

# EDISON Data Science Framework (EDSF): Facilitating Data Science Curricula Development and organisational capacity building

## EDSF Motivation and Background: Preliminary Research, Studies, Reports

Yuri Demchenko, EDISON Project

University of Amsterdam

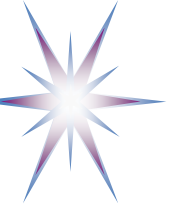
HPCS2019

15 July 2019, Dublin



**mates**





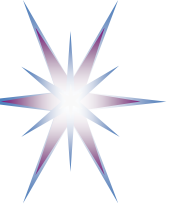
# Workshop Agenda

## **Part 1 13:30 – 15:00**

1. Introduction and bootstrapping (10 min)
  - EDISON Data Science Framework and EDISON Community Initiative
  - Workshop goals and attendees interests
2. Background information and how the EDSF has been done (20 min)
  - EU and International studies on data related competences and skills
  - EU standards, projects, initiatives, associations
3. EDISON Data Science Framework (EDSF) in details and walk through customised curriculum design (60 min)
  - EDSF components: Competence Framework (CF-DS), Body of Knowledge (DS-BoK), Model Curriculum (MC-DS), Data Science Professional Profiles (DSPP)
  - Walk through customised curriculum design: from target professional groups or competences to curriculum suggestions

## **Part 2 15:00-15:30**

1. Data Scientist and Data Steward vacancy and CV analysis and matching from indeed.com – Interactive session and practice (in groups)
  - Analysis of Data Science and Data Steward CVs and Job vacancies
  - Matching exercise
2. Assessment and discussion on Data Science soft skills: workplace and 21st Century
3. Closing: comments, questions

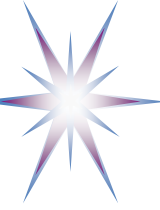


# Outline

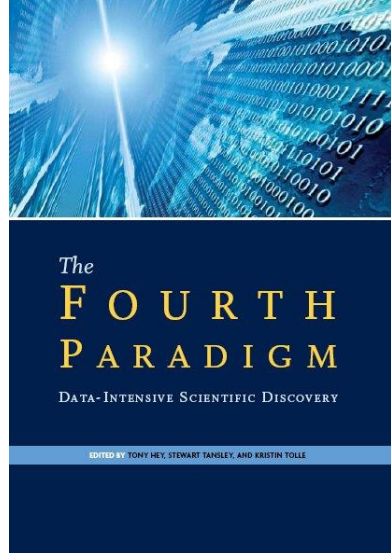
- Data driven research and demand for new skills
- EU Specific Drivers and Recommendations
- International studies and reports
- OECD and UN on Digital Economy and Data Literacy
  - PwC study on Millennials at work (2016)
- Europe: Digital technologies adoption 2016-2017
- Data Driven Victories and Failures – Politics
- EDSF background standards: eCF3.0, CWA 16458 (2012): European ICT Professional Profiles, ESCO



This work is licensed under the Creative Commons Attribution 4.0 International License.  
To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>



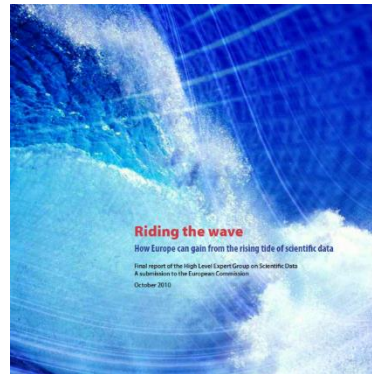
# Visionaries and Drivers: Seminal works, High level reports, Activities



## **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>



Research Data Sharing  
without barriers

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



## **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>

## **HLEG report on European Open Science Cloud**

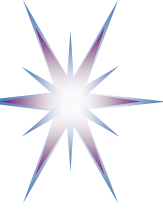
(October 2016)

[https://ec.europa.eu/research/openscience/pdf/realising\\_the\\_european\\_open\\_science\\_cloud\\_2016.pdf](https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf)



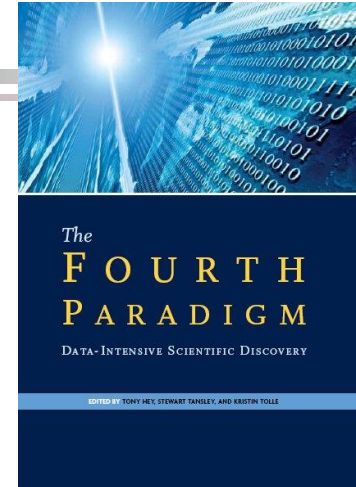
## **Emergence of Cognitive Technologies**

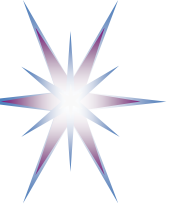
(IBM Watson, Cortana and others)



# The Fourth Paradigm of Scientific Research

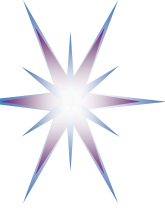
1. Theory, hypothesis and logical reasoning
2. Observation or Experiment, e.g.
  - Newton observed apples falling to design his theory of mechanics
  - Galileo Galilei made experiments with falling objects from the Pisa leaning tower
3. Simulation of theory or model
  - Digital simulation can prove theory or model
4. Data-driven Scientific Discovery (aka Data Science)
  - More data beat hypothesized theory
  - e-Science as computing and Information Technologies empowered science
5. Computer-human - driven science?
  - Machine discovers new patterns and formulates hypothesis in one or multiples knowledge spaces
  - Scientist validates and designs additional texts or experiments





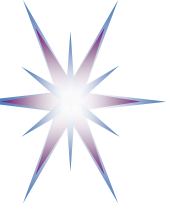
# EU Specific Drivers and Recommendations

---



# Riding the wave (2010): How Europe can gain from the rising tide of scientific data.

- “Unlocking the full value of scientific data”
  - Neelie Kroes, *Vice-President of the European Commission, responsible for the Digital Agenda*
- Just how students will be trained in the future, or how the **profession of “data scientist”** will be developed, are among the questions the resolution of which is still evolving and will present intellectual challenges for both privately and publicly supported research.
  - John Wood, HLEG Chair
- Vision 2030: “Our vision is a scientific e-Infrastructure that supports seamless access, use, re-use and trust of data. In a sense, the physical and technical infrastructure becomes invisible and the data themselves become the infrastructure.”
- Proposed set of actions
  - **4. Train a new generation of data scientists, and broaden public understanding**  
We urge that the European Commission promote, and the member-states adopt, new policies to foster the development of advanced-degree programmes at our major universities for the emerging field of data scientist. We also urge the member-states to include data management and governance considerations in the curricula of their secondary schools, as part of the IT familiarisation programmes that are becoming common in European education.



# The Data Harvest (2014): How sharing research data can yield knowledge, jobs and growth

- Planning the data harvest – John Wood
- The era of data driven science
- We want the right minds, with the right data, at the right time. That's a tall order that requires change in:
  - The way science works and scientists think
  - How scientific institutions operate and interact
  - How scientists are trained and employed

## Recommendation 2

- DO promote **data literacy** across society, from researcher to citizen. Embracing these new possibilities requires training and cultural education – inside and outside universities. **Data science** must be promoted
  - A first-class science: Data sharing provides the foundation for a new branch of science.
  - Data education: Training in the use, evaluation and responsible management of data needs to be embedded in curricula, across all subjects, from primary school to university.
  - Training within EU projects
  - Government and public sector training

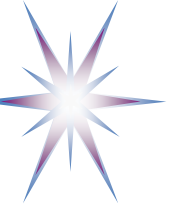


# HLEG EOSC Report Essentials – Core Data Experts [ref]

- **Core Data Experts** is a new class of colleagues with core scientific professional competencies and the communication skills to fill the gap between the two cultures.
  - **Core data experts** are neither computer savvy research scientists nor are they hard-core data or computer scientists or software engineers.
  - They should be technical data experts, though proficient enough in the content domain where they work routinely from the very beginning (experimental design, proposal writing) until the very end of the data discovery cycle
  - Converge two communities:
    - Scientists need to be educated to the point where they hire, support and respect Core Data Experts
    - Data Scientists (Core Data Experts) need to bring the value to scientific research and organisations
- Implementation of the EOSC needs to include instruments to help train, retain and recognise this expertise,
  - In order to support the 1.7 million scientists and over 70 million people working in innovation.

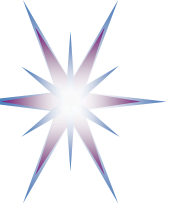


[ref] [https://ec.europa.eu/research/openscience/pdf/realising\\_the\\_european\\_open\\_science\\_cloud\\_2016.pdf](https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf)



# EOSC Report Recommendations – Implementation on training and skills

- **I2.1: Set initial guiding principles to kick-start the initiative as quickly as possible.**
  - A first cohort of core data experts should be trained to translate the needs for data driven science into technical specifications to be discussed with **hard-core data scientists and engineers**.
  - This new class of core data experts will also help translate back to the **hard- core scientists** the technical opportunities and limitations
- **I3: Fund a concerted effort to develop core data expertise in Europe.**
  - Substantial training initiative in Europe to locate, create, maintain and sustain the required core data expertise.
  - **By 2022, to train (hundreds of thousands of) certified core data experts** with a demonstrable effect on ESFRI/e-INFRA activities and prospects for long-term sustainability of this critical human resource
    - Consolidate and further develop assisting material and tools for Data Management Plans and Data Stewardship plans (including long-term preservation in FAIR status)
- **I7: Provide a clear operational timeline to deal with the early preparatory phase of the EOSC.**
  - **Define training needs for the necessary data expertise and draw models for the necessary training infrastructure**



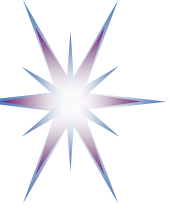
# Initiatives: GO FAIR and IFDS

- Global Open FAIR
  - Findable – Accessible – Interoperable - Reusable
- **IFDS – Internet of FAIR Data and Services = EOSC**
- GO FAIR implementation approach
  - GO-TRAIN: Training of data stewards capable of providing FAIR data services
  - FAIRdICT: Top Sector Health collaboration with top team ICT
- A critical success factor is availability of expertise in data stewardship
  - Training of a new generation of FAIR data experts is urgently needed to provide the necessary capacity

<https://www.dtls.nl/fair-data/>

<https://www.dtls.nl/fair-data/go-fair/>

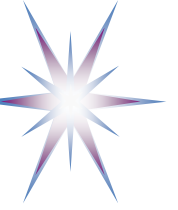
<https://www.dtls.nl/fair-data/fair-data-training/>



# Industrial Data Space: Industrial Data

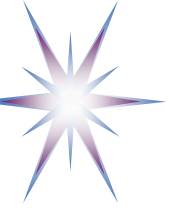
STREAM data principles for industrial and commoditised data

- [S] Sovereign
- [T] Trusted
- [R] Reusable
- [E] Exchangeable
- [A] Actionable
- [M] Measurable
- Other data properties: Important **to commoditise** data
  - Quality, Valuable, Auditable/Trackable, Brandable, Authentic
  - Interoperable, Findable, Accessible, not-Rival, Composable
  - Ownership and IPR
- Leverages FAIR principles for research data
  - Findable – Accessible – Interoperable - Reusable



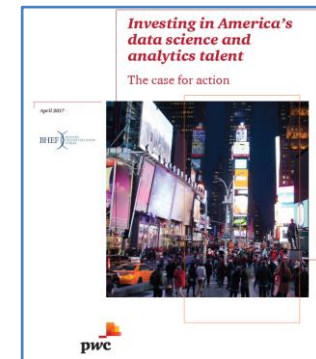
# International and EU studies on data-driven skills

- Final Report on European Data Market Study by IDC (Feb 2017)
- PwC and BHEF report “Investing in America’s data science and analytics talent: The case for action” (April 2017)
- Burning Glass Technology, IBM, and BHEF report (2017)
- PwC study: Millennials at work (2016)



# Industry reports on Data Science Analytics and Data enabled skills demand

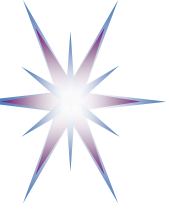
- Final Report on European Data Market Study by IDC (Feb 2017)
  - The EU data market in 2016 estimated EUR 60 Bln (growth 9.5% from EUR 54.3 Bln in 2015)
    - Estimated EUR 106 Bln in 2020
  - Number of data workers 6.1 mln (2016) - increase 2.6% from 2015
    - Estimated EUR 10.4 million in 2020
  - 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
- Burning Glass Technology, IBM, and BHEF report “The Quant Crunch: How the demand for Data Science Skills is disrupting the job Market” (April 2017)
  - <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



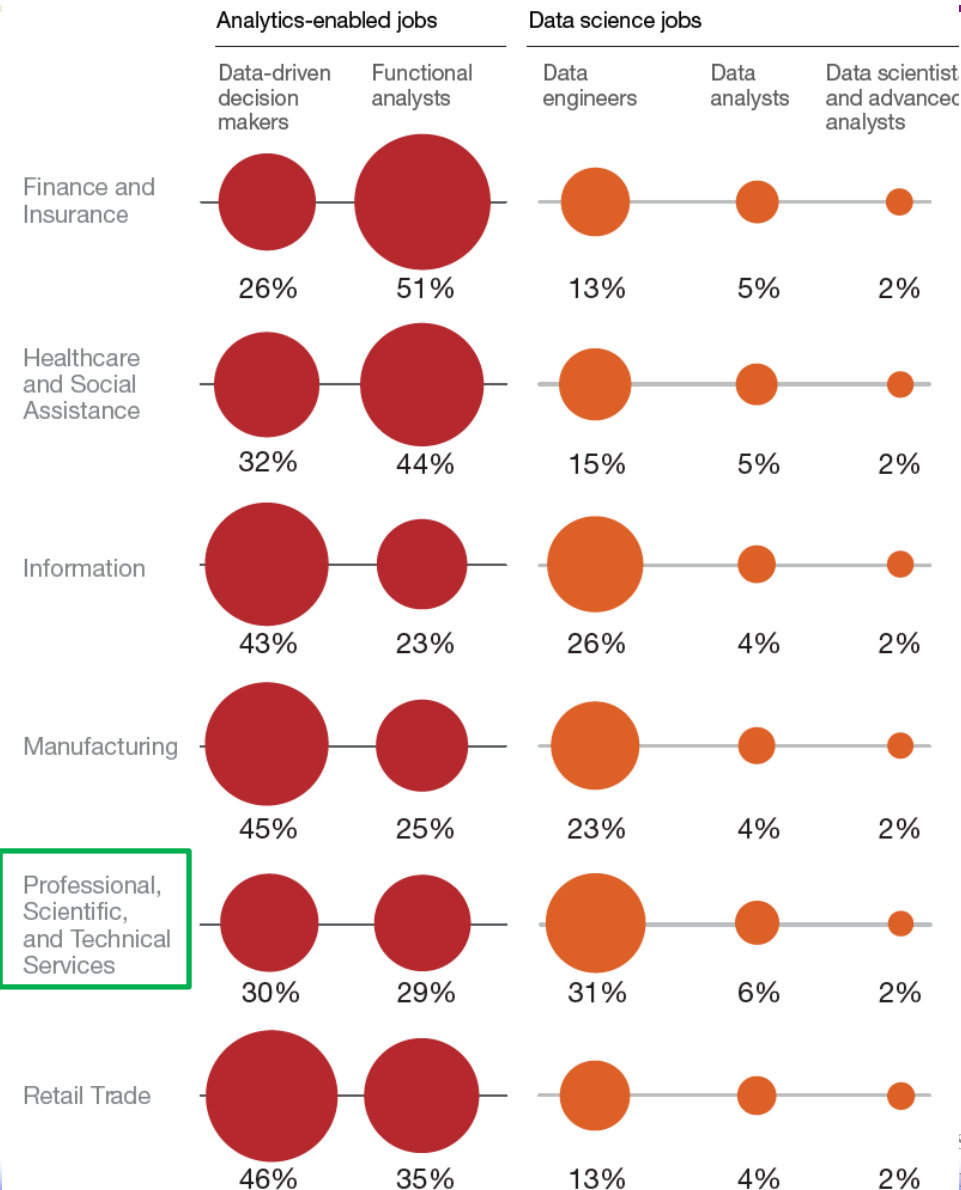
 Citing EDISON and EDSF



 Influenced by EDISON

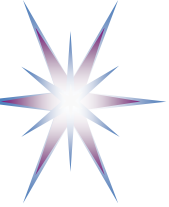


# PwC&BHEF: Demand for DSA enabled jobs

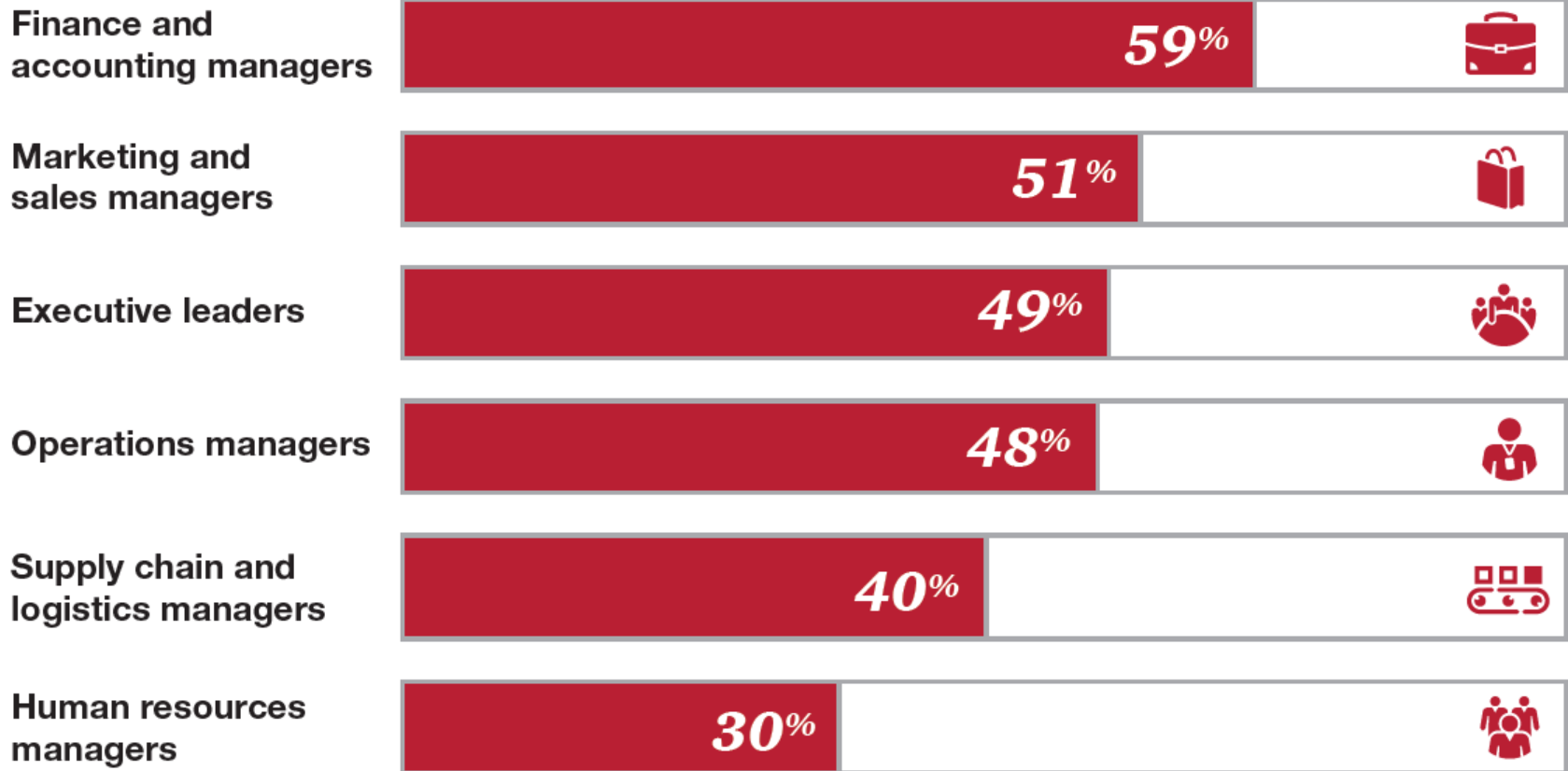


Demand for business people with analytics skills, not just data scientists

- 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



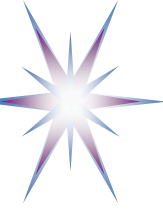
# PwC&BHEF: Data Science and Data Analytics Competences for Managers and Decision Makers



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

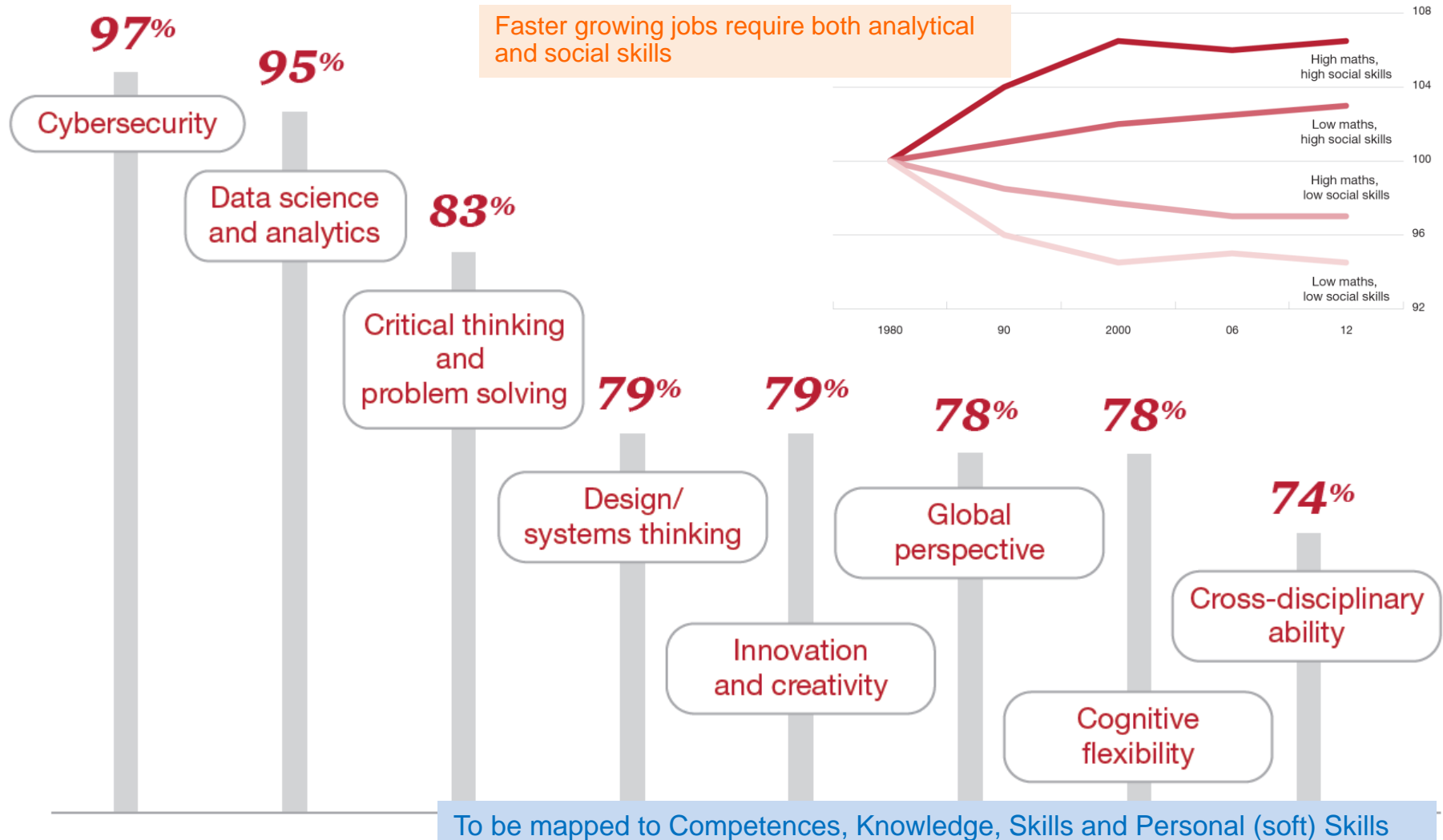
- Source: BHEF and Gallup, *Data Science and Analytics Business Survey* (December 2016).





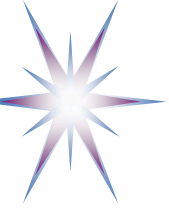
# PwC&BHEF: Skills that are tough to find

Figure 8: The fastest-growing job areas require both analytical and social skills  
US, change in employment skills by skills required, 1980 = 100



Source: Business Roundtable (2017).

EDSF Tutorial 2019



# PwC&BHEF: Data Science and Analytics skills, by 2021: The supply-demand challenge

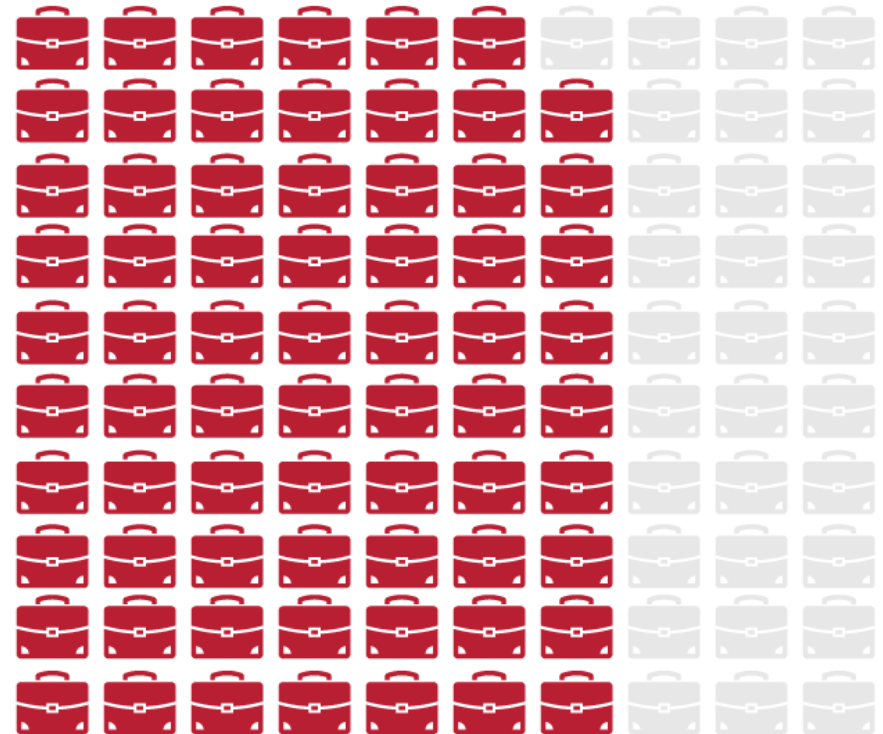
Student supply



**23%**

of educators say all graduates will have data science and analytics skills

Employer demand



**69%**

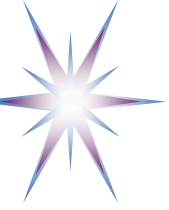
of employers say they will prefer job candidates with these skills over ones without



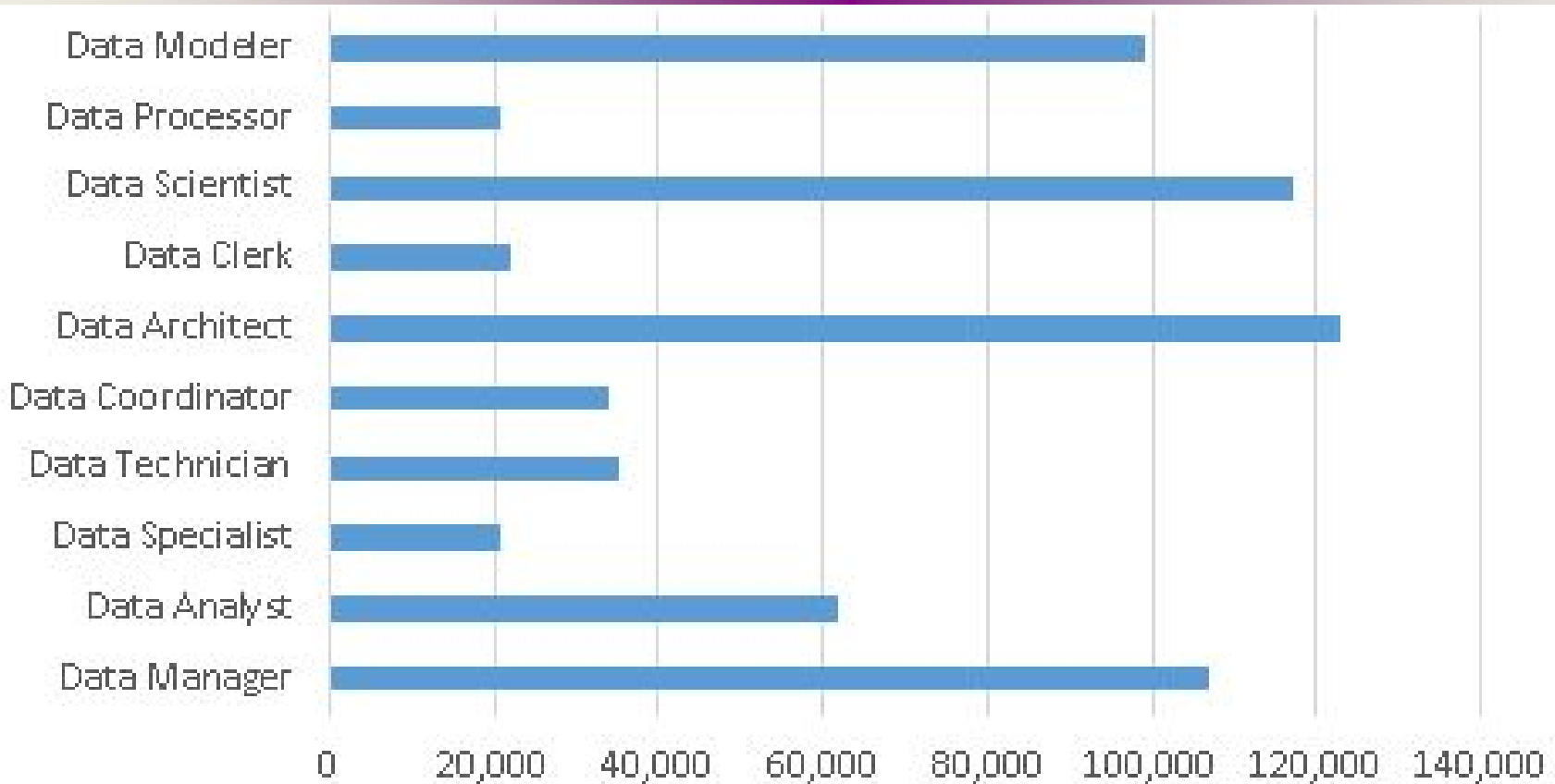
# IBM&BGT: DSA Jobs Time to Fill and Salary (2016-2017)

DSA Framework Category	Top Industries (by Demand Volume)	Average Time to Fill (Days)	Average Annual Salary
Data-Driven Decision Makers	Professional Services	50	\$96,845
	Finance & Insurance	37	\$98,131
	Manufacturing	43	\$93,641
Functional Analysts	Finance & Insurance	35	\$71,937
	Professional Services	48	\$69,135
	Manufacturing	39	\$72,571
Data Systems Developers	Professional Services	51	\$82,447
	Finance & Insurance	35	\$87,039
	Manufacturing	43	\$81,138
Data Analysts	Professional Services	47	\$74,917
	Finance & Insurance	31	\$83,209
	Manufacturing	41	\$72,742
Data Scientists & Advanced Analysts	Professional Services	51	\$97,457
	Finance & Insurance	43	\$106,610
	Manufacturing	45	\$92,543
Analytics Managers	Finance & Insurance	38	\$113,754
	Professional Services	53	\$107,185
	Manufacturing	40	\$106,926

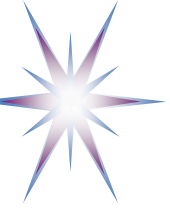
- On average, DSA jobs in Professional Services remain open for 53 days, eight days longer than the overall DSA average. (IBM, BGT 2017 Study)



# Closer look at Data related Jobs and Salaries (2016)



Source: The Job Market for Data Professionals, by Robert R Downs, SciDataCon2016  
<http://www.scidatacon.org/2016/sessions/98/poster/51/>



# OECD and UN on Digital Economy and Data Literacy

OECD (Organisation for Economic Cooperation and Development)

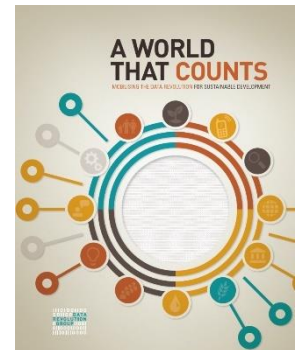
- Demand for new type of *“dynamic self-re-skilling workforce”*
- Continuous learning and professional development to become a shared responsibility of workers and organisations

[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](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/ICCP/IIS(2015)10/FINAL&docLanguage=En)

UN

- Data Revolution Report "A WORLD THAT COUNTS" Presented to Secretary-General (2014)  
<http://www.undatarevolution.org/report/>
- Data Literacy is defined as key for digital revolution and Industry 4.0
- **Data literacy** = critically analyse data collected and data visualised



# PwC study: Millennials at work (2016) - 1

## Confirmed results of previous studies:

- Loyalty-lite to company
  - The power of employer brands and the waning importance of corporate responsibility
- A time of compromise: benefit from individual package negotiation
- Development and work/life balance are more important than position or salary
  - Work/life balance and diversity promises are not being kept
- Financial reward is secondary but cash bonuses are valued
- A techno generation avoiding face time and prefer network communication
- Moving up the ladder faster expectation but often not confirmed by hard work required
- Generational communication but not without tensions



# PwC study: Millennials at work (2016) - 2

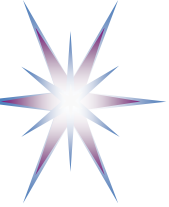
- What organisation is an attractive employer?
  - Opportunities for career progression
  - Competitive wages/other financial incentives
  - Excellent training/development programmes
- Factors most influenced decision to accept your current job?
  - The opportunity for personal development
  - The reputation of the organisation
  - The role itself
- Which three benefits would you most value from an employer?
  - Training and development
  - Flexible working hours
  - Cash bonuses



## What can employers do?

Business leaders and HR need to work together to:

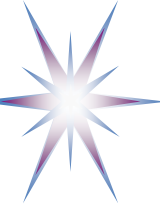
- Understand this generation
- Get the 'deal' right
- Help millennials grow
- Feedback, feedback and more feedback
- Set them free
- Encourage learning
- Allow faster advancement
- Expect millennials to go



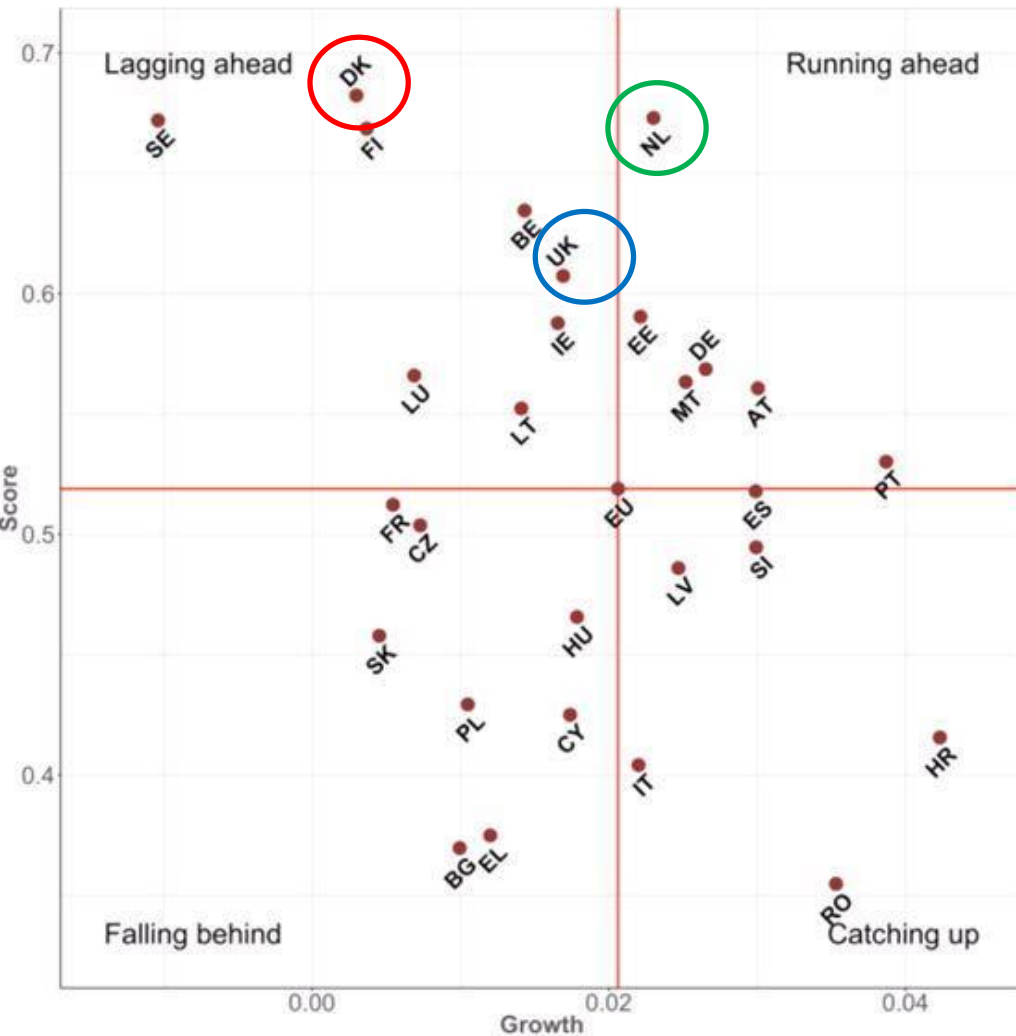
# Europe: Digital technologies adoption 2016-2017

---

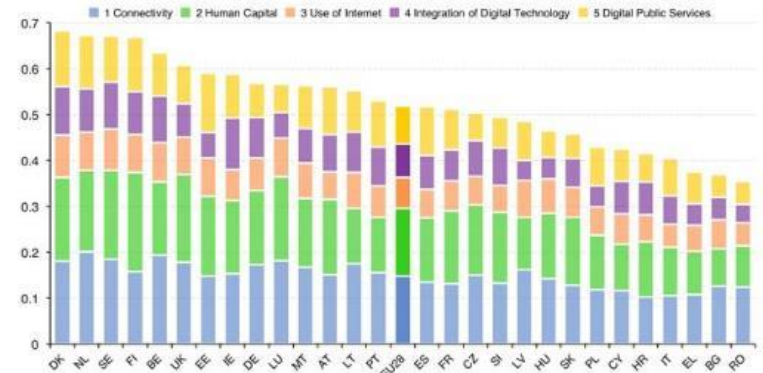




# Europe: Digital Economy and Society Index (EU 2015-2016)



What is the ranking in 2016?

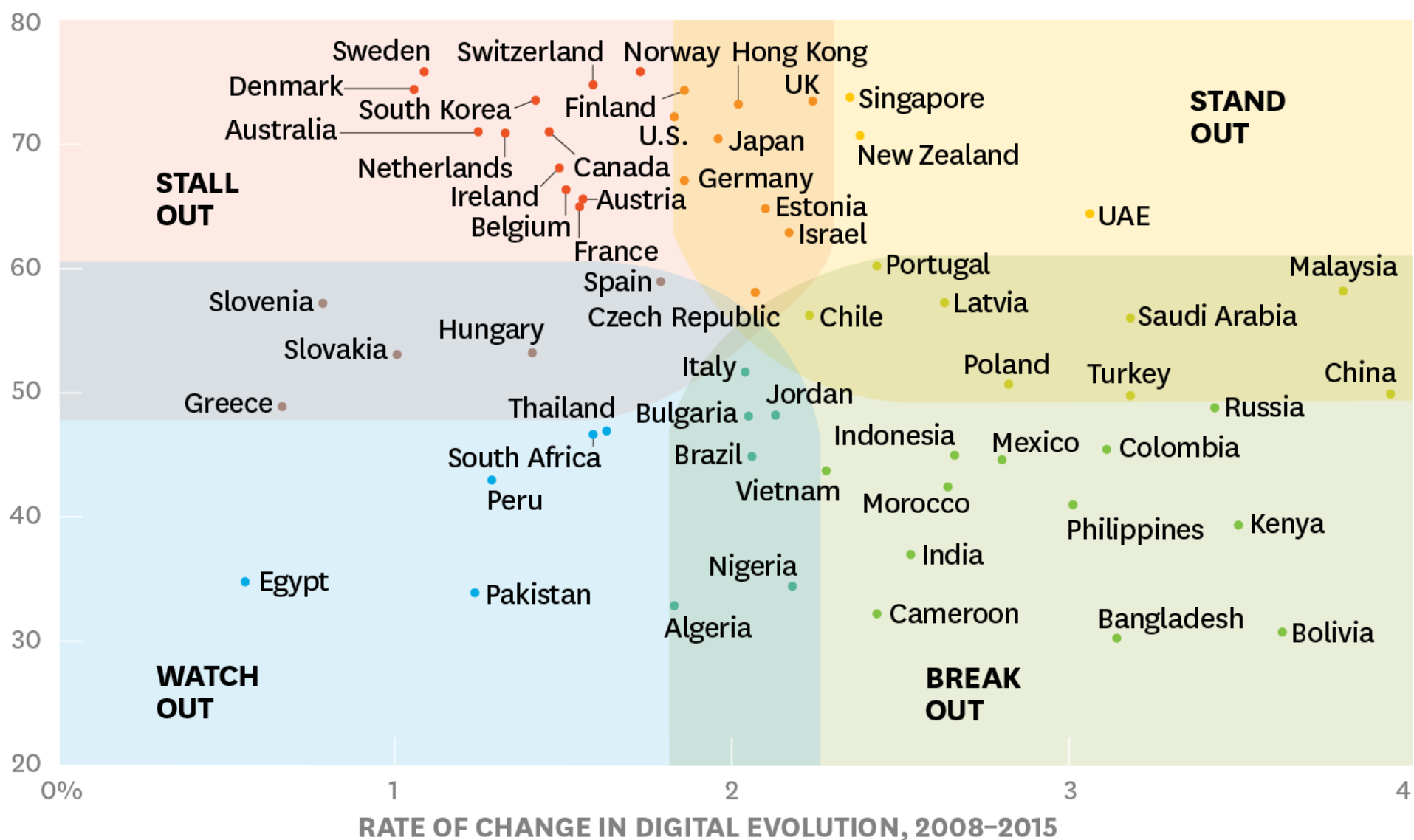


[ref] [http://europa.eu/rapid/press-release\\_MEMO-16-385\\_en.htm](http://europa.eu/rapid/press-release_MEMO-16-385_en.htm)

# Plotting the Digital Evolution Index, 2017

Where the digital economy is moving the fastest, and where it's in trouble.

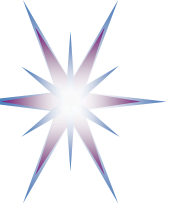
HOW COUNTRIES SCORED ACROSS FOUR DRIVERS ON THE DIGITAL EVOLUTION INDEX (OUT OF 100)



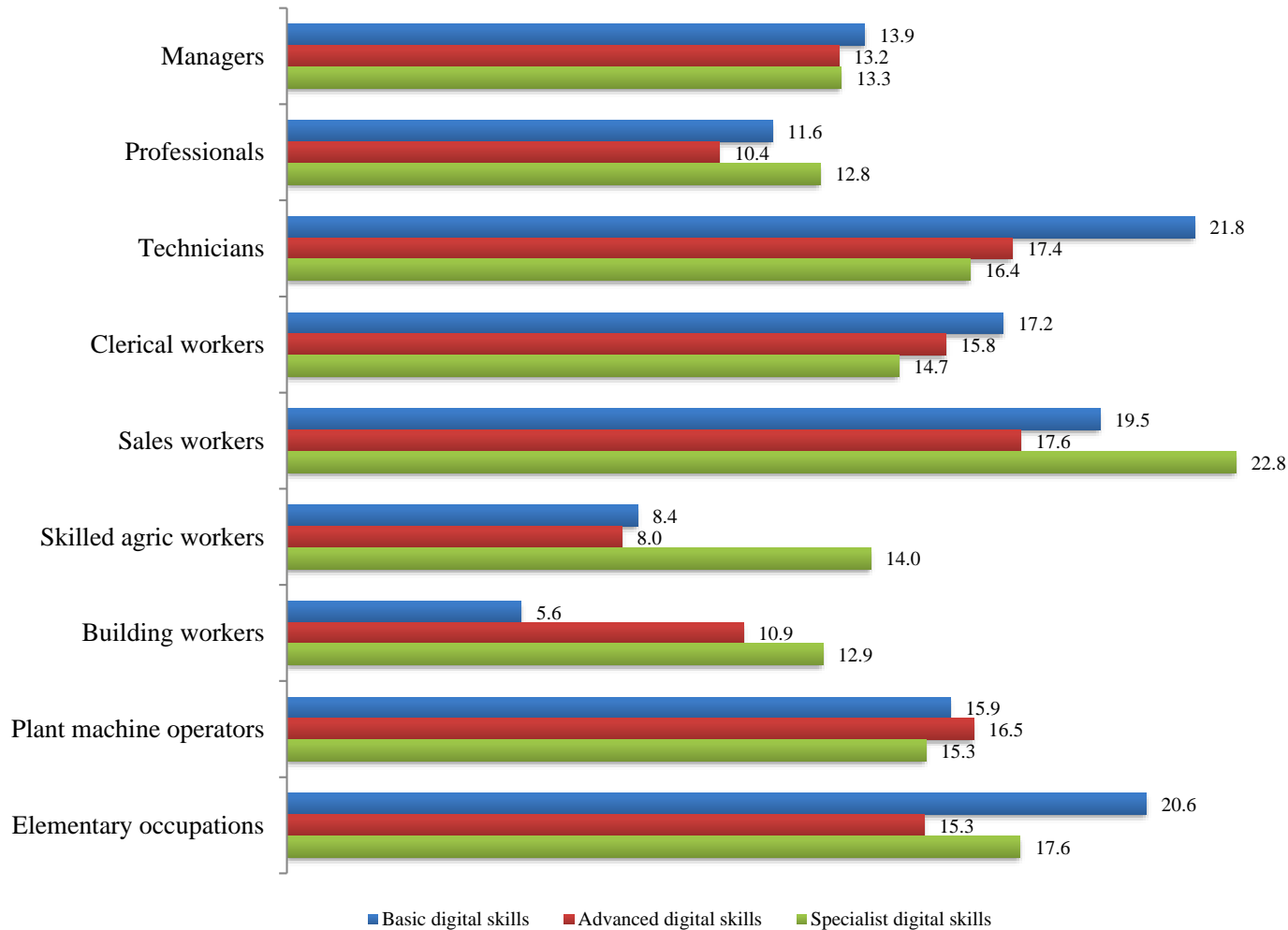
[Ref] <https://hbr.org/2017/07/60-countries-digital-competitiveness-indexed>

SOURCE DIGITAL EVOLUTION INDEX 2017, THE FLETCHER SCHOOL AT TUFTS UNIVERSITY AND MASTERCARD

© HBR.ORG

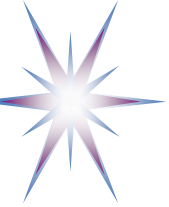


# Deeper look: Digital skills gaps density by occupation and type of digital skills, EU28 (%)



ICT for work: Digital skills in the workplace, Digital Single Market, Reports and studies, May 2017

<https://ec.europa.eu/digital-single-market/en/news/ict-work-digital-skills-workplace>

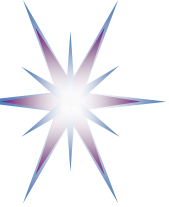


# Workplaces reporting having taken action to tackle digital skill gaps by type of action undertaken, EU28 (% of workplaces with digital skill gaps which undertook actions)



ICT for work: Digital skills in the workplace, Digital Single Market, Reports and studies, May 2017

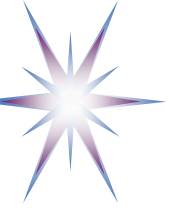
<https://ec.europa.eu/digital-single-market/en/news/ict-work-digital-skills-workplace>



# Data Driven Victories and Failures - Politics

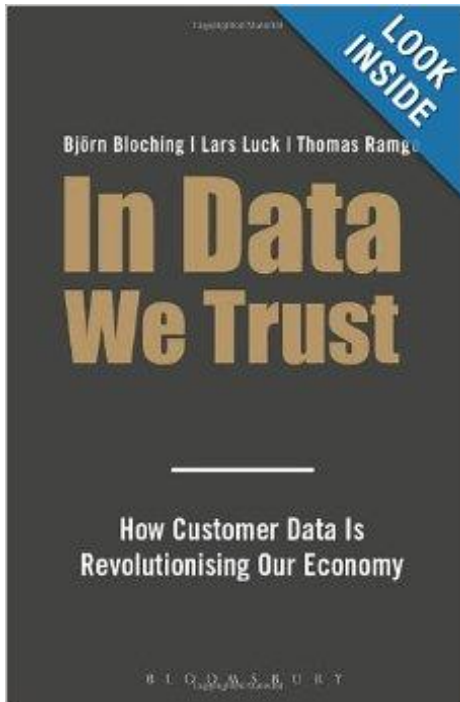
## Very high impact events and facts

- **US Election 2012** – Obama’s campaign and rise of Big Data analytics
  - Micro-targeting and Social Networks analysis
- **Brexit 2016**
  - “Data driven Brexit” – first serious ring for right use of Data Science technologies
- **US Election 2016**
  - Clinton’s campaign – “Data driven” but using only upper layer of Social Network (SN) web
  - Trump’s campaign – Targeting bottom SN web and “forgotten people not to be forgotten”
    - Matt Oczkowski, leader on Trump’s campaign: “If he was going to win this election, it was going to be because of a Brexit style mentality and a different demographic trend than other people were seeing.”
- **France election 2017**
  - Awakening



# Data-Driven Brexit: A Wakeup Call for Analysts

By Barry Devlin, June 28, 2016



Book: In Data We Trust:  
How Customer Data is  
Revolutionising Our  
Economy (Aug 2012)

- A strategy for tomorrow's data world

Data-Driven Brexit: A  
Wakeup Call for Analysts  
By Barry Devlin, June 28,  
2016



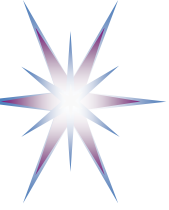
- Article “In Data we trust” by T.Edsall in The New York Times
- Multimillion-dollar contract for data management and collection services awarded May 1, 2013 to Liberty Work (for Republicans) to build advanced list of voters

- There are significant lessons for believers in data-driven business to learn from how data was and wasn't used for decision making before, during, and after the Brexit vote.
- Human attitude -- including emotion, intuition, and social empathy -- and motivation are at the heart of decision making and the action that follows
- Information will only be accepted when it conforms to preconceived notions. Expertise is not sufficient and, *in extremis*, will be dismissed with ridicule.



# US elections 2016 and Data Analytics

- On-going scandal with Cambridge Analytica
- Facebook under fire
- Growing importance of ethical factor
  - Education is essential to tame new element/dimension of our life - Data
- Increasing impact of EU GDPR (General Data Protection Regulation) - in force from 25 May 2018
  - ePrivacy legislation is coming expected 2020
    - Equal services provision with and w/o identification data collection (online services, social network, network games)
    - Strong lobbying against by Big technology companies



# Education and Ethical norms

- Education is essential to tame new element/dimension of our life – Data
- Importance of Ethics is rising due to emerging AI based decision making
- Industry 4.0 and digital transformation

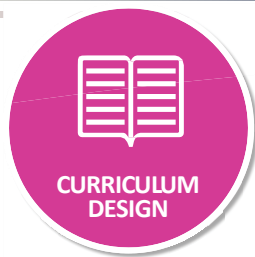


# Industry 4.0 and demand for new skills

WORLD  
ECONOMIC  
FORUM



The Fourth Industrial Revolution, which includes developments in previously disjointed fields such as artificial intelligence and machine-learning, robotics, nanotechnology, 3-D printing, and genetics and biotechnology, will cause widespread disruption not only to business models but also to labour markets over the next five years, with enormous change predicted in the skill sets needed to thrive in the new landscape. This is the finding of a new report, *The Future of Jobs*, published today by the World Economic Forum.

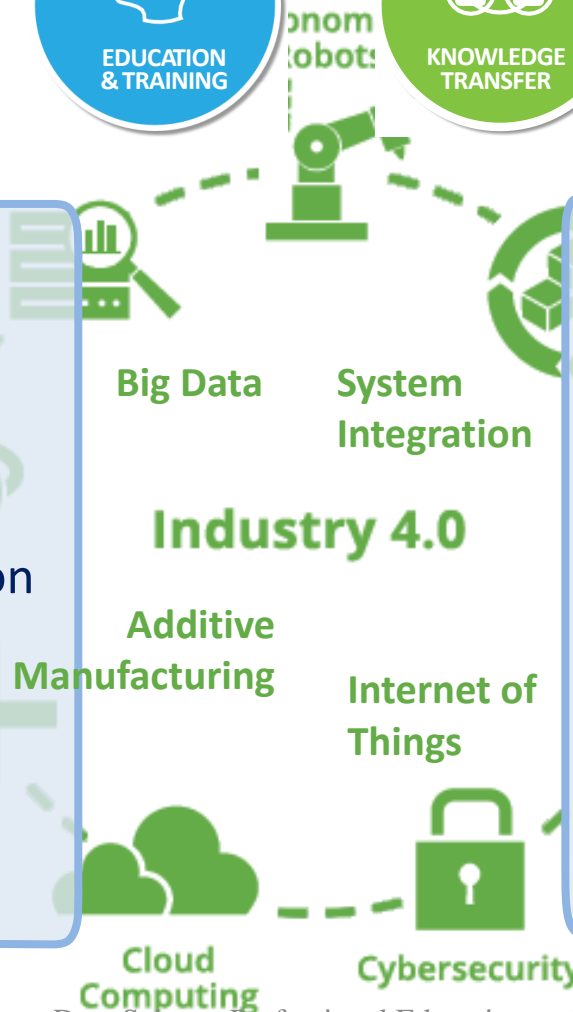


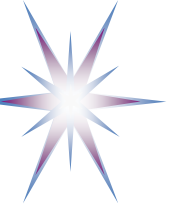
## Digital Transformation

- Digitalisation and IoT
- Intelligent Information
- Data Management
- Digital Assets Manage
- Data Driven Optimisation
- Agile Continuous Improvement
- Customer Experience
- People and skills

## Digital Competences/Skills

- Automation, robotics, electrical vehicles
- Information and data literacy
- Communication and collaboration
- Digital content creation, safety
- Problem solving and critical thinking





# OECD and UN on Digital Economy and Data Literacy

## OECD

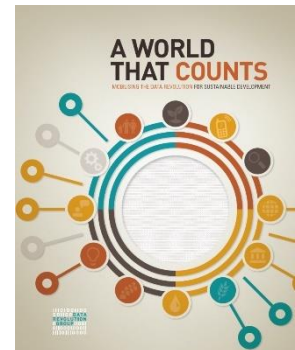
- Demand for new type of *“dynamic self-re-skilling workforce”*
- Continuous learning and professional development to become a shared responsibility of workers and organisations

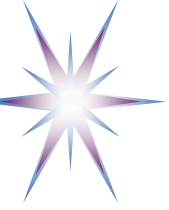
[ref] SKILLS FOR A DIGITAL WORLD, OECD, 25-May-2016

[http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/ICCP/IIS\(2015\)10/FINAL&docLanguage=En](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/ICCP/IIS(2015)10/FINAL&docLanguage=En)

## UN

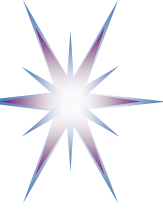
- Data Revolution Report "A WORLD THAT COUNTS" Presented to Secretary-General (2014)  
<http://www.undatarevolution.org/report/>
- Data Literacy is defined as key for digital revolution
- **Data literacy** = critically analyse data collected and data visualised





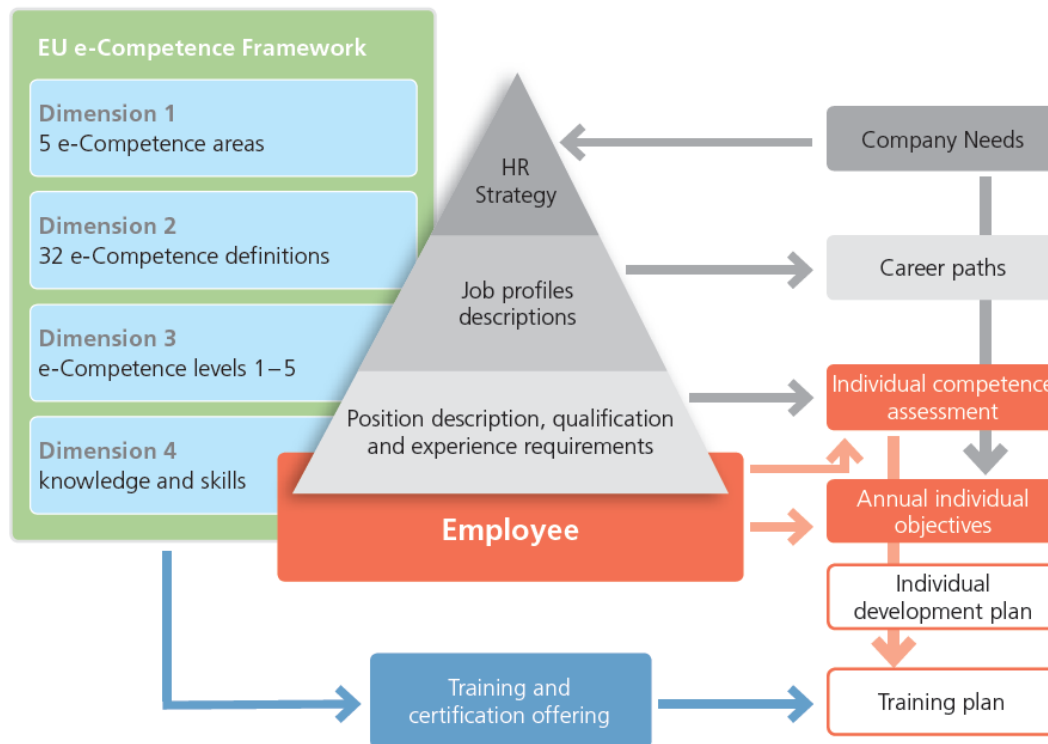
# EDSF: Background standards

- eCFv3.0 – e-Competence Framework
- European ICT Professional Profiles
- European Skills, Competences, Qualifications and Occupations (ESCO) Taxonomy  
<https://ec.europa.eu/esco/portal/#modal-one>
- IEEE/ACM Classification Computer Science (CCS2012)
- ACM Information Technology Competencies Model



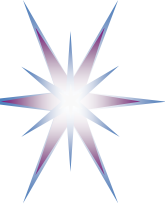
# e-CFv3.0 structure and 4-dimensional model

- European e-Competence Framework for IT (e-CFv3.0) dimension
  - Dimension1: 5 competence areas: Plan, Build, Run, Enable, Manage
  - Dimension2: 32 e-competence definition
  - Dimension3: 5 proficiency levels
  - Dimension4: Knowledge and skills



- Multiple use of e-CFv3.0 within ICT organisations
- Provides basis for individual career path, competence assessment, training and certification

- EDISON CF-DS will be used for defining DS-BoK and MC-DS, linking organizational functions and required knowledge
- Mapping CF-DS and DS-BoK to academic disciplines
- Provide basis for individual (self) training and certification



# e-CFv3.0 Internal Structure: Refactoring for CF-DS

## European e-Competence Framework 3.0 overview

Dimension 1 5 e-CF areas (A – E)	Dimension 2 40 e-Competences identified	Dimension 3 e-Competence proficiency levels e-1 to e-5, related to EQF levels 3–8				
		e-1	e-2	e-3	e-4	e-5
A. PLAN	A.1. IS and Business Strategy Alignment					
	A.2. Service Level Management					
	A.3. Business Plan Development					
	A.4. Product/Service Planning					
	A.5. Architecture Design					
	A.6. Application Design					
	A.7. Technology Trend Monitoring					
	A.8. Sustainable Development					
	A.9. Innovating					
B. BUILD	B.1. Application Development					
	B.2. Component Integration					
	B.3. Testing					
	B.4. Solution Deployment					
	B.5. Documentation Production					
	B.6. Systems Engineering					
C. RUN	C.1. User Support					
	C.2. Change Support					
	C.3. Service Delivery					
	C.4. Problem Management					
D. ENABLE	D.1. Information Security Strategy Development					
	D.2. ICT Quality Strategy Development					
	D.3. Education and Training Provision					
	D.4. Purchasing					
	D.5. Sales Proposal Development					
	D.6. Channel Management					
	D.7. Sales Management					
	D.8. Contract Management					
	D.9. Personnel Development					
	D.10. Information and Knowledge Management					
	D.11. Needs Identification					
	D.12. Digital Marketing					
E. MANAGE	E.1. Forecast Development					
	E.2. Project and Portfolio Management					

- **4 Dimensions**

- Competence Areas
- Competences
- Proficiency levels
- Skills and Knowledge

- **5 Competence Areas** defined by ICT Business Process stages

- Plan
- Build
- Deploy
- Run
- Manage

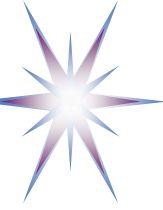
-> Refactor to Scientific Research (or Scientific Data) Lifecycle

- See example of RI manager at IG-ETRD wiki and meeting

- Each competence has **5 proficiency levels**

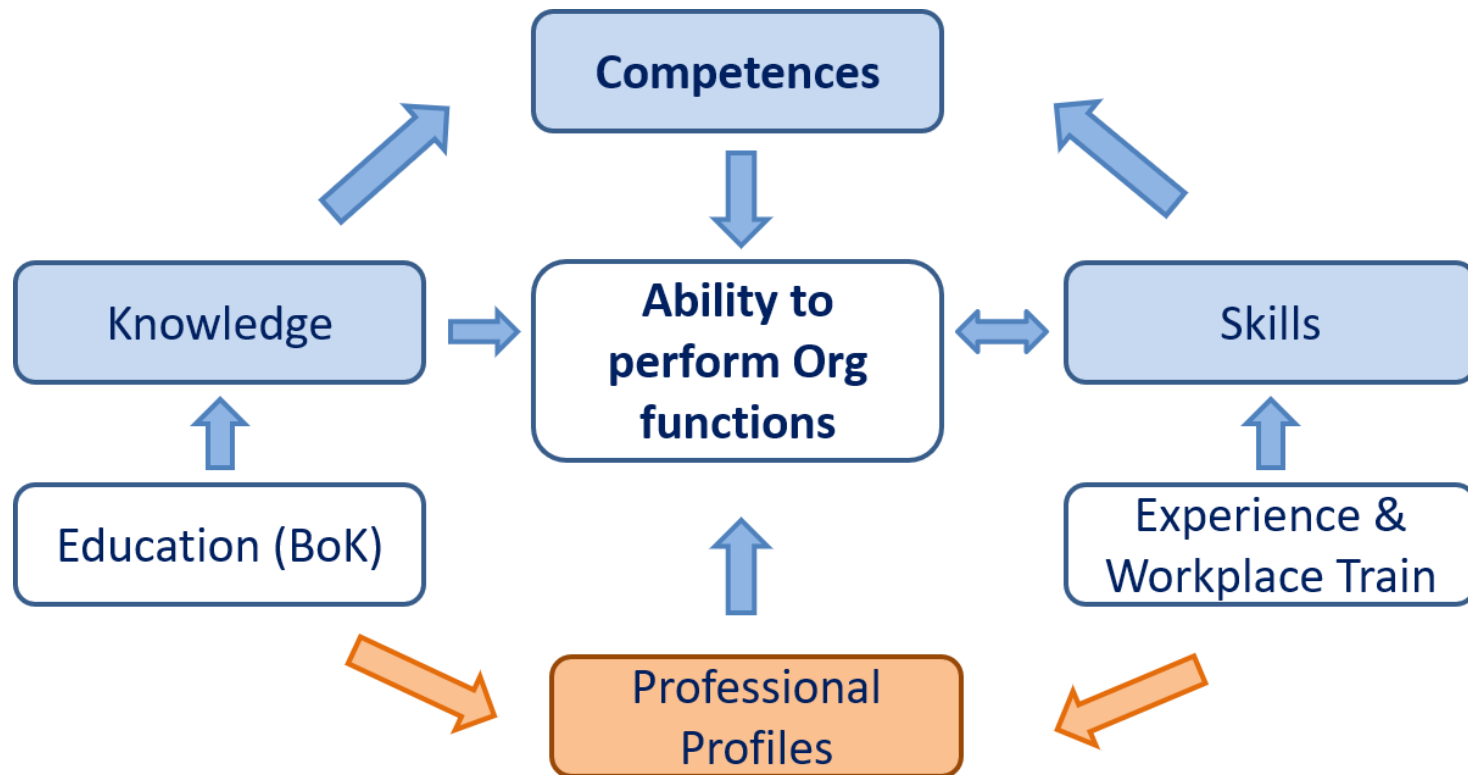
- Ranging from technical to engineering to management to strategist/expert level

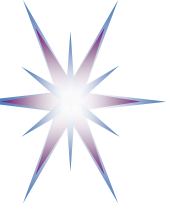
- Knowledge and skills property are defined for/by each competence and proficiency level (not unique)



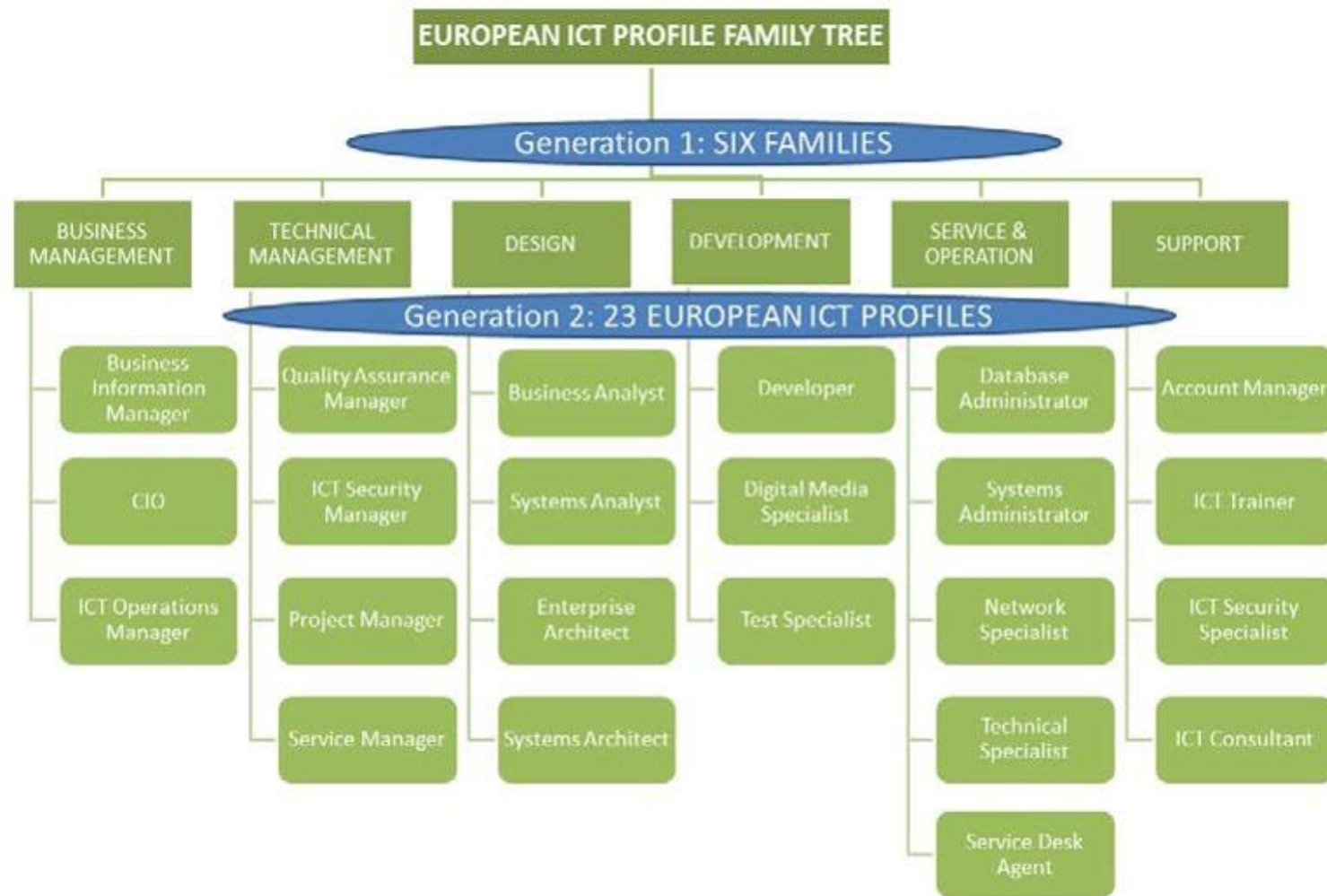
# Competences Map to Knowledge and Skills

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



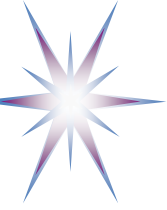


# CWA 16458 (2012): European ICT Professional Profiles

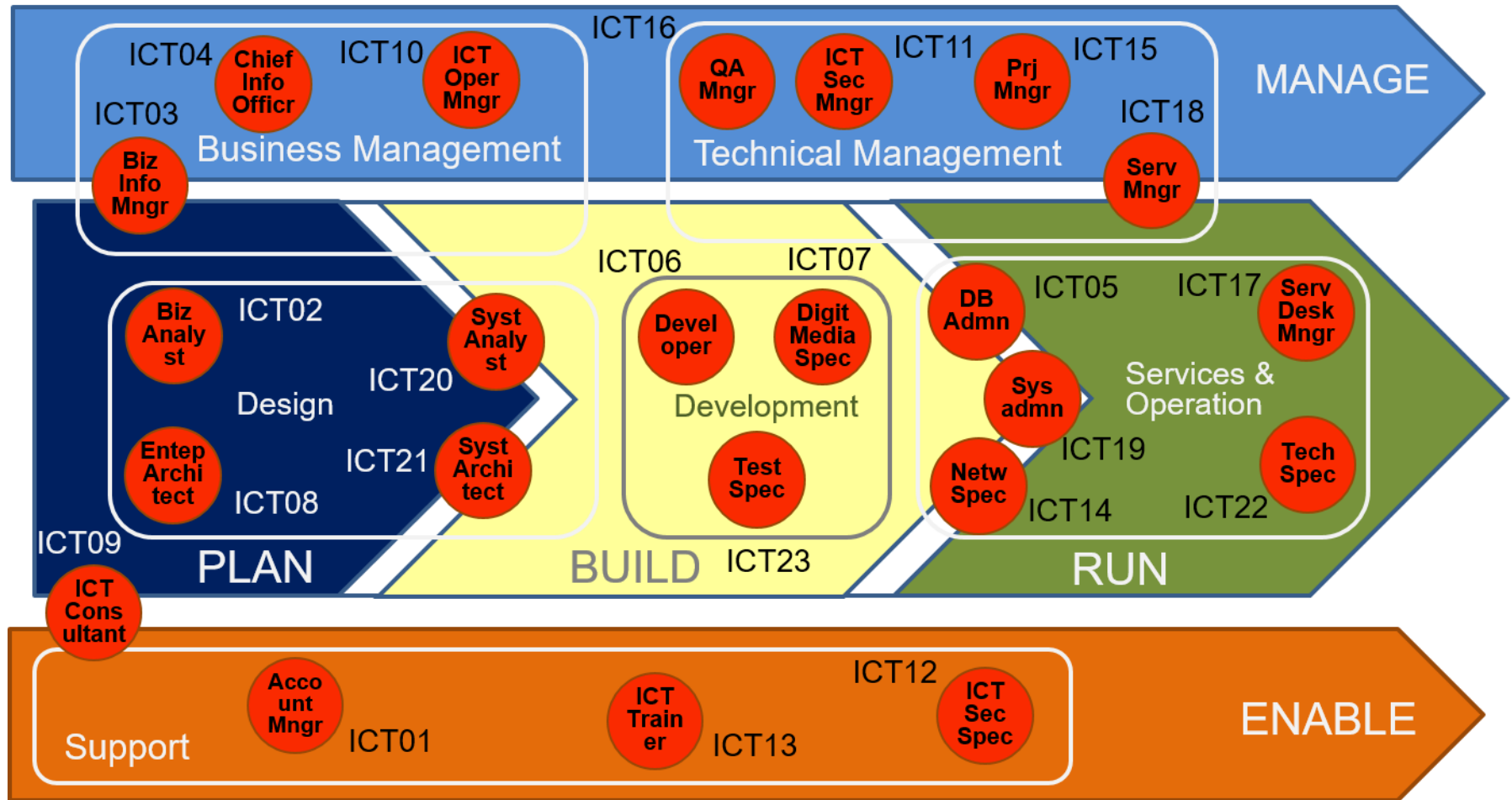


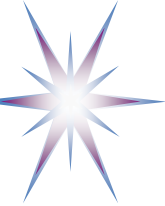
- The CWA defines 23 main ICT profiles the most widely used by organisations





# CWA Professional Profiles and Organisational Workflow

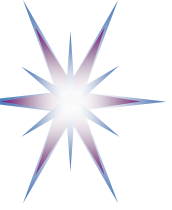




# ESCO Taxonomy (Mandatory since 2017)

<https://ec.europa.eu/esco/portal/#modal-one>

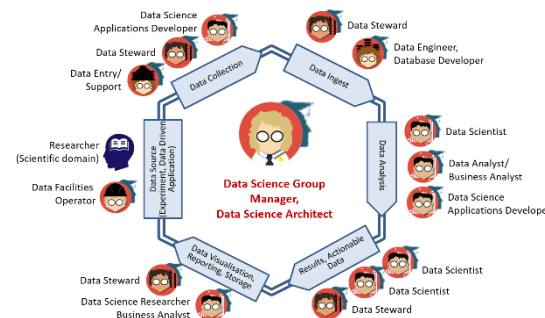
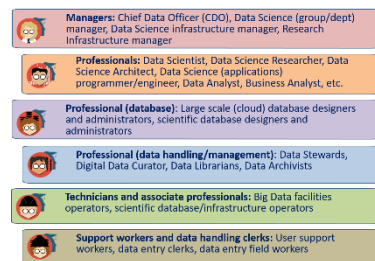
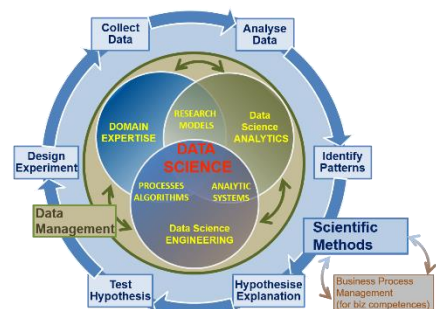
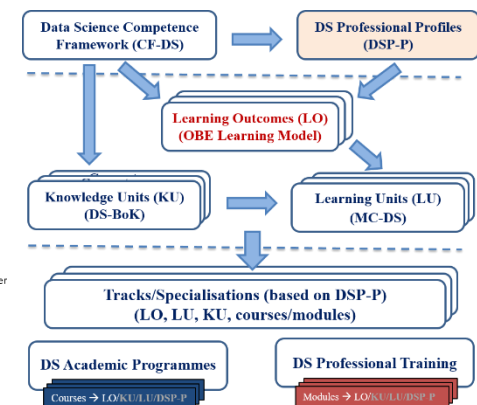
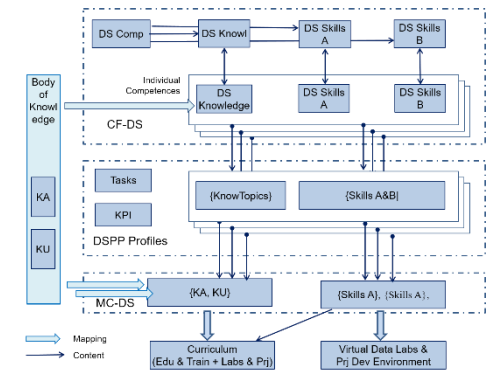
Professionals				
	Science and engineering professionals	<b>Data Science Professionals</b>	Data Science professionals not elsewhere classified	DSP04 Data Scientist
				DSP05 Data Science Researcher
				DSP08 (Big) Data Analyst
Technicians and associate professionals				
	Science and engineering associate professionals	<b>Data Science Technology Professionals</b>	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	<b>Data Science/Big Data Infrastructure Managers</b>		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			
	<b>Data handling support workers (alternative)</b>	<b>Data and information entry and access</b>	Digital Archivists and Librarians	DSP20 Data entry/access desk/terminal workers
				DSP21 Data entry field workers
				DSP22 User support data services



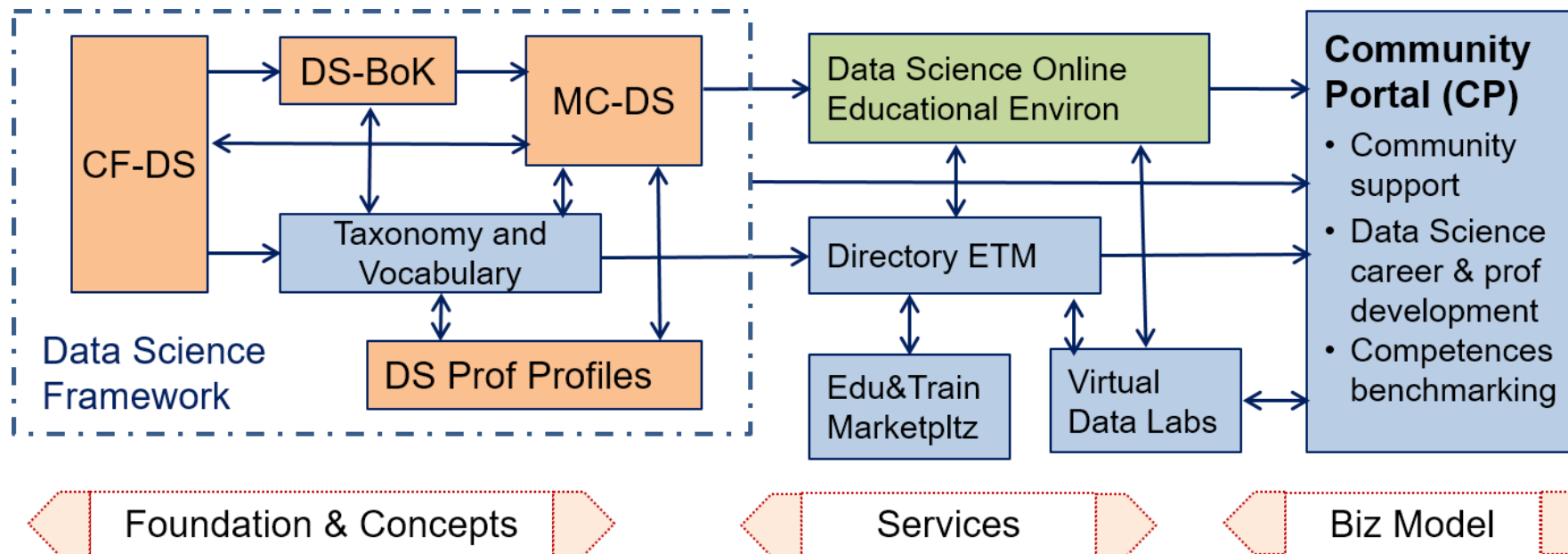
# ICT and Data related Skills shortage - Impact

- Problems with hiring (skilled) ICT related staff
  - At least one year for training and acquiring experience
  - As soon as new employees are confident with their skills, they leave for big(ger) companies or industry
- Open Data Science/Stewards positions stay unfilled longer
  - In research institutions for months and years
  - In industry for months
- Companies/organisations want experienced Data Science workers
  - There is no time to acquire necessary experience
- Millennials factor
  - Do we understand difference of the millennials workforce?
- Challenges: How to obtain, train in shorter period and sustain new digital (ICT and Data related) skills in organisations

- 
- The diagram illustrates the Data Science Framework, organized into two main sections: **Foundation & Concepts** and **Services**.
- Foundation & Concepts (enclosed in a dashed blue box):**
- CF-DS** (Data Science Framework Core) is the central component, connected to **DS-BoK** and **MC-DS**.
  - DS-BoK** (Data Science Body of Knowledge) and **MC-DS** (Master Curriculum for Data Science) are interconnected with **CF-DS** and **Taxonomy and Vocabulary**.
  - Taxonomy and Vocabulary** is connected to **DS-BoK**, **MC-DS**, and **DS Prof Profiles**.
  - DS Prof Profiles** (Data Science Professional Profiles) is connected to **MC-DS** and **Capacity Building and Career Development**.
- Services:**
- Data Science Community Portal** is connected to **MC-DS**.
  - Online Educational Environment** is connected to **MC-DS**.
  - Capacity Building and Career Development** is connected to **DS Prof Profiles**.



# EDISON Data Science Framework (EDSF)

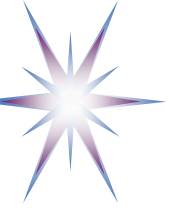


## EDISON Framework components

- 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

- EDSF 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.



# Links to EDISON Resources

- EDISON Data Science Framework Release 3 (EDSF)  
<https://github.com/EDISONcommunity/EDSF>

## Component EDSF documents

CF-DS – Data Science Competence Framework

[https://github.com/EDISONcommunity/EDSF/blob/master/EDISON\\_CF-DS-release3-v09.pdf](https://github.com/EDISONcommunity/EDSF/blob/master/EDISON_CF-DS-release3-v09.pdf)

DS-BoK – Data Science Body of Knowledge

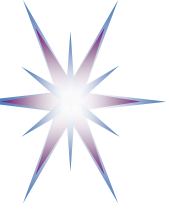
[https://github.com/EDISONcommunity/EDSF/blob/master/EDISON\\_DS-BoK-release3-v04.pdf](https://github.com/EDISONcommunity/EDSF/blob/master/EDISON_DS-BoK-release3-v04.pdf)

MC-DS – Data Science Model Curriculum

[https://github.com/EDISONcommunity/EDSF/blob/master/EDISON\\_MC-DS-release3-v04.pdf](https://github.com/EDISONcommunity/EDSF/blob/master/EDISON_MC-DS-release3-v04.pdf)

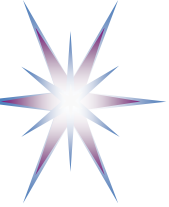
DSPP – Data Science Professional profiles

[https://github.com/EDISONcommunity/EDSF/blob/master/EDISON\\_DSPP-release3-v05.pdf](https://github.com/EDISONcommunity/EDSF/blob/master/EDISON_DSPP-release3-v05.pdf)



# 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](mailto:edison-net@list.uva.nl)
- EDISON project website (legacy) – in move to <http://edison-project.net/>
  - Redirected to <https://github.com/EDISONcommunity/EDSF/wiki/EDSFhome>



# Other related links

- 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>



This work is licensed under the Creative Commons Attribution 4.0 International License.  
To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>