Building the Data Science Profession in Europe

Yuri Demchenko
23 October 2016

EDISON
building the data science profession

Grant 675419 (INFRASUPP-4-2015: CSA)

IBM

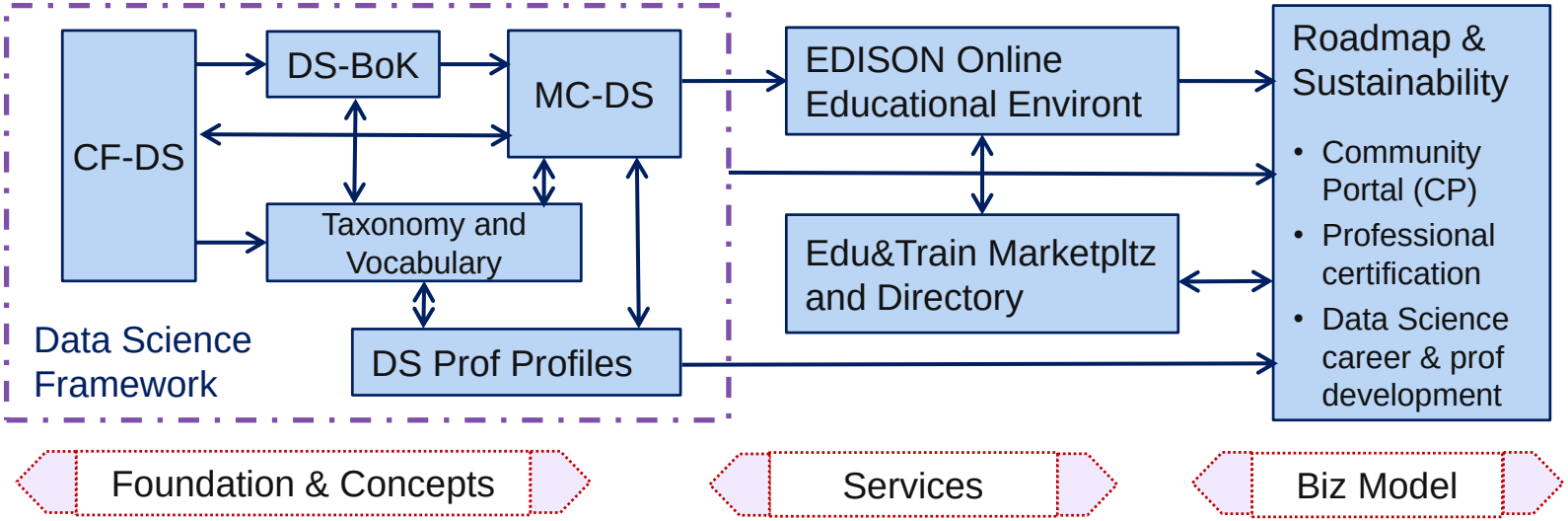
Outline

- Background and motivation
  - Demand for Data Science and European initiatives related to Digital Single Market and Digital Skills
- EDISON Data Science Framework (EDSF)
  - From Data Science Competences to Body of Knowledge and Model Curriculum
- **Data Science Competence Framework**: Essential competences and skills
- Data Science Professional profiles and taxonomy
- Educational and training focus
  - Data Science Body of Knowledge (DS-BoK)
  - Data Science Model Curriculum (MC-DS)
- Further developments to formalise EDSF and Data Science profession establishment
- Discussion and Invitation to cooperation

Grant 675419 (INFRASUPP-4-2015: CSA)

EDISON building the data science profession

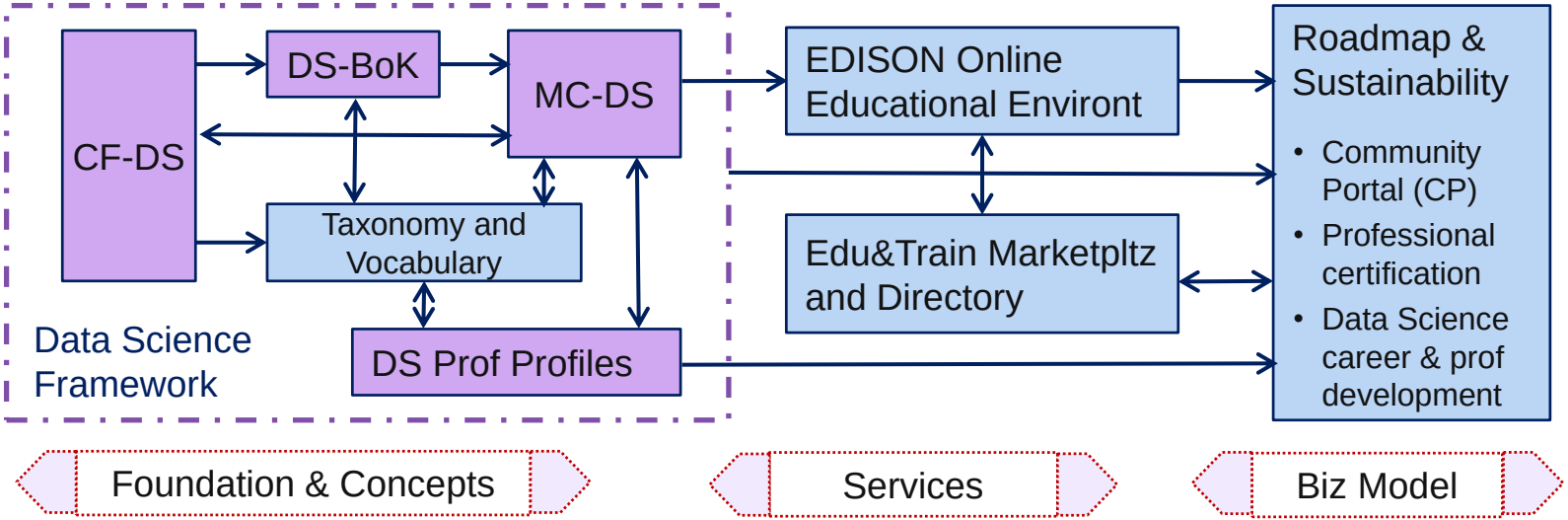
EDISON Data Science Framework (EDSF)



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 definition
- Data Science Taxonomies and Scientific Disciplines Classification

EDISON Data Science Framework (EDSF) Release 1 (Oct 2016)



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 definition
- Data Science Taxonomies and Scientific Disciplines Classification

Background Frameworks and Standards

e-CFv3.0 - European e-Competence Framework for IT

- Structured by 4 Dimensions and organizational processes
  - Competence Areas: Plan – Build – Run – Enable - Manage
  - Competences: total defined 40 competences
  - Proficiency levels: identified 5 levels linked to professional education levels
  - Skills and Knowledge

CWA 16458 (2012): European ICT Professional Profiles Family Tree

- Defines 23 ICT profiles for common ICT jobs

ESCO (European Skills, Competences, Qualifications and Occupations)

- Standard for European job market since 2016
- Expected inclusion of the Data Science occupations family – mid 2017

ACM Classification of Computer Science – CCS (2012)

- ACM Computer Science Body of Knowledge (CS-BoK) and ACM and IEEE Computer Science Curricula 2013 (CS2013)

EDISON Data Science Competence Framework

- How it was made
- 5 main Data Science competences groups
- Skills, tools and languages
- Practical use of the CF-DS
- Data Science Professional profiles and occupations family

Demanded Data Science Competences and Skills: Jobs market analysis and stakeholders

Initial Analysis (period Aug – Sept 2015) -> Continuous monitoring (in development)

– IEEE Data Science Jobs (World but majority US)
  - Collected > 120, selected for analysis > 30, verified with 100s job ads

– LinkedIn Data Science Jobs (NL)
  - Collected > 140, selected for analysis > 30, verified with 100s job ads

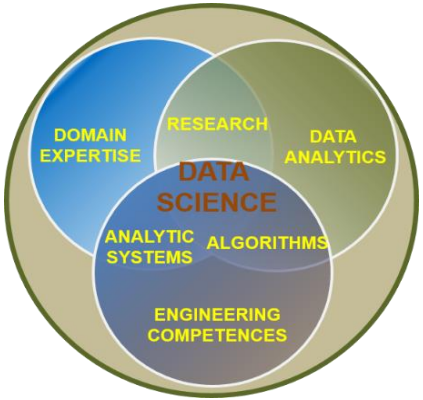
– Existing studies and reports + numerous blogs & forums

Analysis methods

– Data analytics methods: classification, clustering, expert evaluation

– Research methods: Data collection - Hypothesis – Artefact - Evaluation

Data Scientist Definition



[ref] Legacy: NIST BDWG definition of Data Science

Based on the definition by NIST Big Data WG (NIST SP1500 - 2015)

*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*

Other definitions to admit such features as

– Ability to solve variety of business problems, tell “stories”, input to decision making
– Optimize performance and suggest new services for the organisation
– Develop a special mindset and be statistically minded, ***understand raw data** and **“appreciate data as a first class product”***

- ***Data science** is the empirical synthesis of actionable knowledge and technologies required to handle data from raw data through the complete data lifecycle process.*
- ***Big Data** is the technology to build system and infrastructures to process large volume of structurally complex data in a time effective way*

Identified Data Science Competence Groups

Commonly accepted Data Science competences/skills groups include

- Data Analytics or Business Analytics or Machine Learning
- Engineering or Programming
- Subject/Scientific Domain Knowledge

EDISON identified 2 additional competence groups demanded by organisations

- **Data Management, Curation, Preservation**
- **Scientific or Research Methods and/vs Business Processes/Operations**

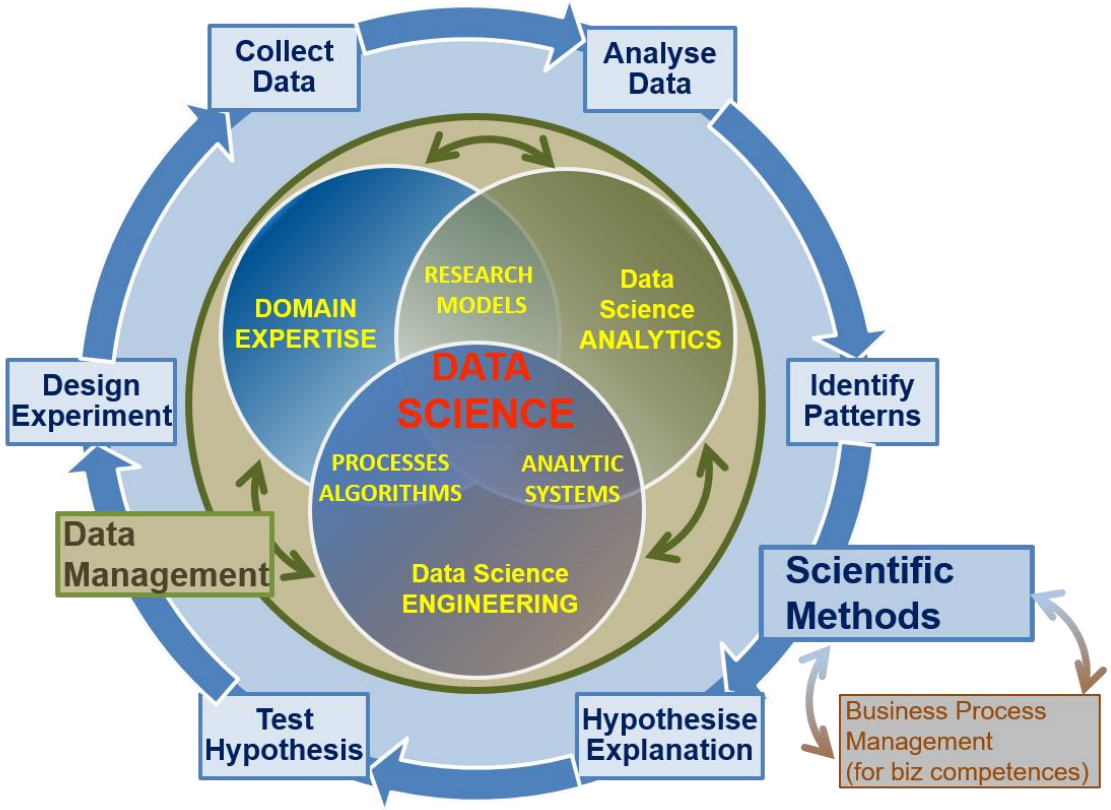
Other skills commonly recognized aka “soft skills” or “social/professional intelligence”

- Inter-personal skills or team work, cooperativeness

Important aspects of integrating Data Scientist into organisation structure

- General Data Science (and Data) **literacy** for all involved roles and management
- Common agreed and understandable way of communication and **information/data presentation**
- ***Role of Data Scientist: Provide a kind of literacy advice and guidance to organisation***

Data Science Competence Groups - Research



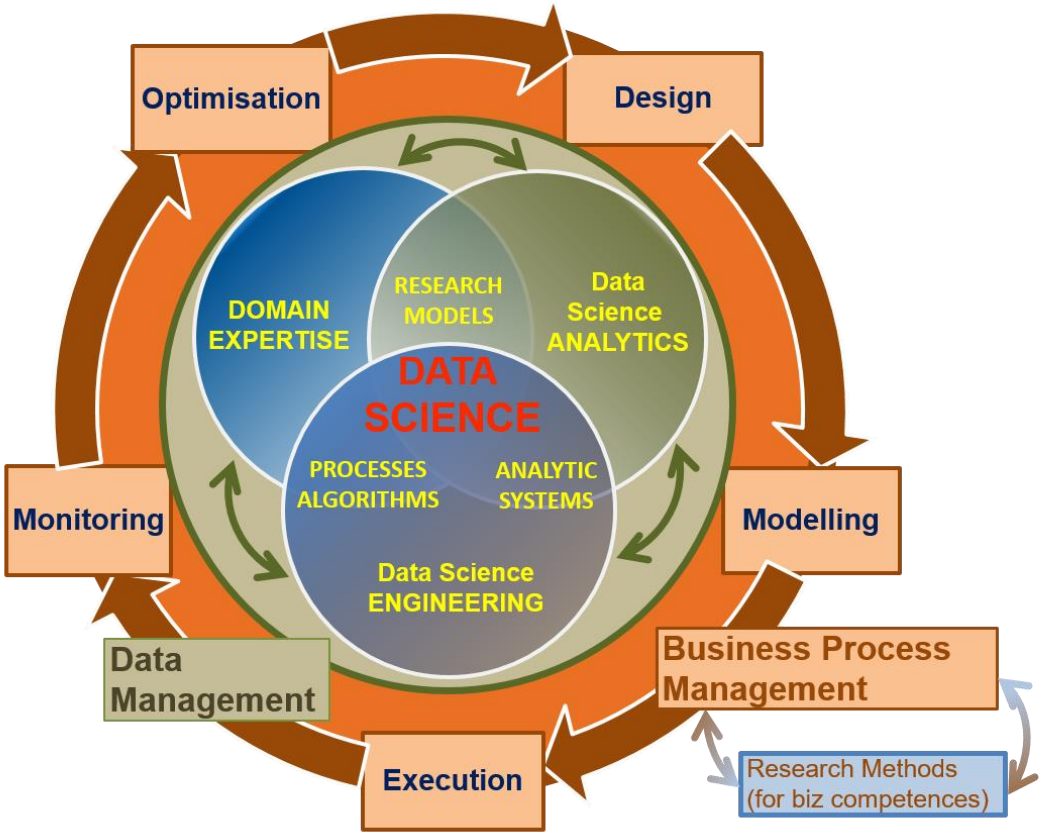
Data Science Competence includes 5 areas/groups

- Data Analytics
- Data Science Engineering
- Domain Expertise
- Data Management
- Scientific Methods (or Business Process Management)

Scientific Methods

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

Data Science Competence Groups - Business



Data Science Competence includes 5 areas/groups

- Data Analytics
- Data Science Engineering
- Domain Expertise
- Data Management
- Scientific Methods (or Business Process Management)

Business Process Operations/Stages

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

Identified Data Science Competence Groups

	Data Science Analytics (DSDA)	Data Management (DSDM)	Data Science Engineering (DSENG)	Research/Scientific Methods (DSRM)	Data Science Domain Knowledge, e.g. Business Processes (DSDK/DSBPM)
0	Use appropriate statistical techniques and predictive analytics on available data to deliver insights and discover new relations	Develop and implement data management strategy for data collection, storage, preservation, and availability for further processing.	Use engineering principles to research, design, develop and implement new instruments and applications for data collection, analysis and management	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	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
1	**DSDA01** Use predictive analytics to analyse big data and discover new relations	**DSDM01** Develop and implement data strategy, in particular, Data Management Plan (DMP)	**DSENG01** Use engineering principles to design, prototype data analytics applications, or develop instruments, systems	**DSRM01** Create new understandings and capabilities by using scientific/ research methods or similar domain related development methods	**DSBPM01** Understand business and provide insight, translate unstructured business problems into an abstract mathematical framework
2	**DSDA02** Use statistical techniq to deliver insights	**DSDM02** Develop data models including metadata	**DSENG02** Develop and apply computational solutions	**DSRM02** Direct systematic study toward a fuller knowledge or understanding of the observable facts	**DSBPM02** Participate strategically and tactically in financial decisions
3	**DSDA03** Develop specialized …	**DSDM03** Collect integrate data	**DSENG03** Develops specialized tools	**DSRM03** Undertakes creative work	**DSBPM03** Provides support services to other
4	**DSDA04** Analyze complex data	**DSDM04** Maintain repository	**DSENG04** Design, build, operate	**DSRM04** Translate strategies into actions	**DSBPM04** Analyse data for marketing
5	**DSDA05** Use different analytics	**DSDM05** Visualise cmplx data	**DSENG05** Secure and reliable data	**DSRM05** Contribute to organizational goals	**DSBPM05** Analyse optimise customer relatio

Identified Data Science *Skills/Experience* Groups

Group 1: Skills/experience related to competences

- Data Analytics and Machine Learning
- Data Management/Curation (both general and scientific)
- Data Science Engineering (hardware and software) skills
- Scientific/Research Methods or Business Process Management
- Application/subject domain related (research or business)
- Mathematics and Statistics

Group 2: 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 3: Programming and programming languages and IDE

- General and specialized development platforms for data analysis and statistics

Group 4: Soft skills or Social Intelligence

- Personal, inter-personal communication, team work, professional network

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

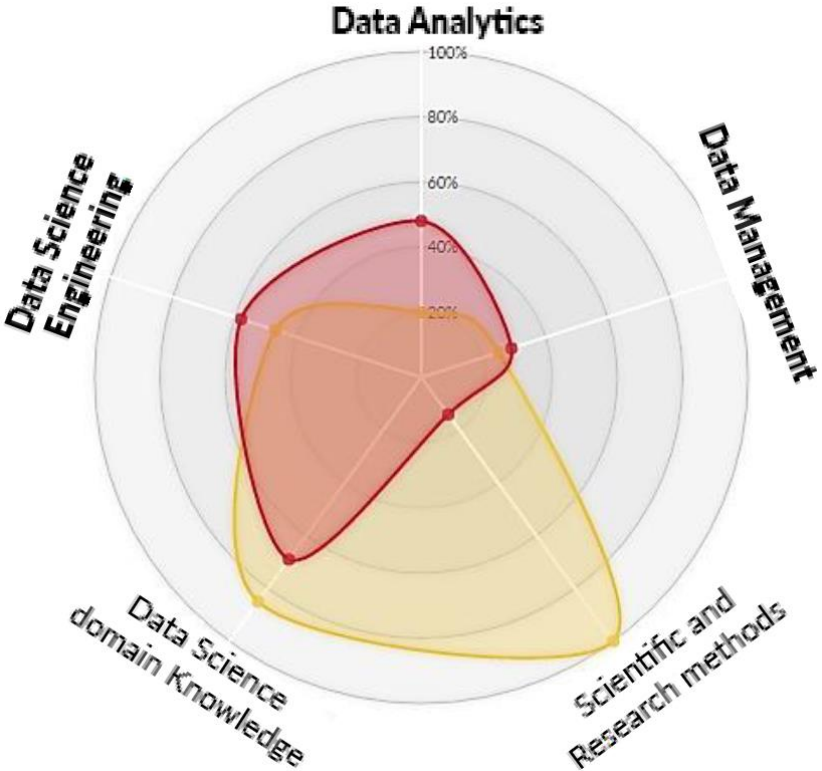
Professional certification

- In combination with DS-BoK professional competences benchmarking

Vacancy construction tool for job advertisement (for HR)

- Using controlled vocabulary and Data Science Taxonomy

Example competence benchmark result



Red polygon indicates the chosen professional profile

Yellow polygon indicates the practitioner competences/skills

Insufficient competences are shown in a bold typeface

[ref] Kim Hee et al (Frankfurt Univ) “Tailored Data Science Education using Gamification” - DTW2016 Workshop EDISON research paper

Data Science Professions Family

Professional profiles groups are defined in compliance with the ESCO taxonomy

Managers: Chief Data Officer (CDO), Data Science (group/dept) manager, Data Science infrastructure manager, Research Infrastructure manager

Professionals: Data Scientist, Data Science Researcher, Data Science Architect, Data Science (applications) programmer/engineer, Data Analyst, Business Analyst, etc.

Professionals (database): Large scale (cloud) database designers and administrators, scientific database designers and administrators

Professionals and associated support (data handling/management): Data Stewards, Digital Data Curator, Digital Librarians, Data Archivists, Support and data entry workers

Technicians and associate professionals: Big Data facilities operators, scientific database/infrastructure operators

Data Science Occupations: Extension for the ESCO (2016) (1)

22 DSP# Enumerated Data Science and Data related profiles defined by EDISON Framework

Professionals				
	Science and engineering professionals	**Data Science Professionals**	Data Science professionals not elsewhere classified	DSP04 Data Scientist
				DSP05 Data Science Researcher
				DSP08 (Big) Data Analyst
				DSP07 Data Science (Application) Programmer
				DSP09 Business Analyst
		Database and network professionals	Large scale (cloud) data storage designers and administrators	DSP14 Large scale (cloud) database designer*)
			Database designers and administrators	DSP15 Large scale (cloud) database administrator*)
			Database and network professionals not elsewhere classified	DSP16 Scientific database administrator*)
	Information and communications technology professionals	**Data Science technology professionals**	Data handling professionals not elsewhere classified	DSP12 Digital Librarian
				DSP13 Data Archivist
				DSP10 Data Steward
				DSP11 Data curator

Data Science Occupations: Extension for the ESCO (2016) (2)

Technicians and associate professionals				
	Science and engineering associate professionals	**Data Science Technology Professionals**	Data Infrastructure engineers and technicians	DSP17 Big Data facilities Operators
				DSP18 Large scale (cloud) data storage operators
			Database and network professionals not elsewhere classified	DSP19 Scientific database operator*)
Managers				
	Production and specialised services managers	**Data Science/Big Data Infrastructure Managers**		DSP01/DSP02 Data Science/Big Data Infrastructure Manager
			Research Infrastructure Managers	DSP03 RI Manager
				DSP03 RI Data storage facilities manager
Clerical support workers				
	General and keyboard clerks			
	Data handling support workers (alternative)	**Data and information entry and access**	Digital Archivists and Librarians	DSP20 Data entry/access desk/terminal workers
				DSP21 Data entry field workers
				DSP22 User support data services

Example Mapping CF-DS groups to the DSP profiles

Relevance of a competence to a DSP profile:

5 – high, 1 - low

Profile ID	Data Science Profile title	DSDA	DSDM	DSENG	DSRM	DSDK
Managers						
DSP01	Data Science (group) Manager	3	4	3	3	2
DSP02	Data Science Infrastructure Manager	2	4	4	2	2
DSP03	Research Infrastructure Manager	2	4	4	3	2
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 Applications Programmer/Engineer	4	2	5	3	4
DSP08	Data Analyst	5	3	3	3	4
DSP09	Business Analyst	5	3	3	4	5
Professional (data handling/ management)						
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
Professional (database)						
DSP14	Large scale (cloud) database designer	2	4	4	3	3
DSP15	Large scale (cloud) database administrator	2	4	3	2	3
DSP16	Scientific database administrator	2	4	3	2	3
Technicians and associate professionals						
DSP17	Big Data facilities Operator	1	4	4	2	3
DSP18	Large scale (cloud) data storage operator	1	4	3	1	1
DSP19	Scientific database operator	1	4	3	2	3
Clerical support workers (general and keyboard workers)						
DSP20	Data entry/access worker		2	1		2
DSP21	Data entry field workers		2	1		2
DSP22	User support data services		3	2		2

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

– Group Manager
– Data Science Architect
– Data Analyst
– Data Science Application programmer
– Data Infrastructure/facilities administrator/operator: storage, cloud, computation
– Data stewards

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)

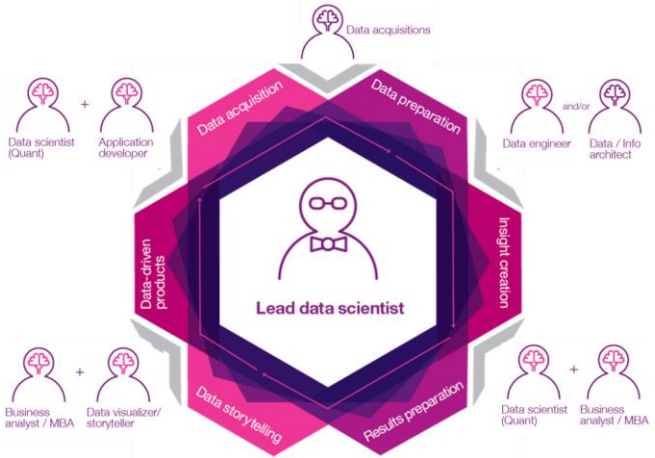
Estimated: Group of 10-12 data specialists for research institution of 200-300 research staff.

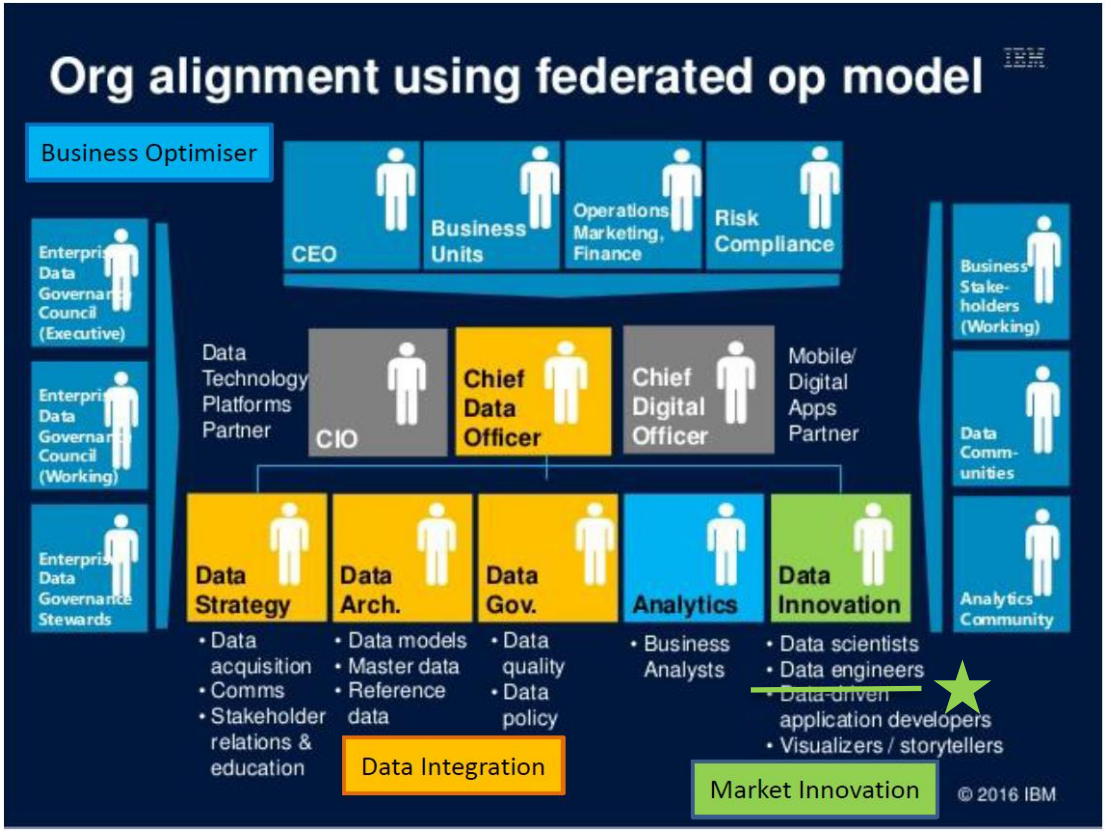
EXAMPLE: IBM emerging professions

[ref] Cortnie Abercrombie, What CEOs want from CDOs and how to deliver on it (2016) [online] http://www.slideshare.net/IBMBDA/what-ceos-want-from-cdos-and-how-to-deliver-on-it

Mastering the art of data science: How to craft cohesive teams that create business results, IBM Institute for Business Value, 2016

– Create data science teams with varied backgrounds and skills





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

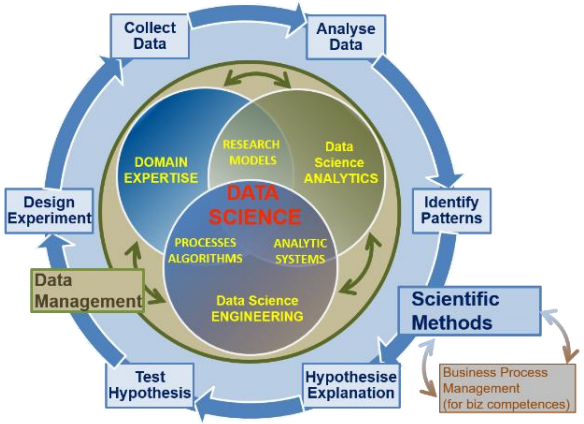
Platforms and environment

- 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
- 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: *Scientific/Research Methods group*

KAG5-DSBP: Business process management group

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
- 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)
  - Based on Bloom's Taxonomy, Outcome Based Learning, etc

Example DS-BoK Knowledge Areas definition and mapping to existing BoKs and CCS (2012)

Knowledge Area Groups (KAG)	Knowledge Areas (KA)	Suggested Knowledge Units (KU)	Mapping to CCS2012 (including suggested Data Science extensions) and existing BoKs
KAG1-DSDA: Data Analytics group (including Machine Learning, statistical methods)	Theory of computation	Design and Analysis of Algorithms	CCS2012: Theory of computation Design and analysis of algorithms Data structures design and
		Machine Learning Theory	

Knowledge Area Groups (KAG)	Knowledge Areas (KA)	Suggested Knowledge Units (KU)	Mapping to CCS2012 (including suggested Data Science extensions) and existing BoKs
KAG2-DSENG: Data Science Engineering group including Software an infrastructu engineering	Computer systems organisation for Big Data	Parallel and Distributed Computer Architecture	**CCS2012: Computer systems organization** Architectures Parallel architectures
		Computer networks: architectures	

Knowledge Area Groups (KAG)	Knowledge Areas (KA)	Suggested Knowledge Units (KU)	Mapping to CCS2012 (including suggested Data Science extensions) and existing BoKs
	Data Management and Enterprise data infrastructure	Data management, including Reference and Master Data	**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.
		Data Warehousing and Business Intelligence	
		Data storage and operations	
		Data archives/storage compliance and certification	
		Metadata, linked data, provenance	
		Data infrastructure, data registries and data factories	
		Data security and protection	
		Data governance, data quality, data Integration and Interoperability	

Example MC-DS Mapping Learning Units to DS-BoK and CCS (2012)

KAG/ LU# *)	Learning Unit (course name) [2]	Type/relevance [3]				Map to DS-BoK, CCS2012 and known BoKs	
		Tier 1	Tier 2	Elective	Pre requisite	CCS2012 based academic subjects	DS-BoK and other BoKs
	Software requirements and design					Extensions are suggested from SWEBOK	SWEBOK selected KAs • Software requirements

KAG/ LU# *)	Learning Unit (course name) [2]	Type/relevance [3]				Map to DS-BoK, CCS2012 and known BoKs	
		Tier 1	Tier 2	Elective	Pre requisite	CCS2012 based academic subjects	DS-BoK and other BoKs
	Information theory					Mathematical analysis	
	Mathematical analysis						
	Extensibility point for adding new courses						
	Artificial Intelligence					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	No specific BoK are defined
	Natural Language Processing						
	Knowledge Representation and Reasoning						
	Data mining and knowledge discovery						
	Text analysis, Data mining						
	Text analytics including statistical, linguistic, and structural techniques to analyse structured and unstructured data						
	Machine Learning theory and algorithms						
	Classification methods						

Further Developments and Actions

- Run surveys and key interviews for target communities
  - Primarily experts, visionaries and top executives at universities and companies
- Work with champion universities to practically validate the proposed EDSF
- Run a serial of the Champion Universities conferences: Spring 2017 and Summer 2017
  - Extend scope to Champions, Adopters, Ambassadors and Professionals
- Involve academic and industry experts and professional organisations to the definition of DS-BoK following from CF-DS
- Formally provide suggestions to ESCO for the definition of the Data Science professional profiles (occupations) family
- Formally provide suggestions for e-CF3.0 extensions for Data Science to CEN/PC 428
  - Involve national e-CF bodies and adopters where available
- Suggest required ACM CCS(2012) Classification extensions and proposal for Data Science curriculum definition

Discussion

Questions & Comments

Invitation to contribution and cooperation:

– Forum, EDISON Liaisons Groups, Champions Conference (Spring & Summer 2017)

EDISON project website http://edison-project.eu/

EDISON Data Science Framework Release 1 (EDSF)
http://edison-project.eu/edison-data-science-framework-edsf

Survey Data Science Competences: Invitation to participate
https://www.surveymonkey.com/r/EDISON_project_-_Defining_Data_science_profession

Links to EDSF and other EDISON resources

EDISON project website http://edison-project.eu/

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

Survey Data Science Competences: Invitation to participate
https://www.surveymonkey.com/r/EDISON_project_-_Defining_Data_science_profession

Definitions (according to e-CFv3.0)

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

– Competence vs Competency (e-CF vs ACM)
  - Competence is ability acquired by training or education (linked to learning outcome)
  - Competency is similar to skills or experience (acquired feature of a person)

Competence is not to be confused with process or technology concepts such as, ‘Cloud Computing’ or ‘Big Data’. These descriptions represent evolving technologies and in the context of the e-CF, they may be integrated as elements within knowledge and skill examples.

Knowledge in the context of competence definition is treated as something to know, to be aware of, familiar with, and obtained as a part of education.

Skills is treated as provable ability to do something and relies on the person’s experience.

Data Scientist and Subject Domain Specialist

Subject domain components

- Model (and data types)
- Methods
- Processes
- Domain specific data and presentation/visualization methods
- Organisational roles and relations

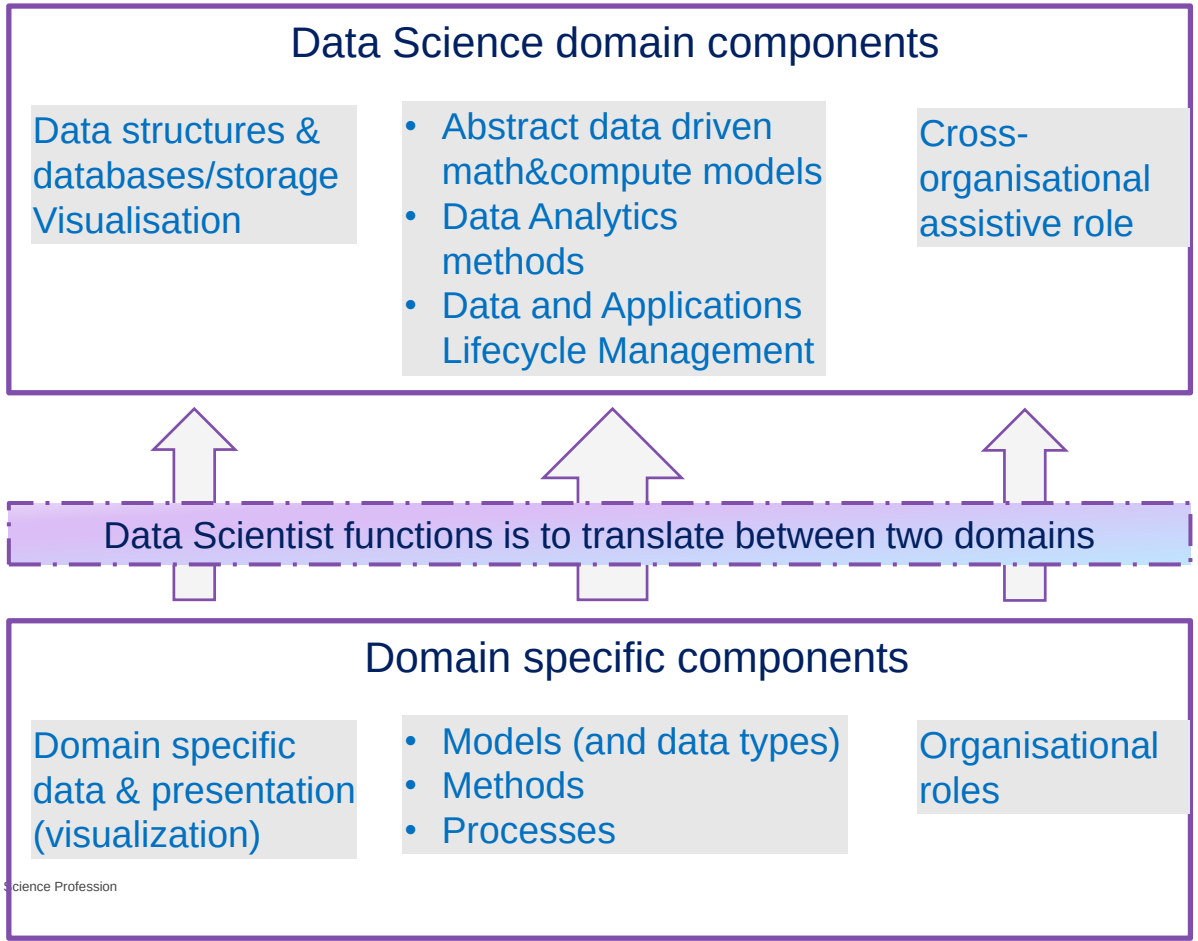
Data Scientist is an assistant to Subject Domain Specialists

- Translate subject domain Model, Methods, Processes into abstract data driven form
- Implement computational models in software, build required infrastructure and tools
- Do (computational) analytic work and present it in a form understandable to subject domain
- Discover new relations originated from data analysis and advice subject domain specialist
- Present/visualise information in domain related actionable way
- Interact and cooperate with different organizational roles to obtain data and deliver results and/or actionable data

Data Science and Subject Domains

Data Scientist role is to maintain the Data Value Chain (domain specific):

Data Integration => Organisation/Process/Business Optimisation => **Innovation**



KAG3-DSDM: *Data Management group: data curation, preservation and data infrastructure*

DM-BoK version 2 “Guide for performing data management”
– 11 Knowledge Areas

(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
(11) Data Quality

Other Knowledge Areas motivated by RDA, European Open Data initiatives, European Open Data Cloud

(12) PID, metadata, data registries
(13) Data Management Plan
(14) Open Science, Open Data, Open Access, ORCID
(15) Responsible data use

- Highlighted in red: Considered (Research) Data Management literacy (minimum required knowledge)

Research Data Management Model Curriculum – Part of the EDISON Data Literacy Training

A. Use cases for data management and stewardship
– Preserving the Scientific Record

B. Data Management elements (organisational and individual)
– Goals and motivation for managing your data
– Data formats
– Creating documentation and metadata, metadata for discovery
– Using data portals and metadata registries
– Tracking Data Usage
– Handling sensitive data
– Backing up your data
– Data Management Plan (DMP) - to be a part of hands on session

C. Responsible Data Use Section (Citation, Copyright, Data Restrictions)
D. Open Science and Open Data (Definition, Standards, Open Data use and reuse, open government data)
– Research data and open access
– Repository and self- archiving services
– ORCID identifier for data
– Stakeholders and roles: engineer, librarian, researcher
– Open Data services: ORCID.org, Altmetric Doughnut, Zenodo

Collaboration with the Research Data Alliance (RDA) on developing model curriculum on Research Data Literacy:
- Modular, Customisable, Localised, Open Access
- Supported by the network of trainers via resource swap board

E. Hands on:
a) Data Management Plan design
b) Metadata and tools
c) Selection of licenses for open data and contents (e.g. Creative Common and Open Database)

Demand for Data Science and data related professions

McKinsey Global Institute on Big Data Jobs (2011 and 2014)

IDC Report on European Data Market (2015)

European initiatives and directives

- Digitising European Industry: Reaping the full benefits of a **Digital Single Market**. COM(2016) 180 final, Brussels, 19.4.2016
- European Cloud Initiative - Building a competitive data and knowledge economy in Europe, COM(2016) 178 final, Brussels, 19.4.2016
- A New Skills Agenda for Europe, COM(2016) 381 final Brussels, 10.6.2016
- Launch **Digital Skills and Jobs Coalition** (1st December 2016, Brussels) to develop comprehensive national digital skills strategies by mid-2017

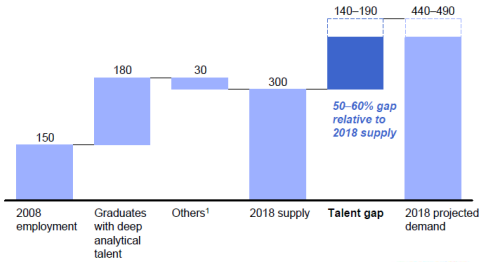
HLEG report on European Open Science Cloud (Oct. 2016) identified need for data experts and data stewards

- Estimation: More than 80,000 data stewards (1 per every 20 scientists)
- Core data experts need to be trained and their career perspective improved

OECD on Skills for a Digital World (25 May 2016)

- Demand for new type of *"dynamic self-re-skilling workforce"*

Demand for Data Science and data related professions

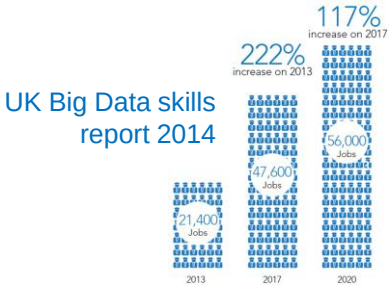


McKinsey Global Institute on Big Data Jobs (2011)
http://www.mckinsey.com/mgi/publications/big_data/index.asp

– Estimated gap of 140,000 - 190,000 data analytics skills by 2018

UK Big Data skills report 2014

– 6400 UK organisations with 100+ staff will have implemented Big Data Analytics by 2020
– Increase of Big Data jobs from 21,400 (2013) to 56,000 (2017)



UK Big Data skills report 2014

IDC Report on European Data Market (2015)

– Number of data workers 6.1 mln (2014) - increase 5.7% from 2013
– Average number of data workers per company 9.5 - increase 4.4%
– Gap between demand and supply 509,000 (2014) or 7.5%

HLEG report on European Open Science Cloud (2016) identified need for data experts and data stewards

– Recommendation: Allocate 5% grant funding for Data management and preservation
– Estimation: More than 80,000 data stewards (1 per every 20 scientists)
– Core data experts need to be trained and their career perspective improved

Recent European Commission Initiatives 2016

Digitising European Industry: Reaping the full benefits of a **Digital Single Market**. COM(2016) 180 final, Brussels, 19.4.2016

The need for new multidisciplinary and digital skills in particular Data Scientist

– Expected rapidly growing demand will lead to more than 800 000 unfilled vacancies by 2020

European Cloud Initiative - Building a competitive data and knowledge economy in Europe, COM(2016) 178 final, Brussels, 19.4.2016

European Open Science Cloud (EOSC) and European digital research and data infrastructure

– To offer 1.7 million European researchers and 70 million professionals in science and technology open and seamless services for ***storage, management, analysis and re-use*** of research data

Address growing demand and shortage of data-related skills

A New Skills Agenda for Europe, COM(2016) 381 final Brussels, 10.6.2016

Addresses the need for digital and complementary skills, ensure young talents flow into data driven research and industry

Launch **Digital Skills and Jobs Coalition** (1st December 2016, Brussels) to develop comprehensive national digital skills strategies by mid-2017

OECD on Digital Economy

Demand for new type of *“dynamic self-re-skilling workforce”*

Continuous learning and professional development to become a shared responsibility of workers and organisations

[ref] SKILLS FOR A DIGITAL WORLD, OECD, 25-May-2016
http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/ICCP/IIS(2015)10/FINAL&docLanguage=En