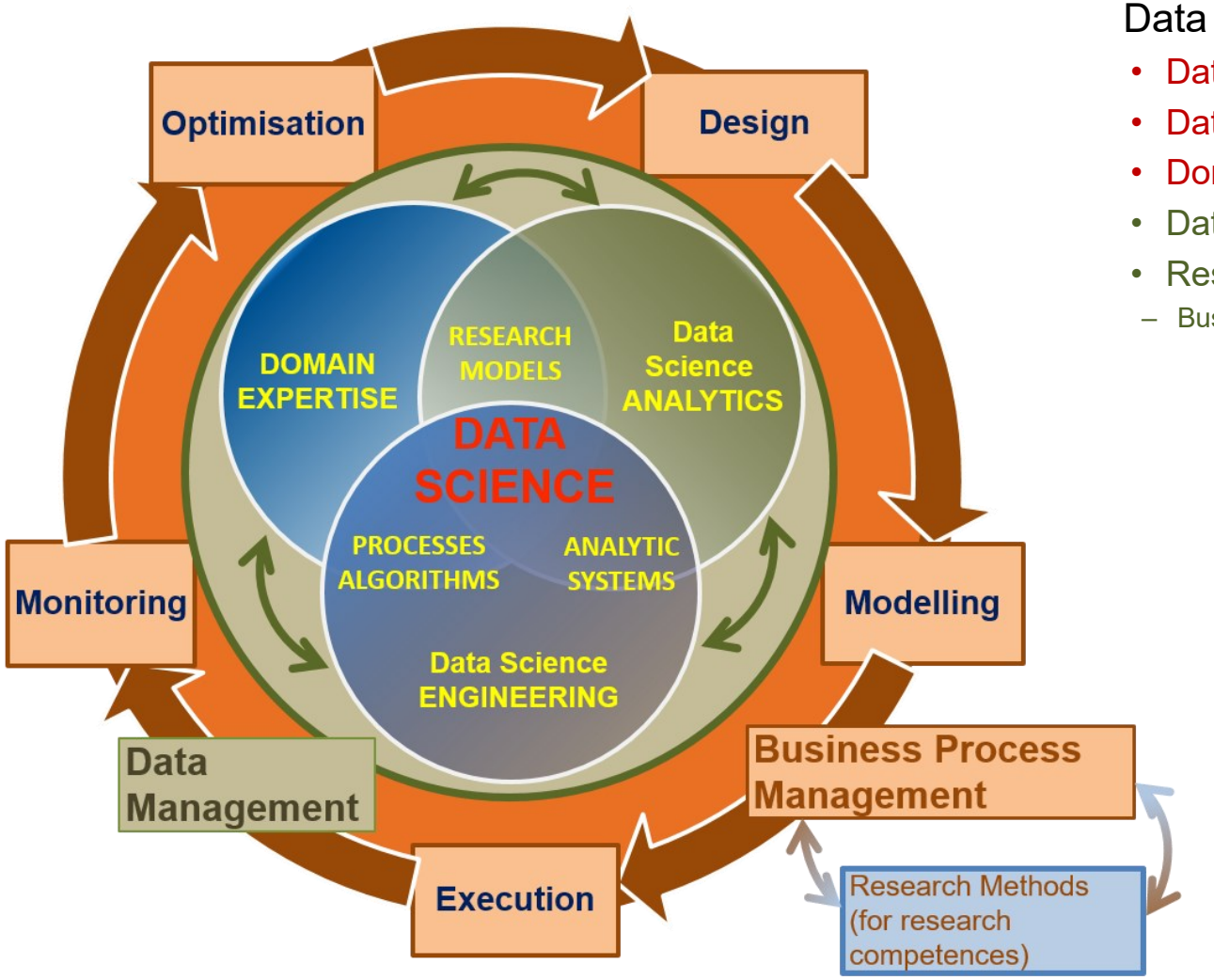
Scientific Methods

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

Business Operations

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

Data Science Competences Groups – Business



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
- Hypothesise Explanation
- Test Hypothesis

Business Process Operations/Stages

- Design
- Model/Plan
- Deploy & Execute
- Monitor & Control
- Optimise & Re-design

Identified Data Science Competence Groups

	Data Science Analytics (DSDA)	Data Science Engineering (DSENG)	Data Management and Governance (DSDM)	Research/Scientific Methods and Project Management (DSRMP)	Data Science Domain Knowledge, e.g. Business Analytics (DSDK/DSBPM)
0	Use appropriate data analytics and statistical techniques on available data to deliver insights into research problem or org. processes and support decision making	Use engineering principles and modern computer technology to research, design, implement new data analytics applications, develop experiments, processes, instruments, systems and infrastructures to support data handling during the whole data lifecycle	Develop and implement data management strategy for data collection, storage, preservation, and availability for further processing.	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	DSDK/DSBA 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
1	**DSDA01** Effectively use variety of data analytics techniques	**DSENG01** Use engineering principles (general and software) to research, design, develop and implement new instruments and applications	DSDM01 Develop and implement data strategy, in particular, Data Management Plan (DMP)	**DSRMP01** Create new understandings and capabilities by using scientific/ research methods	**DSBPM01** Understand business and provide insight, translate unstructured business problems into an abstract mathematical framework
2	**DSDA02** Apply designated quantitative techniques	**DSENG02** Develop and apply computer methods to domain related problems	DSDM02 Develop data models including metadata	**DSRMP02** Direct systematic study toward a fuller knowledge or understanding of the observable facts	**DSBPM02** Participate strategically and tactically in financial decisions
3	**DSDA03** Pull together data from diff sources …	**DSENG03** Develop and prototype data analytics applications	DSDM03 Collect integrate data	**DSRMP03** Undertakes creative work	**DSBPM03** Provides support services to other
4	**DSDA04** Use diff perform techniques	**DSENG04** Develop, deploy operate Big Data storage	DSDM04 Maintain repository	**DSRMP04** Translate strategies into actions	**DSBPM04** Analyse data for marketing
5	**DSDA05** Develop analytics applic	**DSENG05** Apply security mechanisms	DSDM05 Visualise cmplx data	**DSRMP05** Contribute to organis goals	**DSBPM05** Analyse optimise customer relatio
6	**DSDA06** Visualise results of analysis, dashboards	**DSENG06** Design, build, operate SQL and NoSQL	**DSRM06** Develop and manage prolicies	**DSRMP06** Develop and guide data driven projects	**DSBPM06** Analyse data for marketing

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) platforms and tools**
  - 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 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 of 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 (after DARE & BHEF & EDISON)

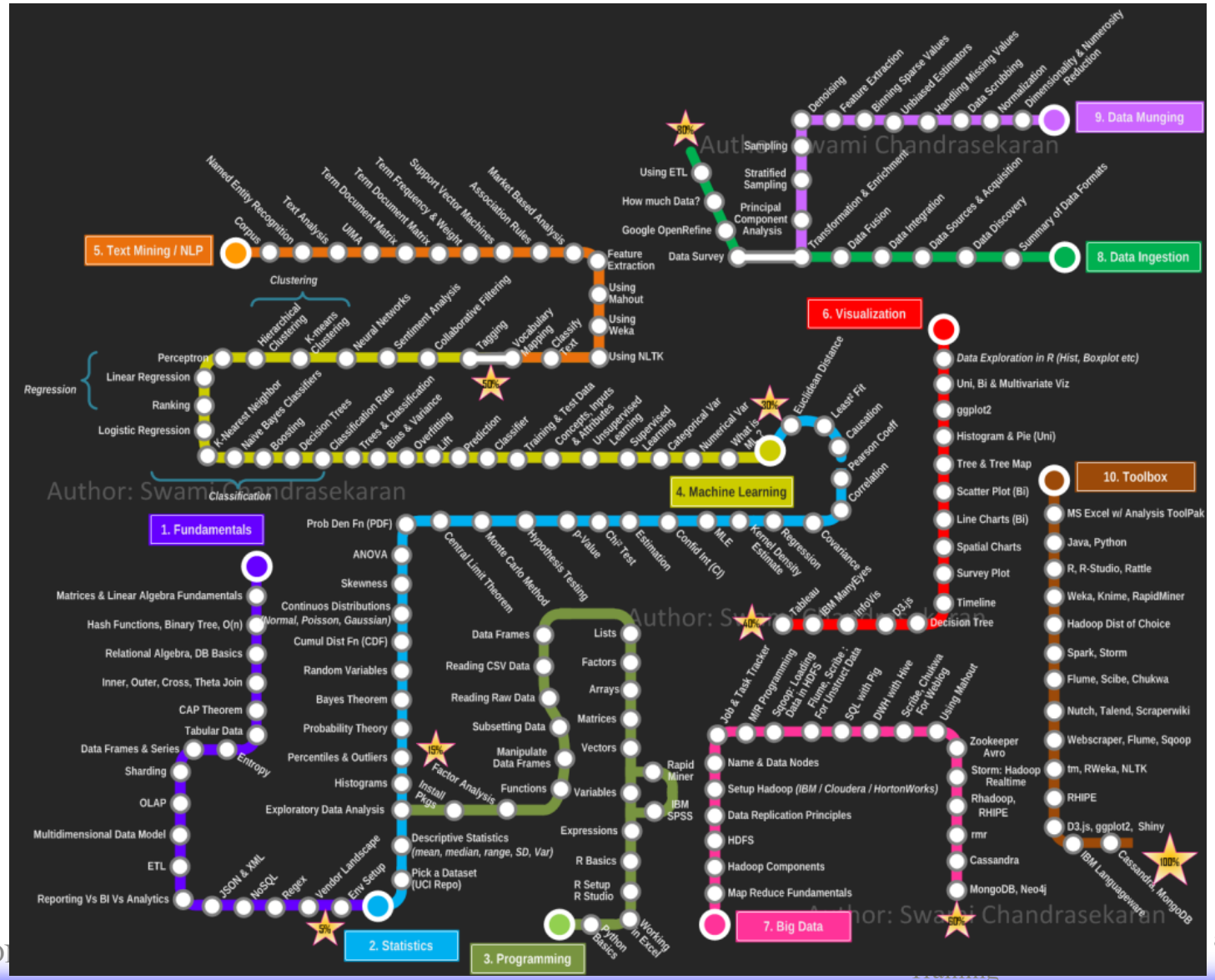
1. **Critical Thinking:** Demonstrating the ability to apply critical thinking skills to solve problems and make effective decisions
2. **Communication:** Understanding and communicating ideas
3. **Collaboration:** Working with other, appreciation of multicultural difference
4. **Creativity and Attitude:** Deliver high quality work and focus on final result, initiative, intellectual risk

5. **Design Thinking:** Structural problem solving and implementation aware solutions
6. **Planning & Organizing:** Planning and prioritizing work to manage time effectively and accomplish assigned tasks

7. **Business Fundamentals:** Having fundamental knowledge of the organization and the industry
8. **Customer Focus:** Actively look for ways to identify market demands and meet customer or client needs
9. **Working with Tools & Technology:** Selecting, using, and maintaining tools and technology to facilitate work activity
10. **Dynamic (self-) re-skilling:** Continuously monitor individual knowledge and skills as shared responsibility between employer and employee, ability to adopt to changes
11. **Professional networking:** Involvement and contribution to professional network activities
12. **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

Becoming a Data Scientist by Swami Chandrasekaran (2013)

http://nirvacana.com/thoughts/becoming-a-data-scientist/

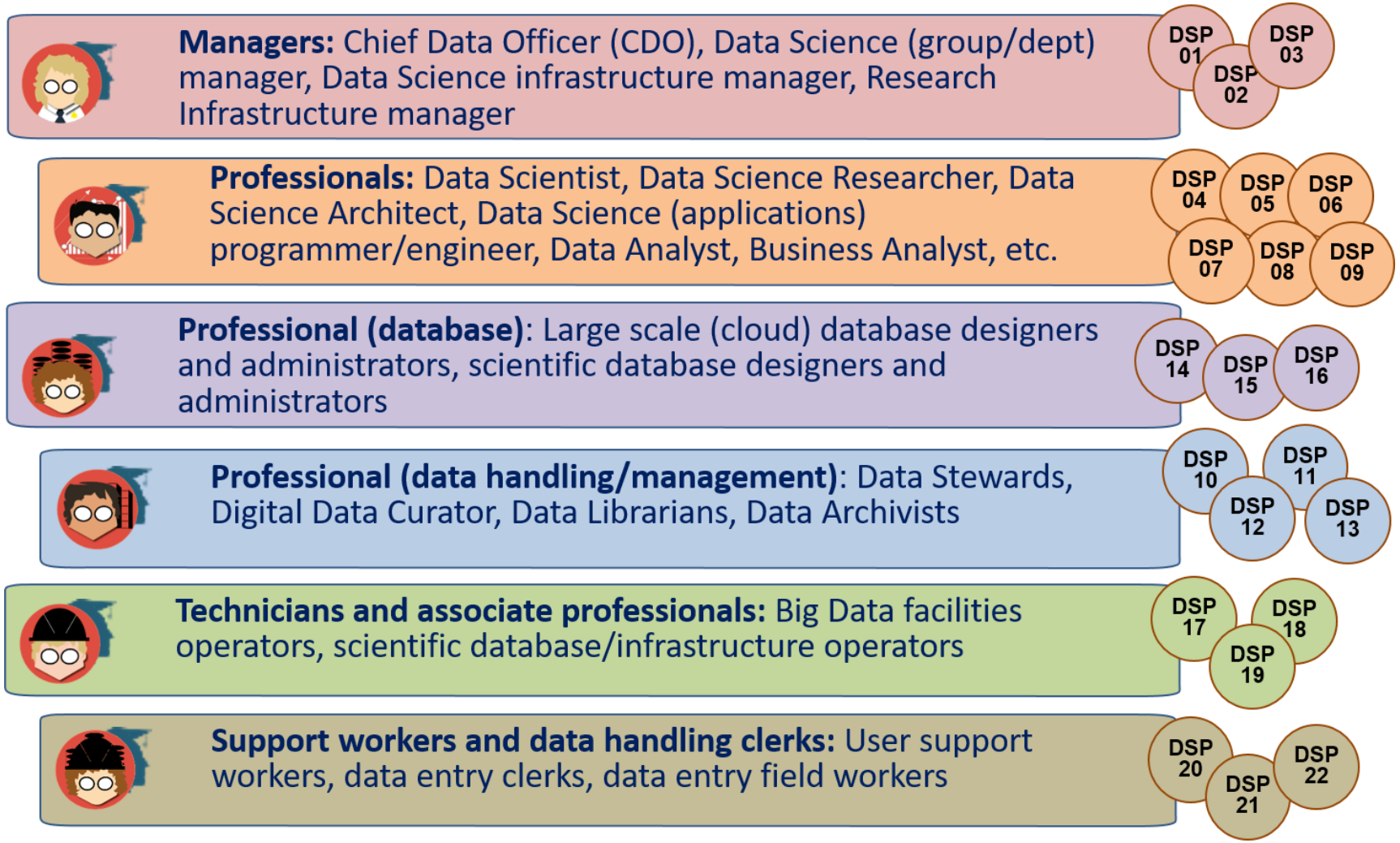


- Good and practical advice how to learn Data Science, step by step
- Follow the route
- Understand and build your route

Practical Application of the CF-DS

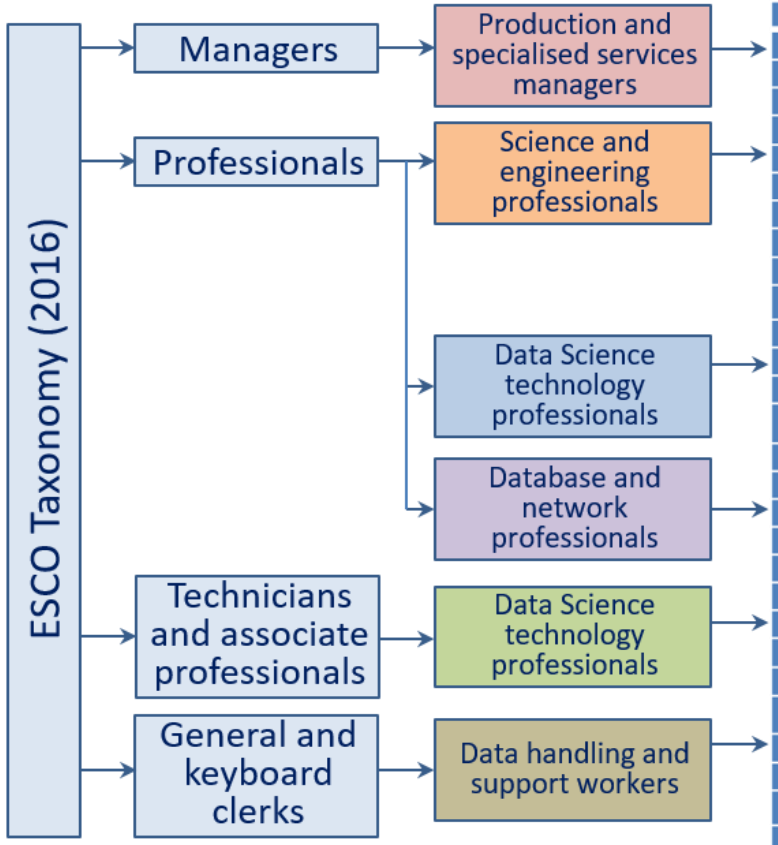
- Basis for the definition of the Data Science Body of Knowledge (DS-BoK) and Data Science Model Curriculum (MC-DS)
  - CF-DS => Learning Outcomes (MC-DS) => Knowledge Areas (DS-BoK)
  - CF-DS => Data Science taxonomy of scientific subjects and vocabulary
- Data Science professional profiles definition
  - 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



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

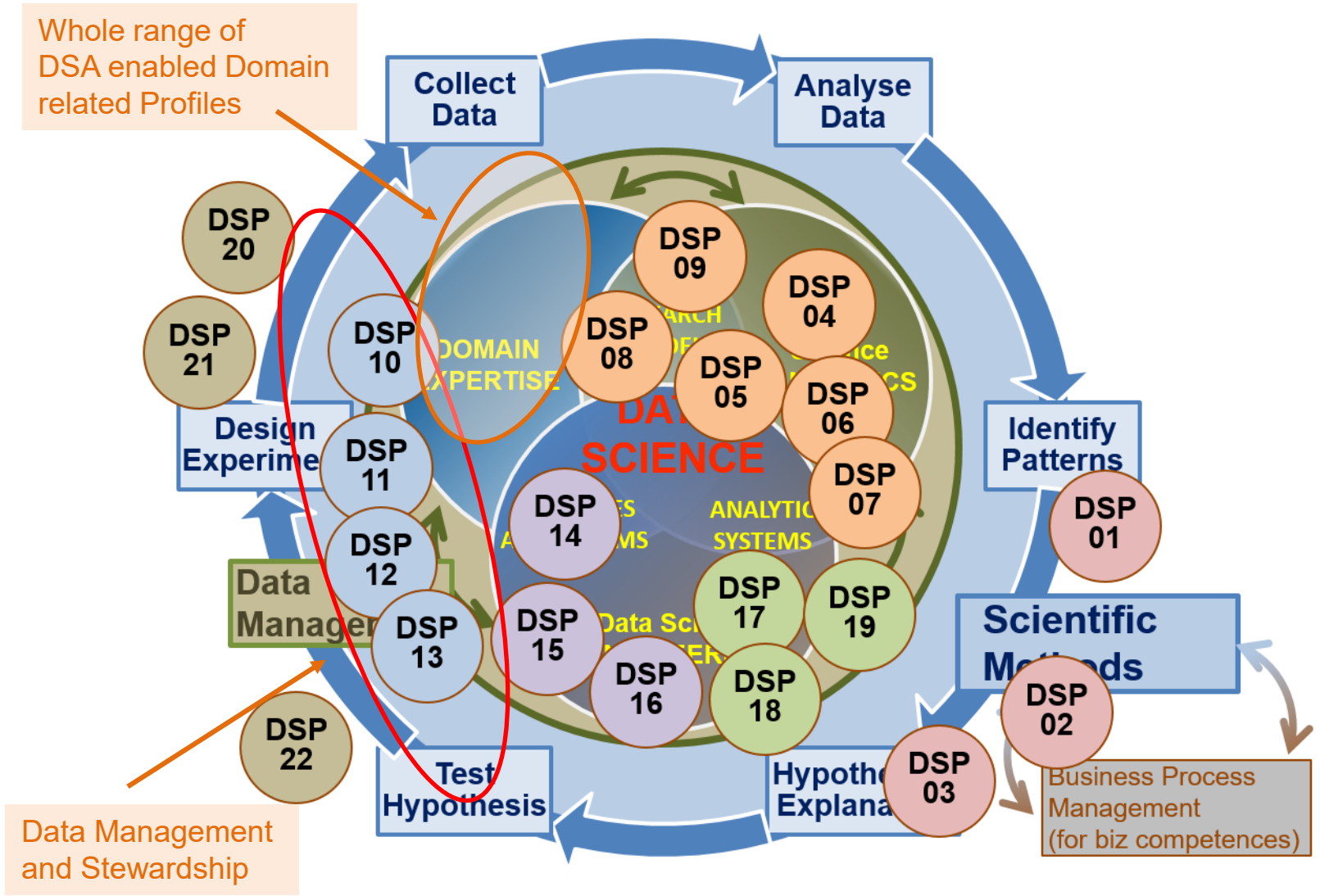
DSP Profiles mapping to ESCO Taxonomy High Level Groups



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
  - Relevance level from 5 – maximum to 1 – minimum

CF-DS and Data Science Professional Profiles



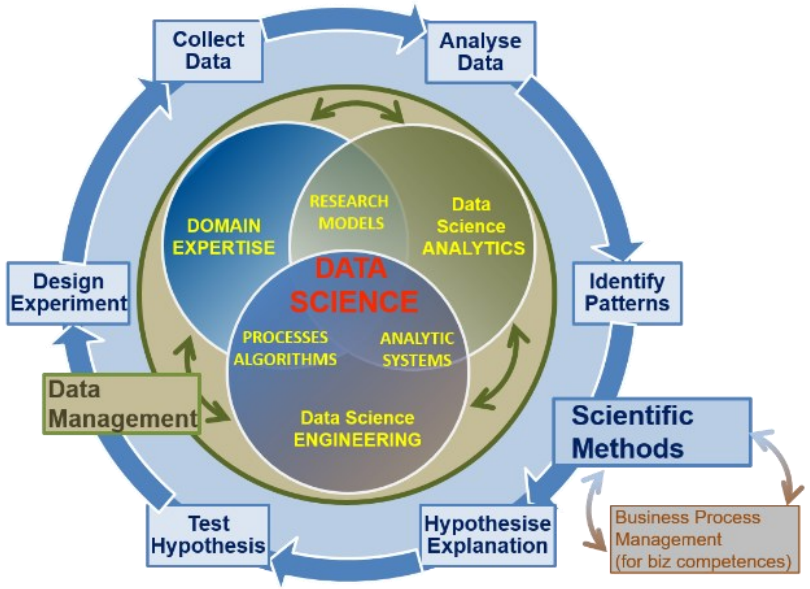
EDSF for Education and Training

- 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

Data Science Body of Knowledge (DS-BoK)

DS-BoK Knowledge Area Groups (KAG)

- 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*
- KAG4-DSRM: *Research Methods and Project Management group*
- KAG5-DSBA: Business Analytics and Business Intelligence



- KAG* - DSDK: Data Science domain knowledge to be defined by related expert groups

Data Science Model Curriculum (MC-DS)

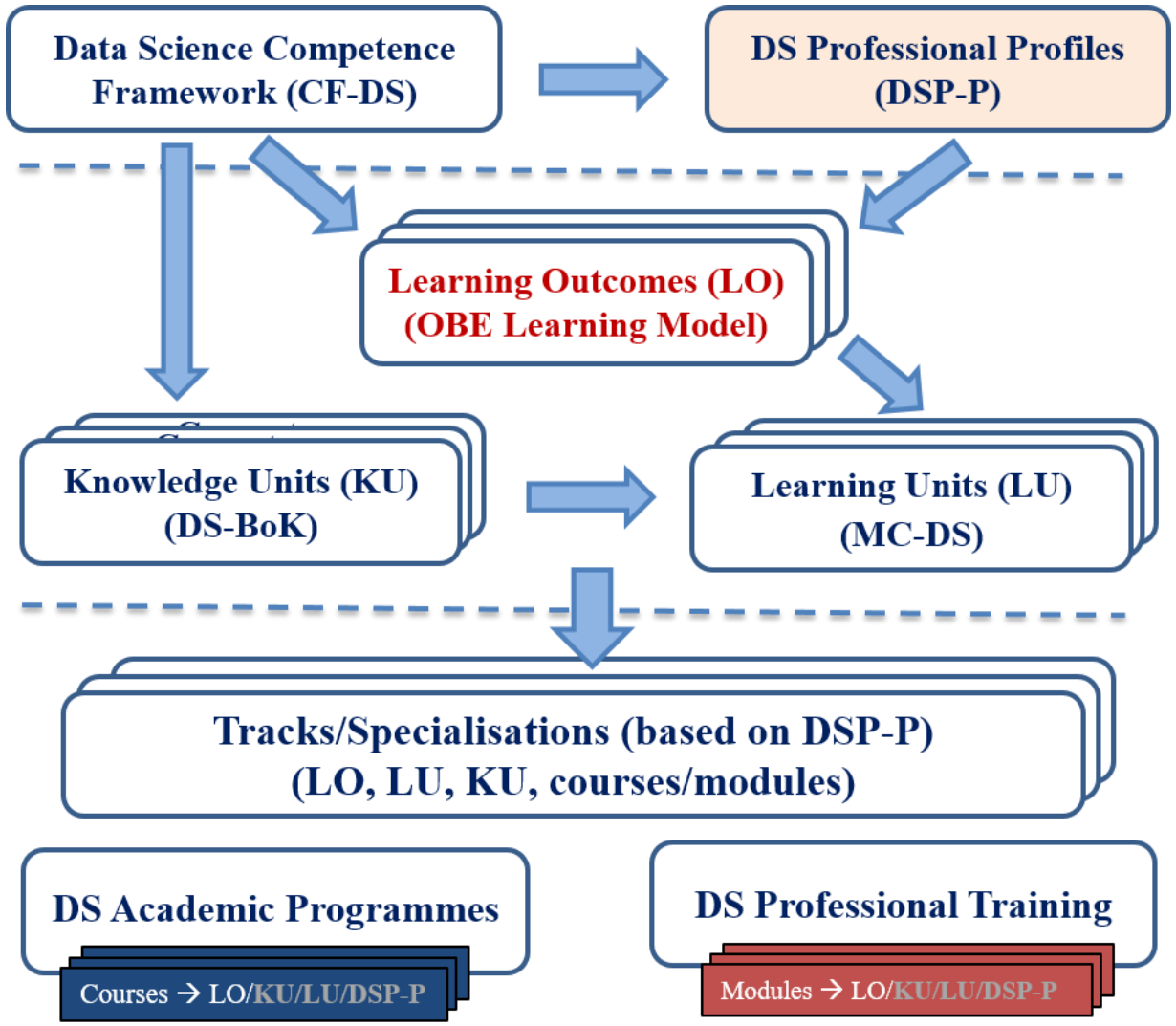
Data Science Model Curriculum includes

- Learning Outcomes (LO) definition based on CF-DS
  - LOs are defined for CF-DS competence groups and for all enumerated competences
  - Knowledge levels: Familiarity, Usage, Assessment (based in Bloom's Taxonomy)
- LOs mapping to Learning Units (LU)
  - LUs are based on CCS(2012) and universities best practices
  - Data Science university programmes and courses inventory (interactive) http://edison-project.eu/university-programs-list
- LU/course relevance: Mandatory Tier 1, Tier 2, Elective, Prerequisite
- Learning methods and learning models (in progress)

Knowledge levels for Learning Outcomes (defined based on Bloom's Taxonomy)

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

Outcome Based Educations and Training Model: Addressing target competences for the profession



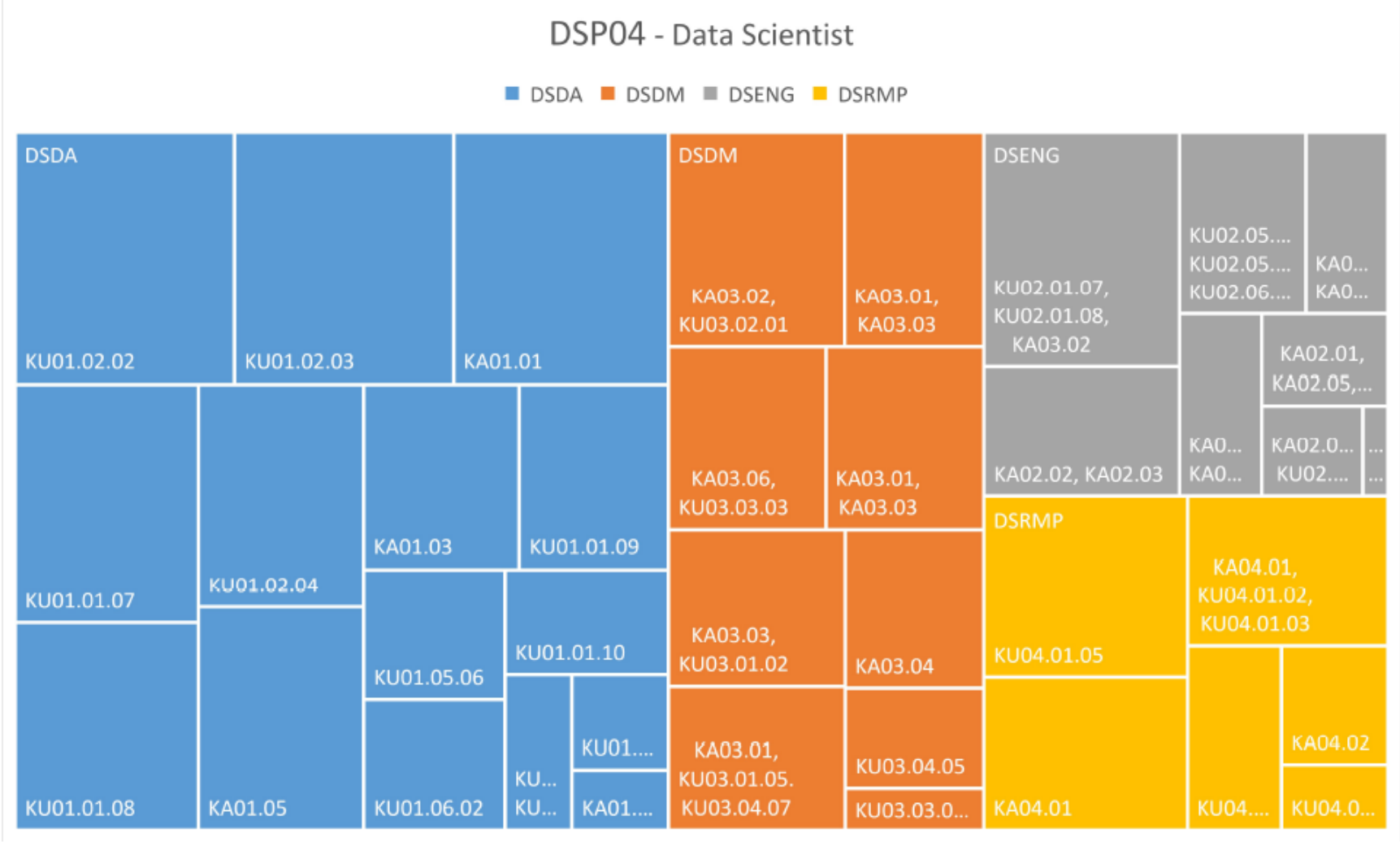
From Competences and DSP Profiles

to Learning Outcomes (LO) and

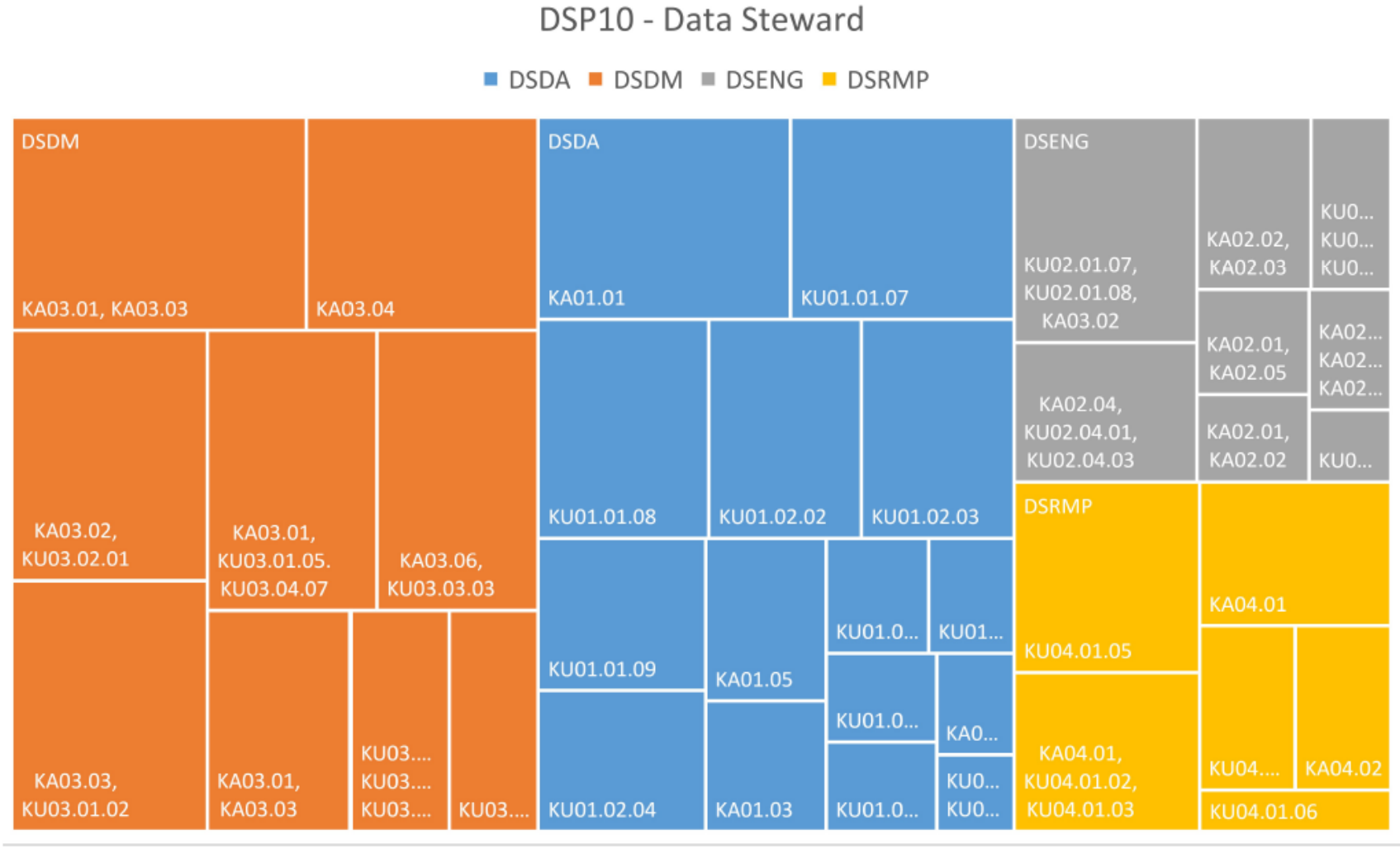
to Knowledge Unites (KU) and Learning Units (LU)

- EDSF allow for customized educational courses and training modules design

DSP04 – Data Scientist MC structure



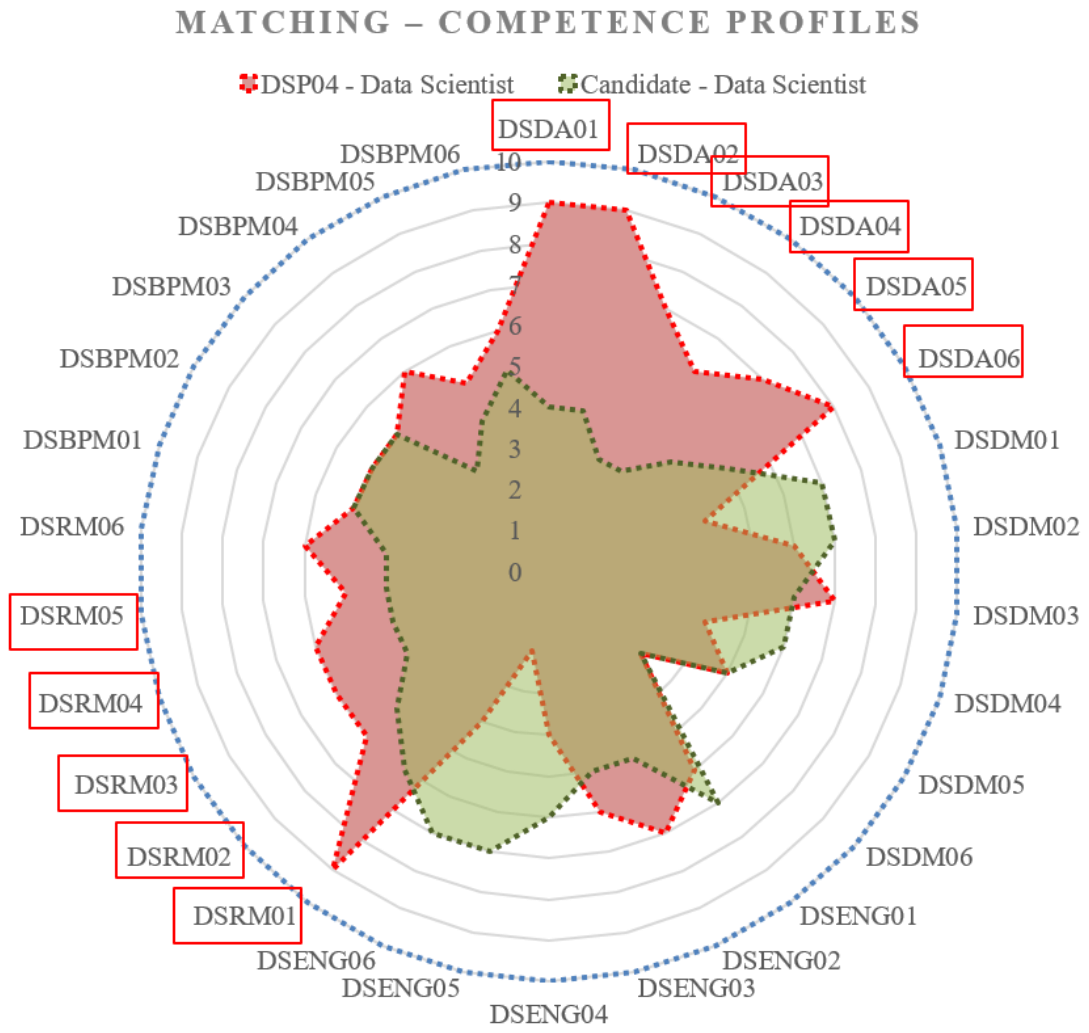
DSP10 – Data Steward MC structure



EDSF Application: Competences assessment and Team building

- Data Science competences assessment (benchmarking)
- Data Science team building

Individual Competences Benchmarking

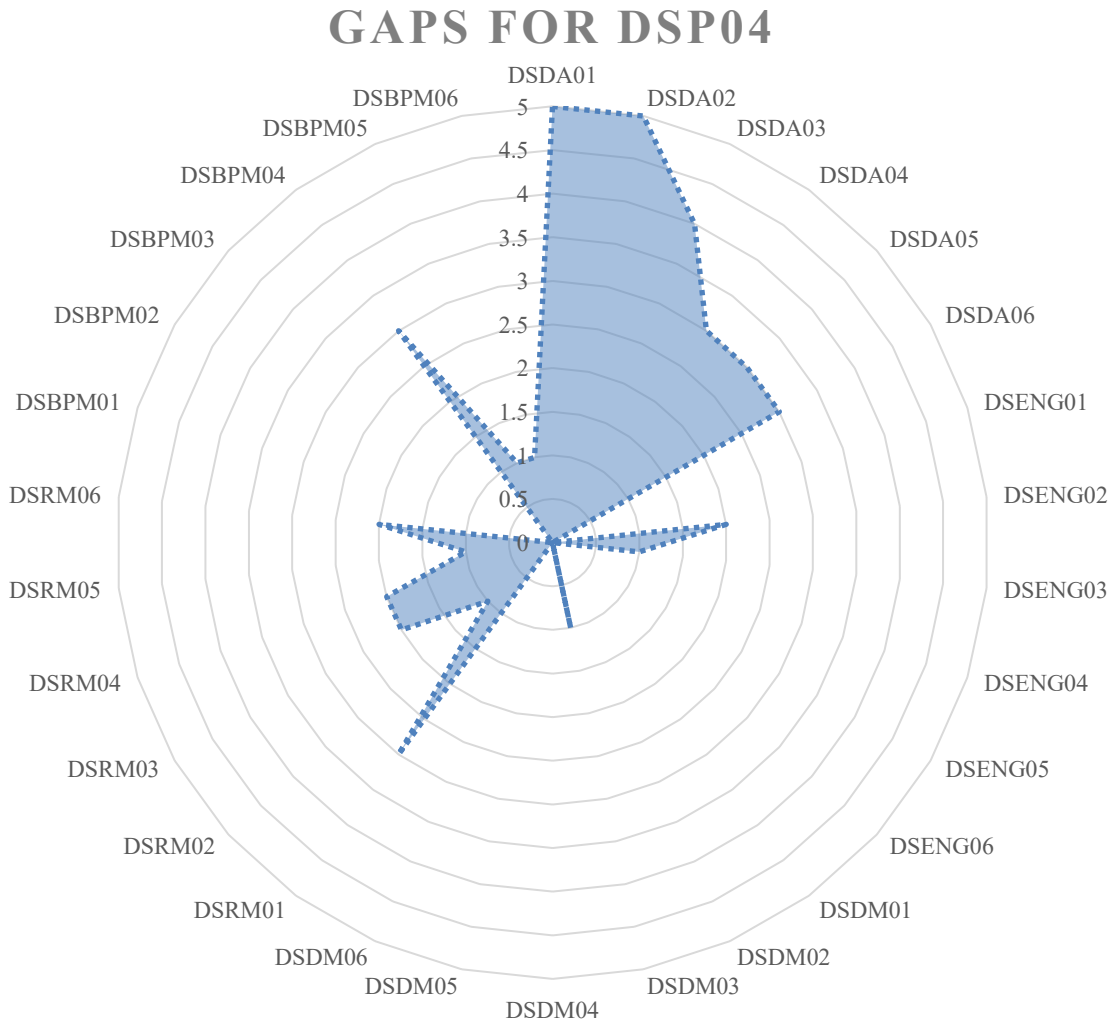


Individual Education/Training Path based on Competence benchmarking

- 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.

Competence/Knowledge gap -> Suggested LUs/courses



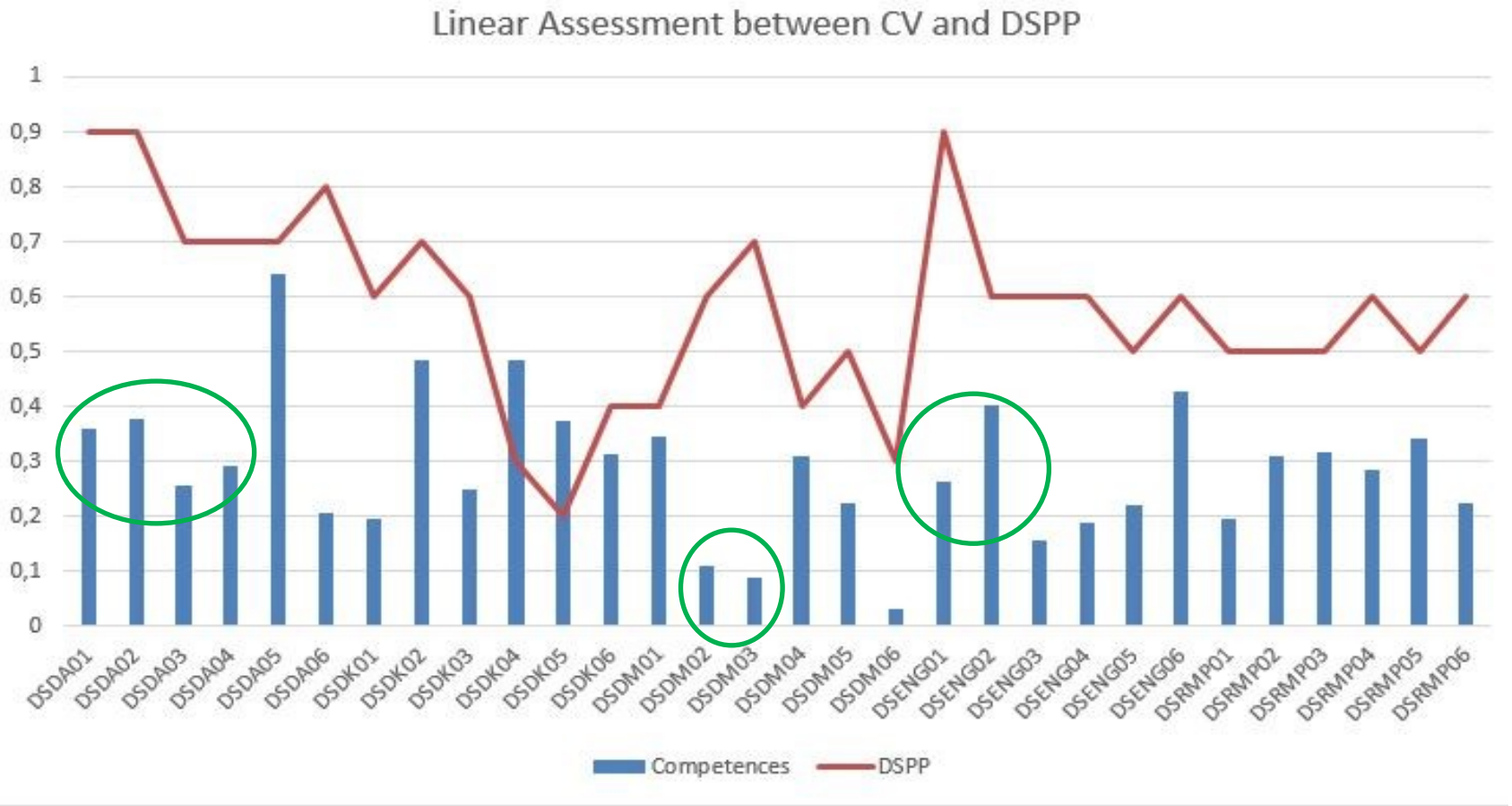
Recommended courses DSDA

- Statistical Methods
- Machine Learning
- Predictive and Quantitative analytics
- Graph Data Analysis
- Data preparation and preprocessing
- Performance Analysis

Recommended courses DSRM

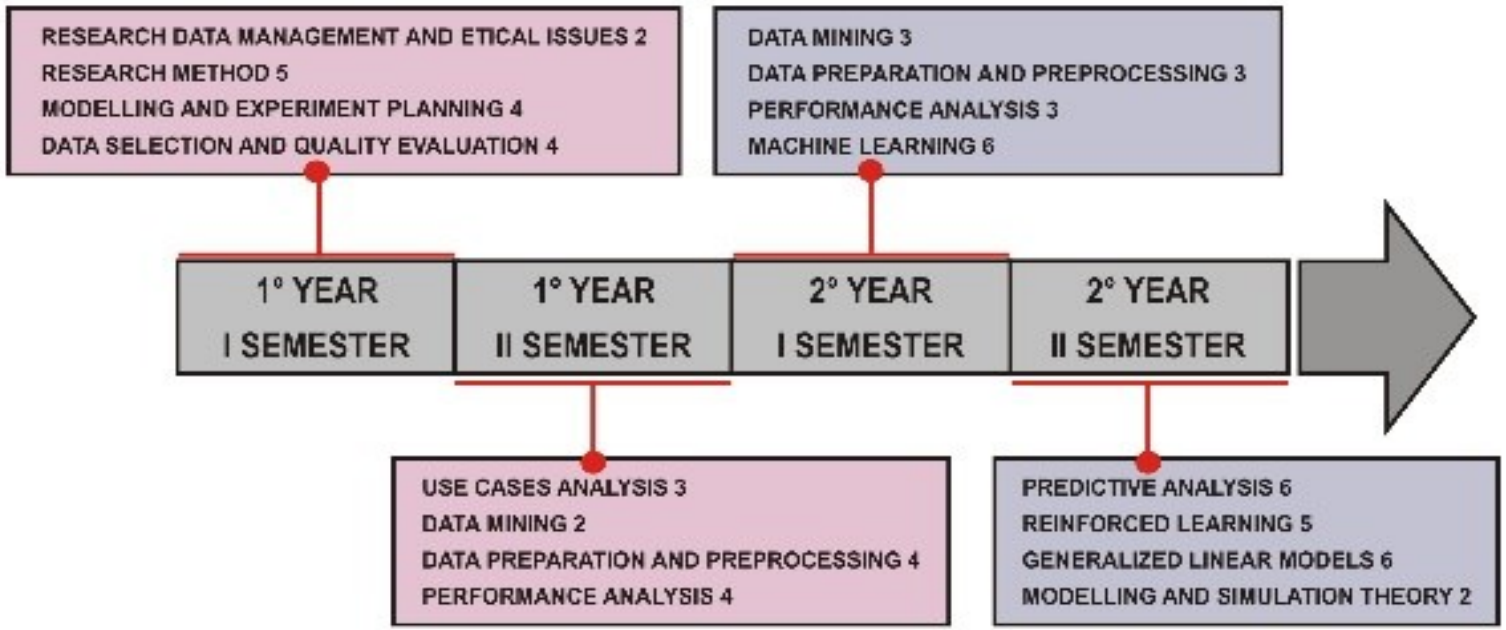
- Research Methods and Project Management

From Competence gap to Proficiency Level and Curriculum Timing – In progress



- Linear skills gap visualisation

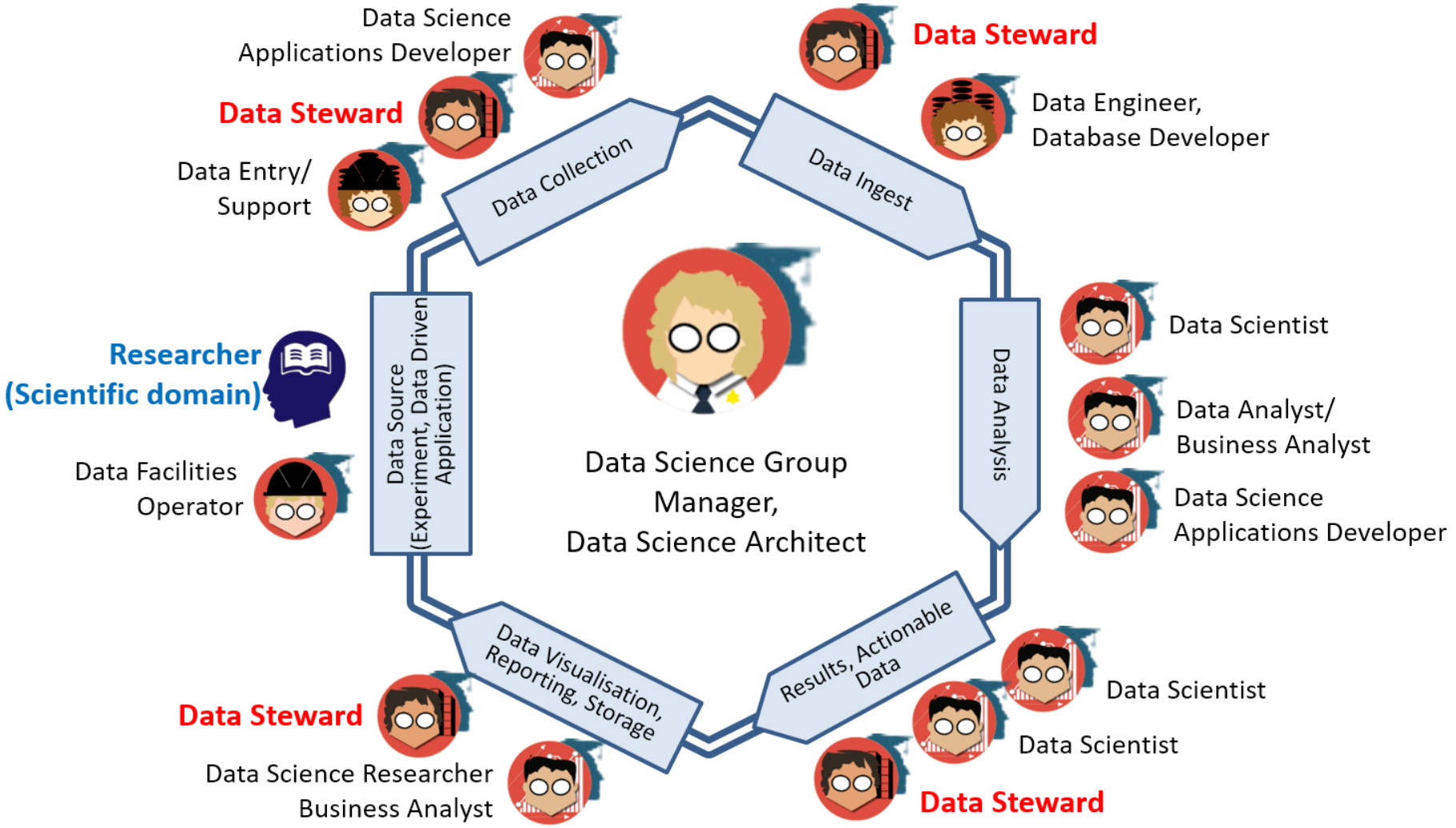
Example curriculum planning: Based on implied courses duration and DSPP profile proficiency levels



Taking into account the learning and cognitive dynamics

- It is essential for key courses to have time span more than 2 months
- Split essential courses over two periods and two semester
- Sequence theory and practice

Building a Data Science Team



Data Science or Data Management Group/Department: Organisational structure and staffing - EXAMPLE

Data Science or Data Management Group/Department

>> Reporting to CDO/CTO/CEO
- Providing cross-organizational services

- (Managing) Data Science Architect (1)
- Data Scientist (1), Data Analyst (1)
- Data Science Application programmer (2)
- Data Infrastructure/facilities administrator/operator: storage, cloud, computing (1)
- **Data stewards**, curators, archivists (3-5)

Estimated: Group of 10-12 data specialists for research institution of 200-300 research staff.

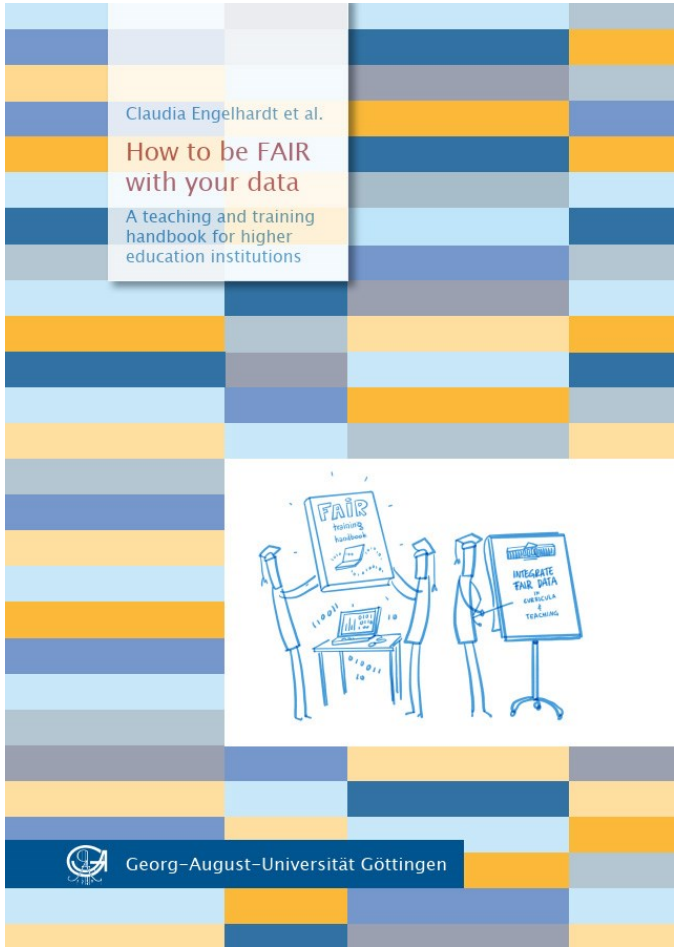
Growing role and demand for Data Stewards and data stewardship

Data Stewardship

- **Creating and managing core Metadata**: Definition and management of business terminology, valid data values, and other critical Metadata.
- **Documenting rules and standards**: Definition/documentation of business rules, data standards, and data quality rules.
  - High quality data are often formulated in terms of rules rooted in the business processes that create or consume data.
  - Stewards help surface these rules and ensure their consistent use.
- **Managing data quality issues**: Stewards are often involved with the identification and resolution of data related issues or in facilitating the process of resolution.
- **Executing operational data governance activities**: Stewards are responsible for ensuring that, day-today and project-by-project, data governance policies and initiatives are adhered to. They should influence decisions to ensure that data is managed in ways that support the overall goals of the organization.

“Best Data Steward is not made but found” DMBOK1 (2009)

EDSF in FAIRsFAIR Project: Data Stewardship Professional Competence Framework (DSP-CF) and FAIR Teaching Handbook



Competence groups extended (by FAIRsFAIR Project)

- DSDM – Data Management and Governance
  - Extended DSDM01 - DSDM06
  - Introduced new
  - DSDM07* Manage Data Management/Data Stewards team
  - DSDM08* Develop organisational policy and implementation of the FAIR data principles and Open Science
  - DSDM09* Specify requirements to and supervise the organisational data management and tools
- DEENG – Data Science Engineering
  - Extended DSENG03-DSENG06
- DSBA – Data Science Business Analytics
  - Extended DSBA01-DSBA04
  - Introduced new DSBA07 Coordinate intra organisational activities related to data management

FAIRSFAIR
Fostering Fair Data Practices in Europe

Project Title	Fostering FAIR Data Practices in Europe
Project Acronym	FAIRsFAIR
Grant Agreement No	831558
Instrument	H2020-INFRAEOSC-2018-4
Topic	INFRAEOSC-05-2018-2019 Support to the EOSC Governance
Start Date of Project	1st March 2019
Duration of Project	36 months
Project Website	www.fairsfair.eu

D7.3 FAIR Competence Framework for Higher Education (Data Stewardship Professional Competence Framework)

Work Package	WP7, FAIR Competences for Higher Education
Lead Author (Org)	Yuri Demchenko (UvA)
Contributing Author(s) (Org)	Lennart Stoy (EUA), Claudia Engelhardt (UGOE), Vinciane Gaillard (EUA)
Due Date	28.02.2021
Date	24.02.2021
Version	1.0
DOI	https://doi.org/10.5281/zenodo.4562088

Ongoing activities and developments
https://github.com/EDISONcommunity/EDSF

- **EDISON Data Science Framework Release 4** – Published 31 December 2022
  - Final Release
  - Call for maintainer role
- **EDISON Data Science Framework Release 3** – Published 31 December 2018 (obsolete)
- **Industry digitalisation projects and data literacy skills training**
  - **MATES project funded by EU ERASMUS Programme (2018-2022) -** EU Maritime industry digital transformation, data skills development and Data + Ocean literacy
  - Skills for SME: Data Science, IoT, Cybersecurity: Prepare to Data Economy and Industry 4.0
- **FAIRsFAIR project (2019-2022)**
  - Data Stewardship and FAIR data principles in the university curricula
  - Data Stewardship competences and Body of Knowledge definition
- **DARE project by APEC** (Asia Pacific Economic Cooperation)
  - Continuing cooperation for Asia Pacific region – Workshop 15-16 July 2019, AP
  - Recommended Data Science and Analytics Competences published August 2017 - https://www.apec.org/Press/Features/2017/0620_DSA

EDISON Initiative Online Presence

- EDSF github project - https://github.com/EDISONcommunity/EDSF
  - Component documents CF-DS, DS-BoK, MC-DS, DSPP

- EDISON Community work area and discussions - https://github.com/EDISONcommunity/EDSF/wiki/EDSFhome
- Mailing list - edison-net@list.uva.nl

- EDISON project website
  - Legacy information to be moved to http://edison-project.net/
  - Old domain *edison-project.eu* expired and contains outdated information

Links to EDISON Resources

EDISON Data Science Framework Release 4 (EDSF2022)
https://github.com/EDISONcommunity/EDSF

Component EDSF documents

CF-DS – Data Science Competence Framework
https://github.com/EDISONcommunity/EDSF/blob/master/EDISON01_CF-DS-release4-v11.pdf

DS-BoK – Data Science Body of Knowledge
https://github.com/EDISONcommunity/EDSF/blob/master/EDISON02_DS-BoK-release4-v07.pdf

MC-DS – Data Science Model Curriculum
https://github.com/EDISONcommunity/EDSF/blob/master/EDISON03_MC-DS-release4-v07.pdf

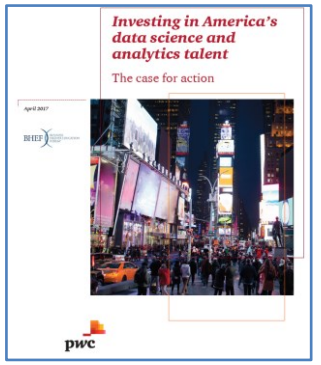
DSPP – Data Science Professional profiles
https://github.com/EDISONcommunity/EDSF/blob/master/EDISON04_DSPP-release4-v08.pdf

Additional materials

- EDSF Reviews, citations and derivative developments
- Data Stewardship Professional Competence Framework (DSP-CF) and FAIR curricula guidelines
- MATES Project: Education and Training on Digital and Data competences for Maritime industry

Industry reports on Data Science Analytics and Data enabled skills demand

- 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
  - 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) - Edited
  - https://public.dhe.ibm.com/common/ssi/ecm/im/en/iml14576usen/IML14576USEN.PDF
  - DSA enabled jobs takes 45-58 days to fill: 5 days longer than average
  - Commonly required work experience 3-5 yrs

Influenced by EDISON

EDSF: Derivatives and Citations

- Evaluation of EDISON's Data Science Competency Framework through a Comparative Literature Analysis (2021)
  - https://par.nsf.gov/servlets/purl/10314377
  - https://www.aimsciences.org/article/doi/10.3934/fods.2021031
- Data Science in Perspective, Paper 2021 by Rogério Rossi https://arxiv.org/ftp/arxiv/papers/2201/2201.05852.pdf
- Toward Foundations for Data Science and Analytics: A Knowledge Framework for Professional Standards, *by Usama Fayyad and Hamit Hamutcu*, 30 June 2020
  - https://hdsr.mitpress.mit.edu/pub/6wx0qmkl/release/4
- Institute of Data Science Management: Certification Standard: The Credentialing Framework (2020)
  - https://www.datascienceinstitute.net/
  - https://www.datascienceinstitute.net/about-the-data-science-institute
- Computing Competencies for Undergraduate Data Science Curricula, Initial Draft, January 2019 ACM Data Science Task Force - https://dstf.acm.org/DSReportInitialFull.pdf
- EDISON at CERN eLearning wiki (Maria Dimou, 2017) - https://twiki.cern.ch/ELearning/Edison

MATES ED2MIT: Education and Training for Data Driven Maritime Industry

- ED2MIT: Education and Training for Data Driven Maritime Industry Layman report
  - https://zenodo.org/record/6728328#.Y67ggnbMK38
- Training materials developed for digital and data skills compliant with DigComp 2.2
- Three online courses delivered
  - Big Data Technologies – Data Management – Data Science Analytics Foundation