

# The Data Talents Gap Panel

Big Data and Analytics Innovation Summit

2-3 Nov 2017, London



## MODERN DATA SCIENTIST

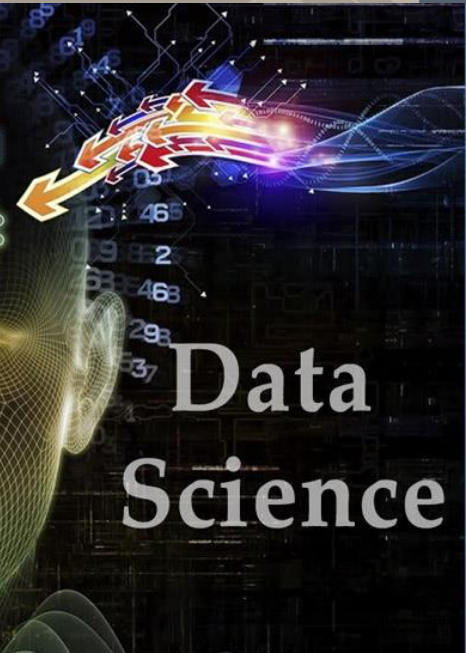
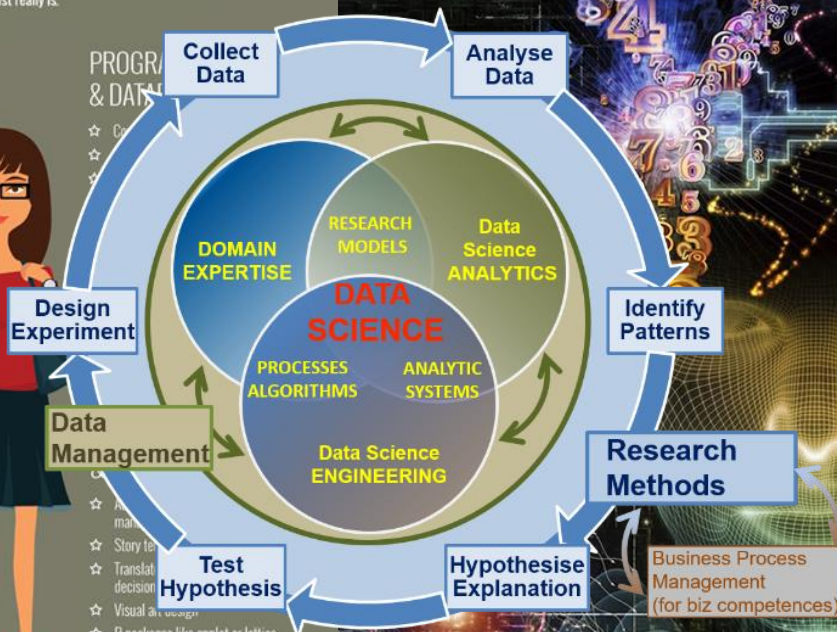
Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

### MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

### DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative,



# Data Science



# The Data Talents Gap Panel

- Introduction to the Panel
  - Panel Moderator - Yuri Demchenko (EDISON project coordinator, University of Amsterdam)
  - Data Science and Analytics competences and skills demand and market
- Statements by Panel members
  - Christoph Best (Google)
  - Joshua Ryan-Saha (The Data Lab)
  - Robert Monné (The Analytics Academy, ORTEC Data Science)
  - Sue Daley (techUK, University of Birmingham) – Apologies
- Questions and answers: Interactive discussion



# Intro to Panel on Data Talents Gap

## What we are going to discuss

- What does the right talent look like?
- What does the right management look like?
- Is the economy producing the requisite number & quality of people?
- So what are the strategies for finding the right people and the incentives needed to keep them engaged over time?
- How can we retain the talent we already have?
- How can we upskill those already there and lastly, ensure a pipeline of candidates so the area can continue to grow and develop?



# Introduction to the Talents Gap

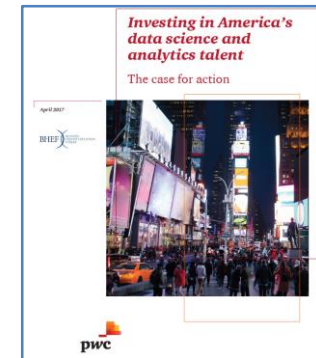
- Recent reports, studies and facts
- Existing Initiatives, projects, frameworks
- EDISON Data Science Framework (EDSF)
- Data Science competences and skills
- Essential Data Scientist professional skills: Thinking and Acting like Data Scientist





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

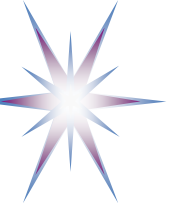
- Final Report on European Data Market Study by IDC (Feb 2017)
  - The EU data market in 2016 estimated EUR 60 Bln (growth 9.5% from EUR 54.3 Bln in 2015)
    - Estimated EUR 106 Bln in 2020
  - Number of data workers 6.1 mln (2016) - increase 2.6% from 2015
    - Estimated EUR 10.4 million in 2020
  - 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



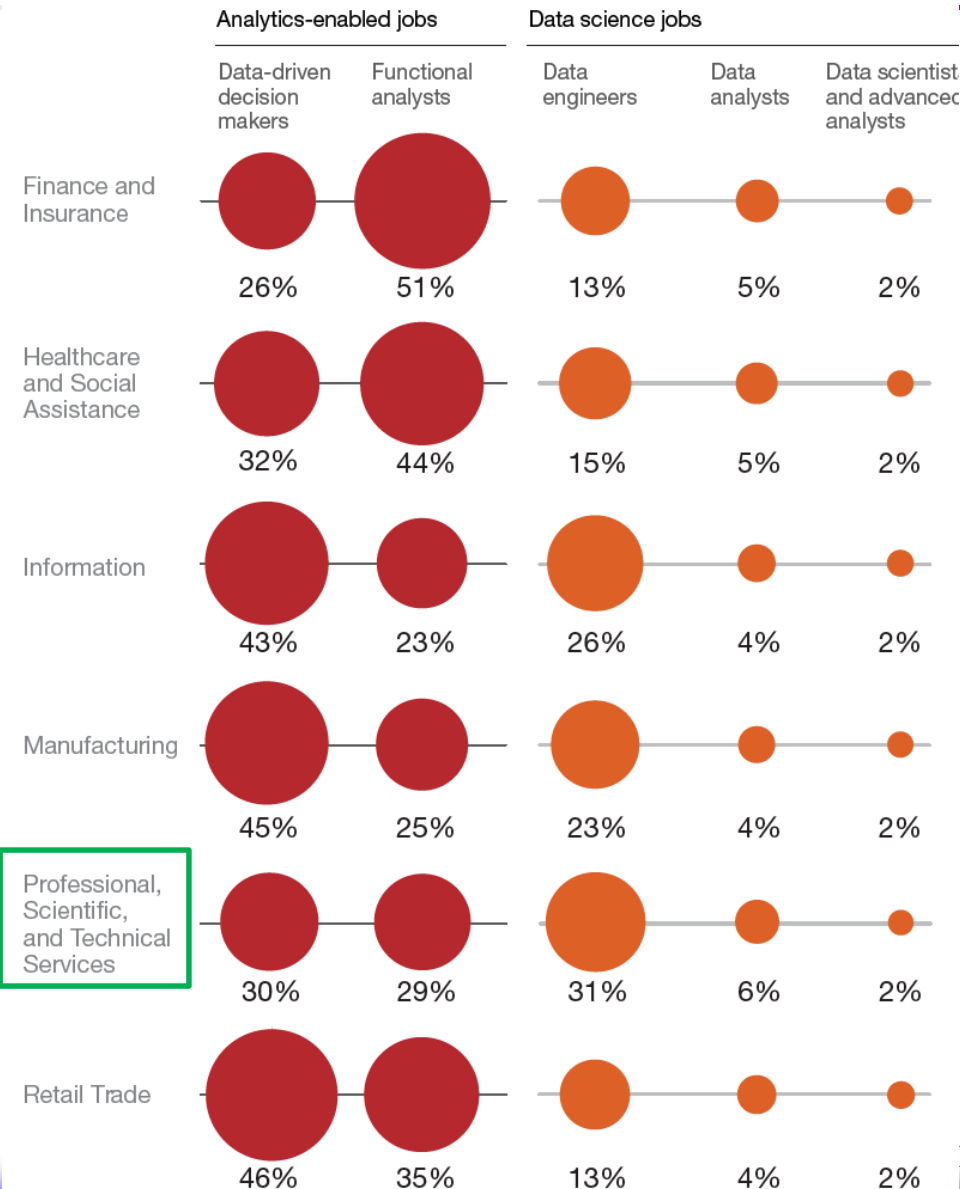
 Citing EDISON and EDSF



 Influenced by EDISON

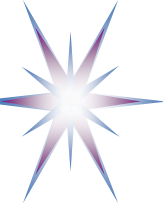


# PwC&BHEF: Demand for DSA enabled jobs



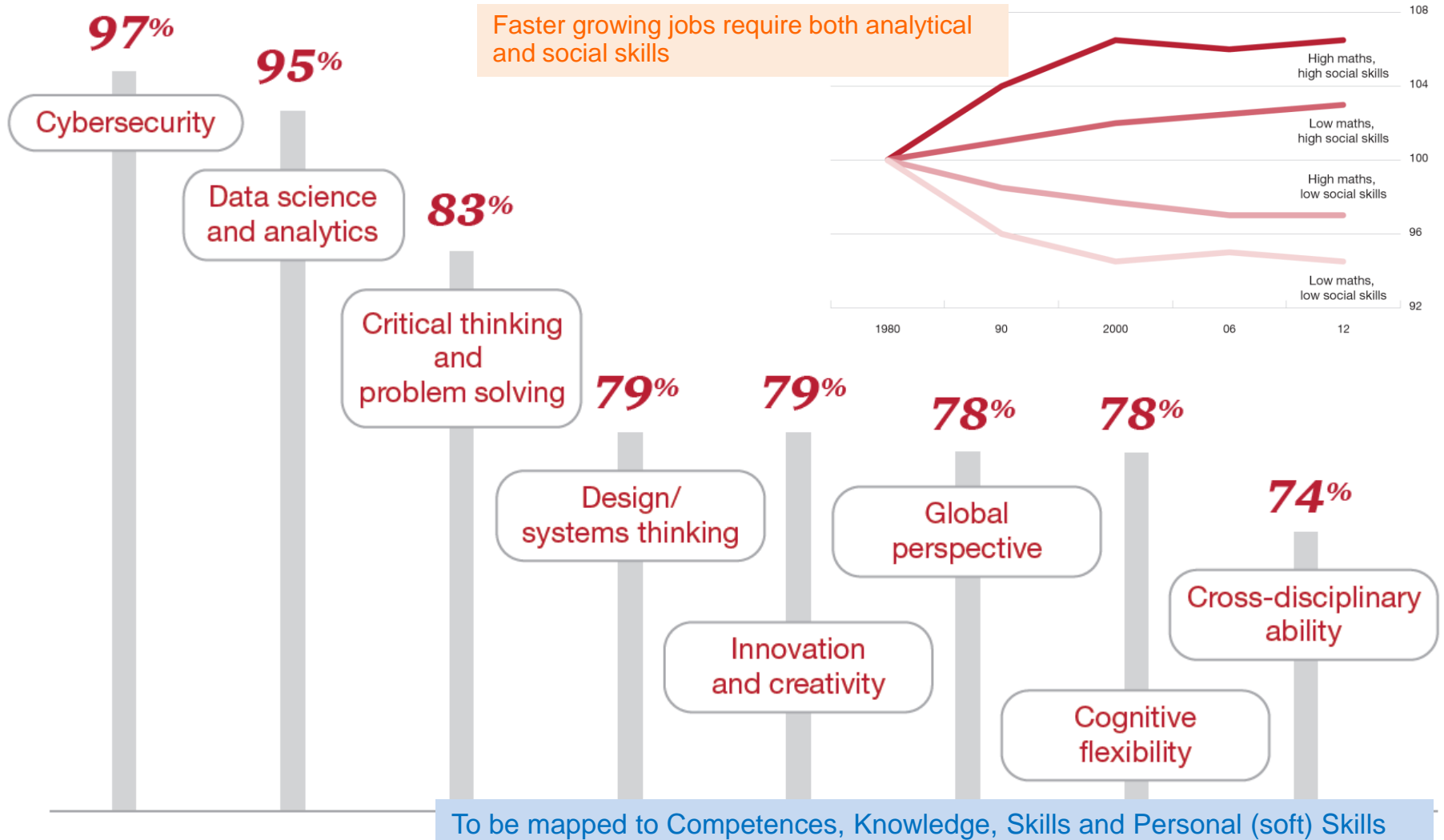
## Demand for business people with analytics skills, not just data scientists

- 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: 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



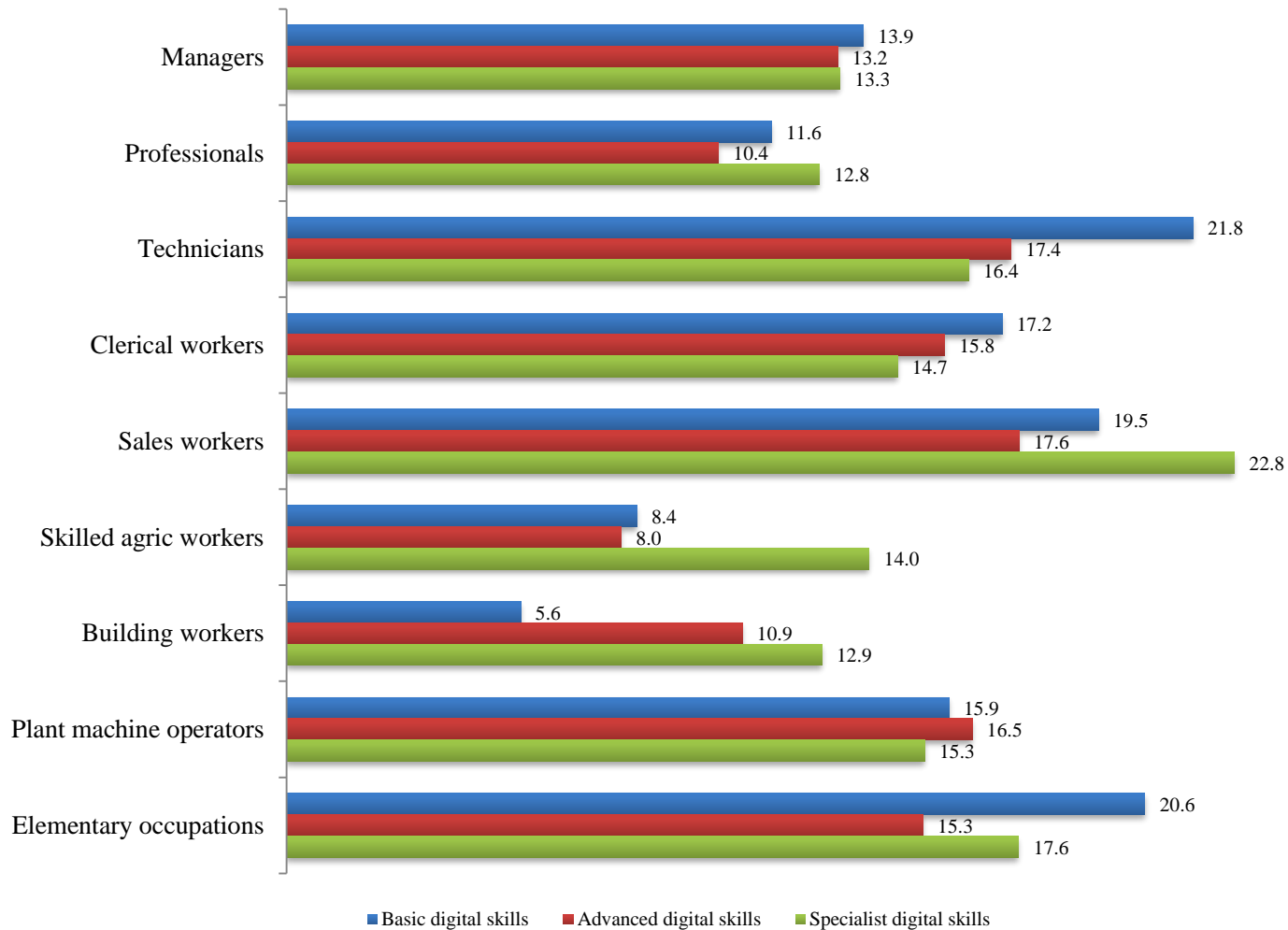
Source: Business Roundtable (2017).

BDAI Summit 2017

The Data Talents Gap



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



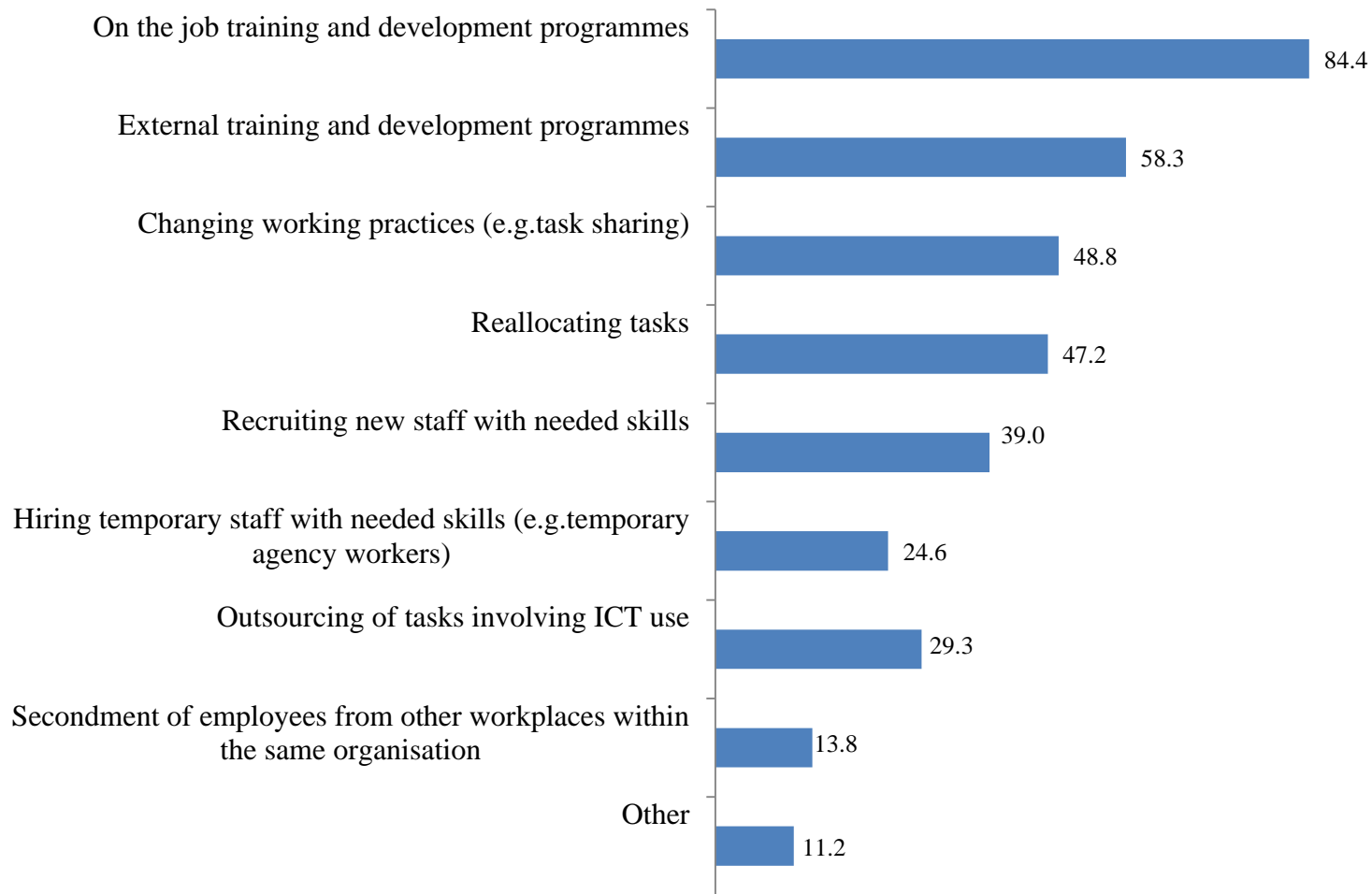
ICT for work: Digital skills in the workplace, Digital Single Market, Reports and studies, May 2017

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





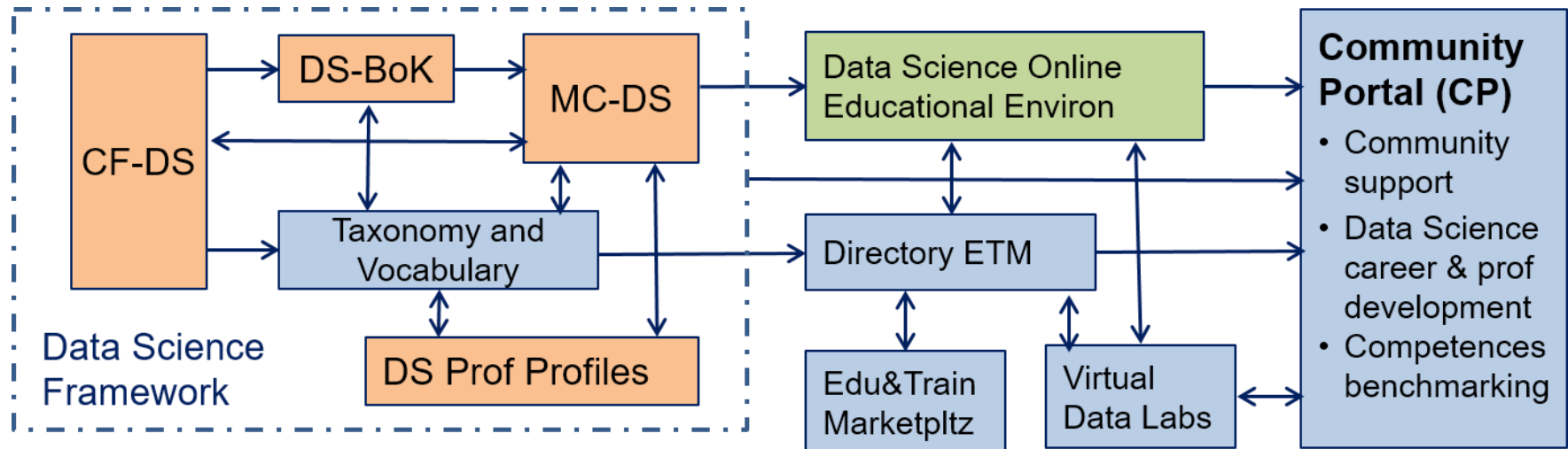
# Workplaces reporting having taken action to tackle digital skill gaps by type of action undertaken, EU28 (% of workplaces with digital skill gaps which undertook actions)



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>

# EDISON Data Science Framework (EDSF)



Foundation & Concepts

Services

Biz Model

## EDISON Framework components

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

## Methodology

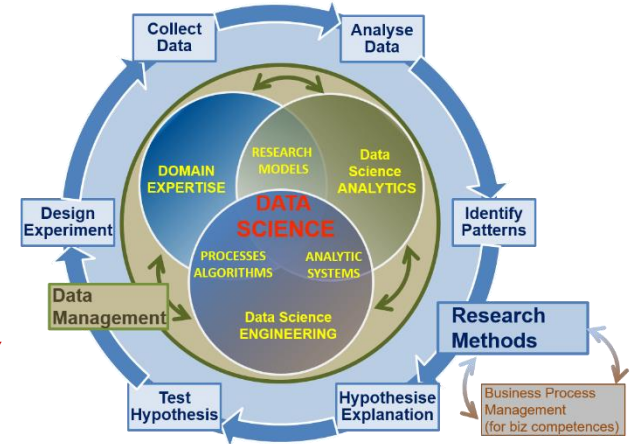
- ESDF development based on job market study, existing practices in academic, research and industry.
- Review and feedback from the ELG, expert community, domain experts.
- Input from the champion universities and community of practice.



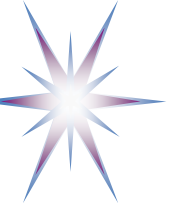
# Data Scientist definition

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

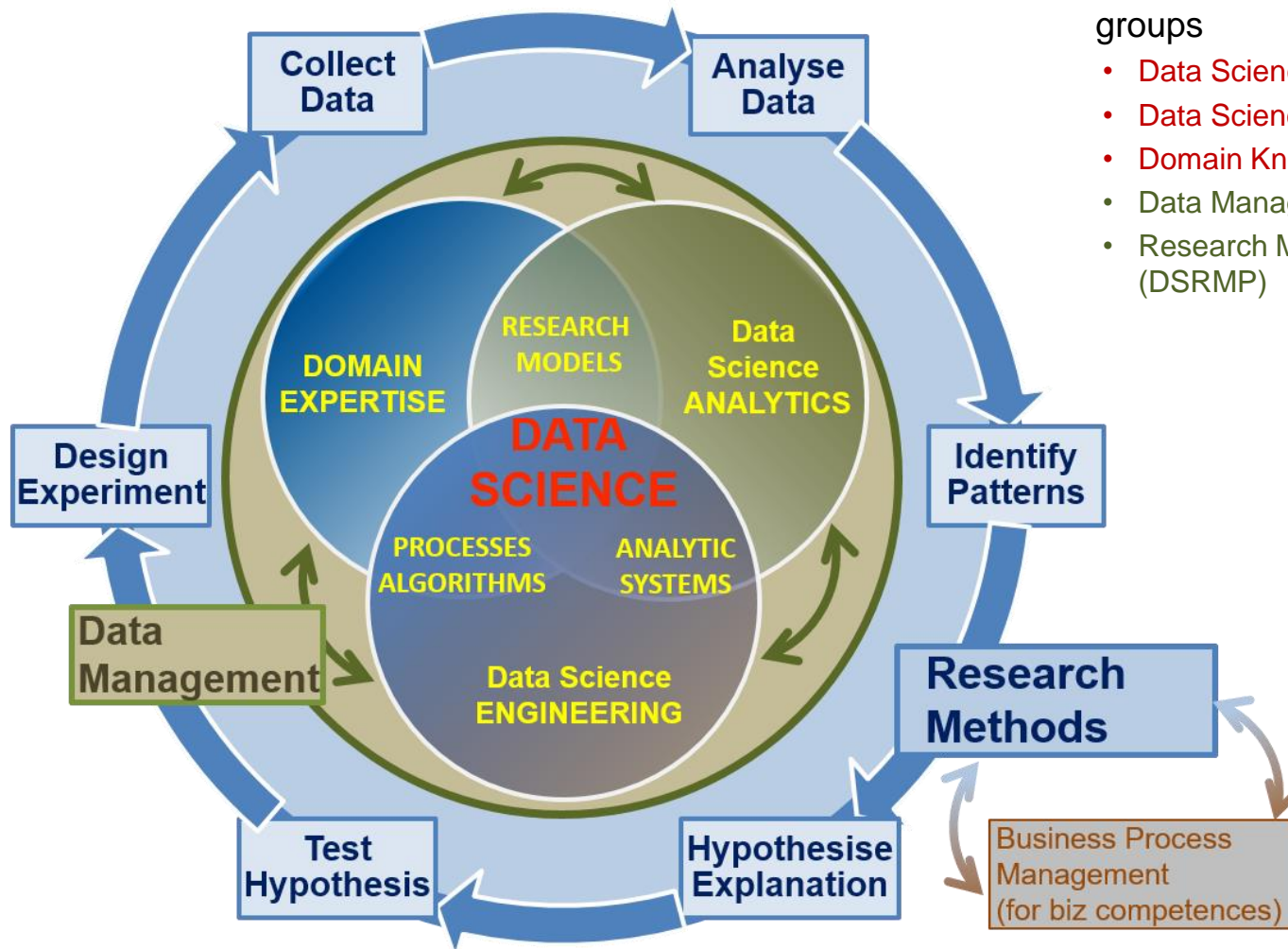
- A **Data Scientist** is a practitioner who has sufficient knowledge in the overlapping regimes of expertise in **business needs, domain knowledge, analytical skills, and programming and systems engineering expertise** to manage the end-to-end scientific method process through each stage in the **big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



- Core Data Science competences and skills groups
  - **DSDA - Data Science Analytics** (including Statistical Analysis, Machine Learning, Business Analytics)
  - **DSENG - Data Science Engineering** (including Software and Applications Engineering, Data Warehousing, Big Data Infrastructure and Tools)
  - **DSDM - Data Management, Data Governance, Stewardship, Curation, Preservation**
  - **DSRMP - Research Methods and Project Management**
  - **DSDK - Domain Knowledge and Expertise** (Subject/Scientific domain related)
- 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 (DSDA)
- Data Science Engineering (DSENG)
- Domain Knowledge and Expertise (DSDK)
- Data Management DSDM
- Research Methods Project Management (DSRMP)

## Research Methods

- Hypothesis Formulation
- Design Experiment
- Collect Data
- Analyse Data
- Identify Patterns
- Hypothesis Evaluation

## Business Process Management

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



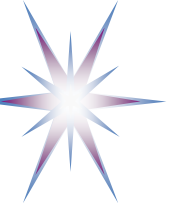
# Identified Data Science *Skills/Experience* Groups

## Skills Type A – Based on knowledge acquired

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

## Skills Type B – Base on practical or workplace experience

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



# Data Science Professional Skills: Thinking and Acting like Data Scientist

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



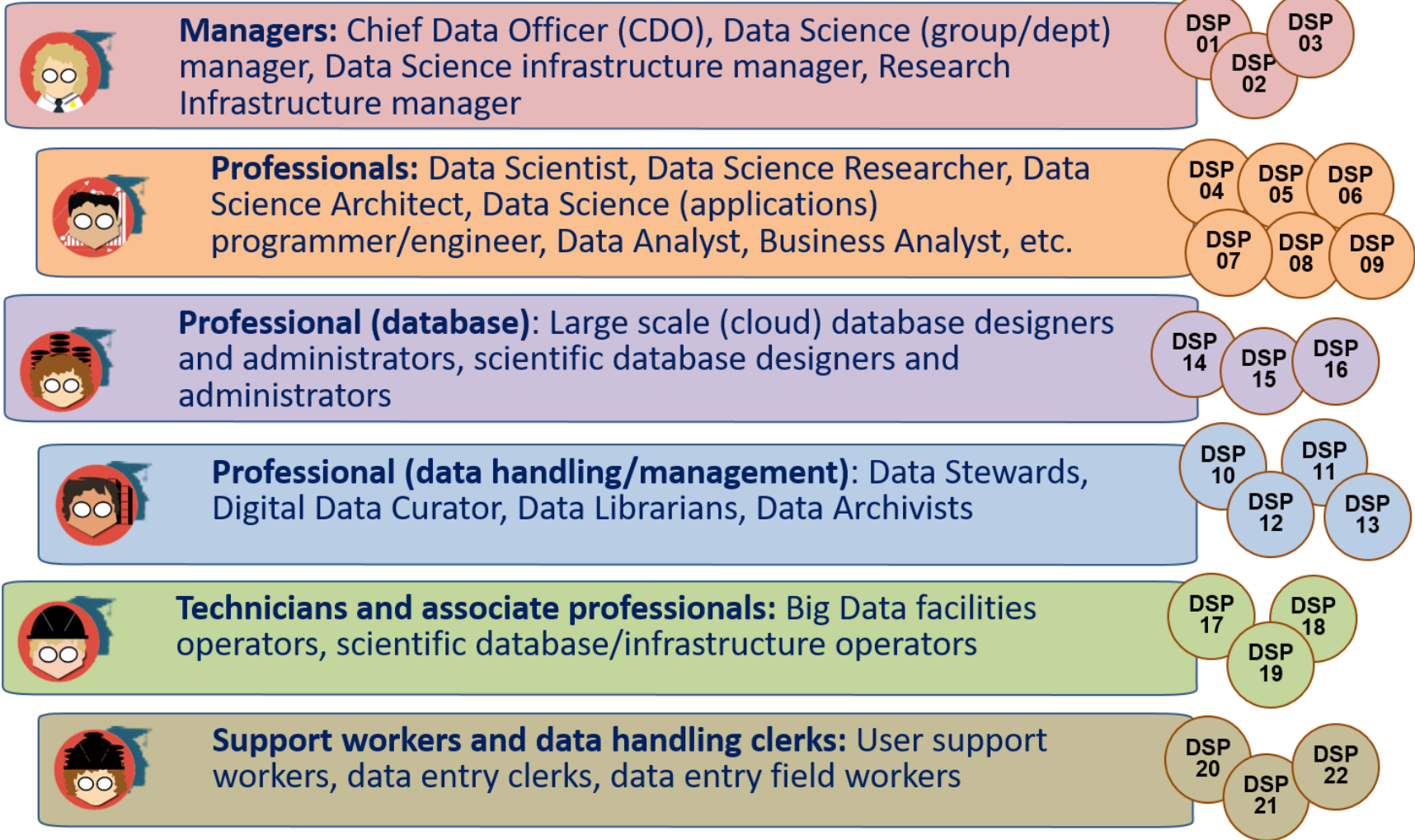


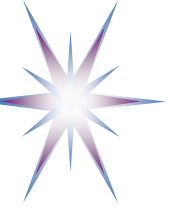
# 21st Century Skills (DARE & BHEF & EDISON)

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

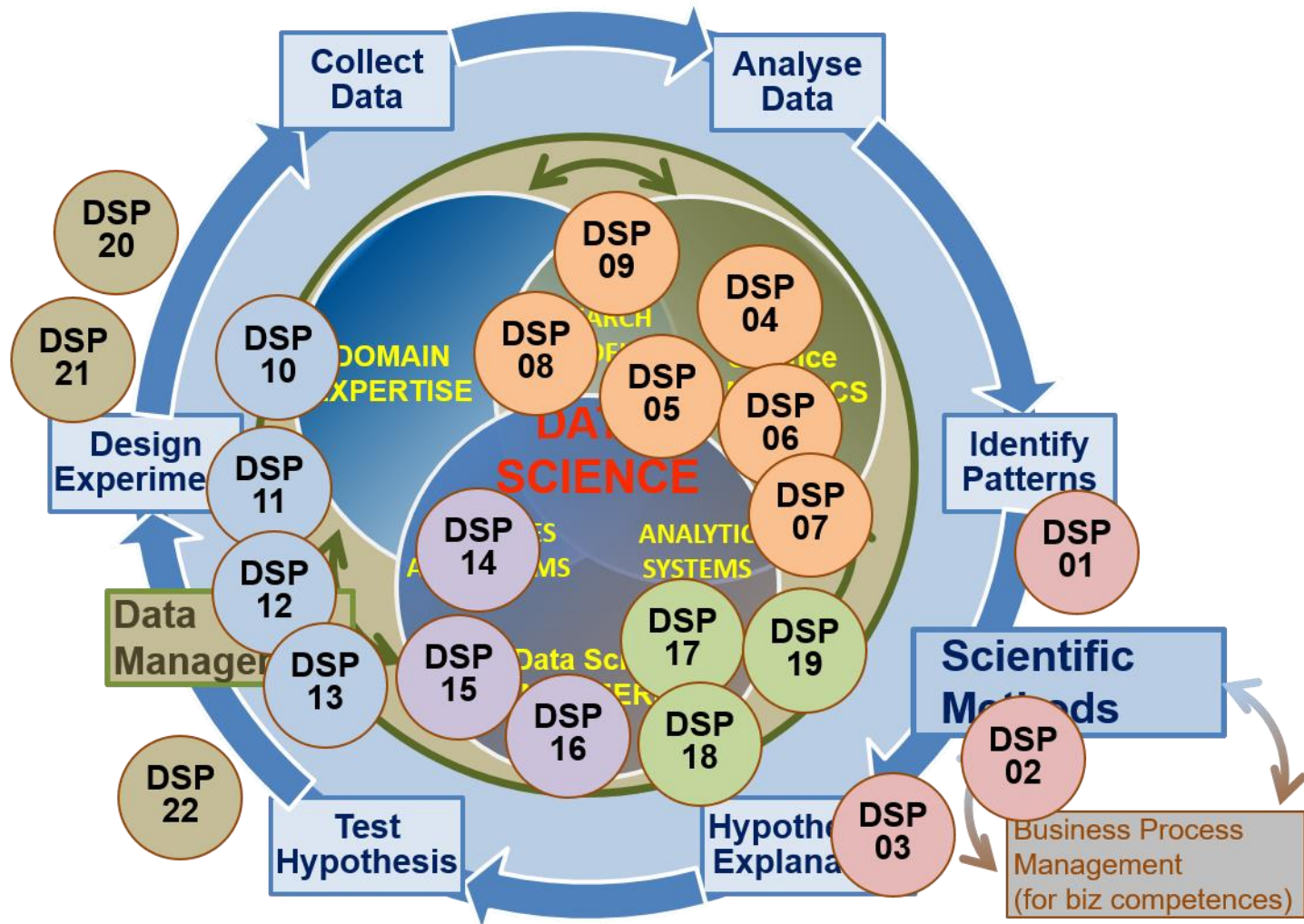


# Data Science Professions Family





# CF-DS and Data Science Professional Profiles

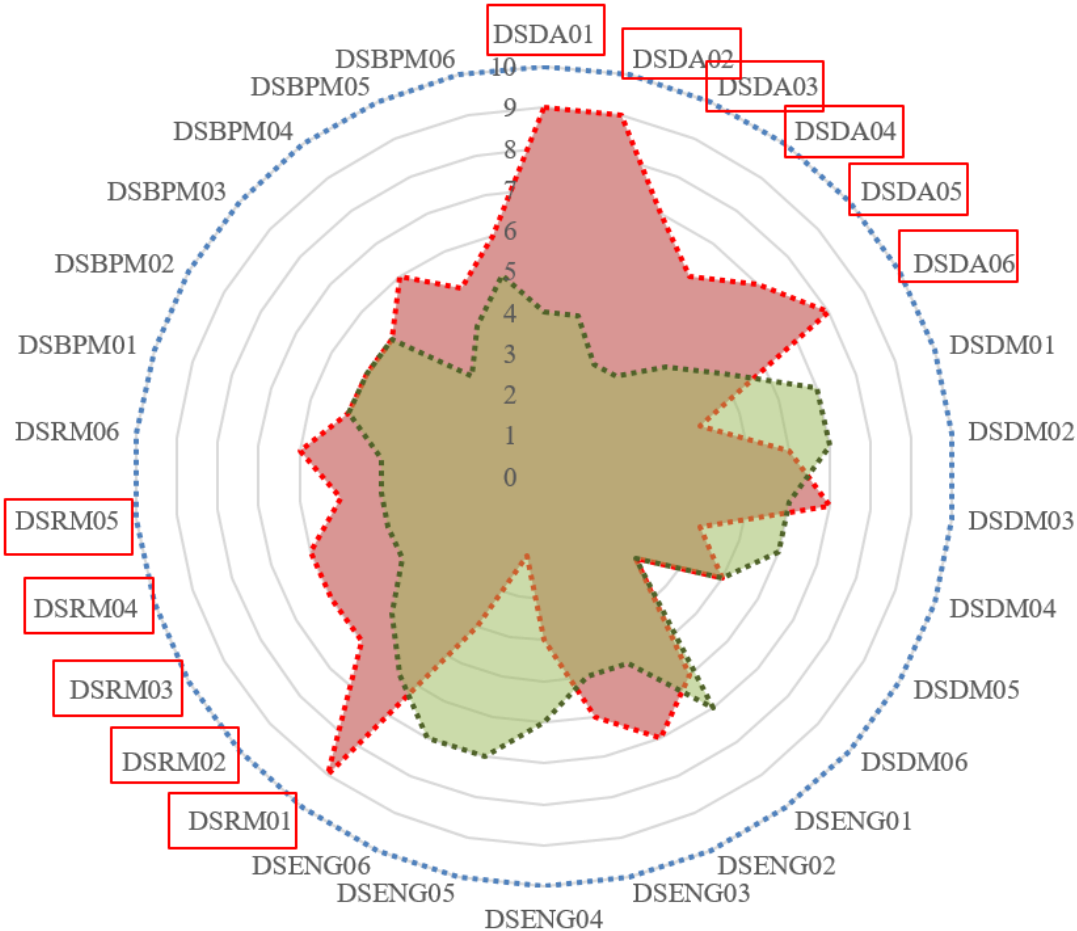




# Individual Competences Benchmarking Based on EDSF Competence Framework

## MATCHING – COMPETENCE PROFILES

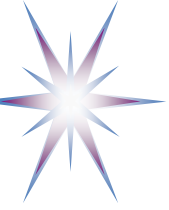
■ DSP04 - Data Scientist   ■ Candidate - Data Scientist



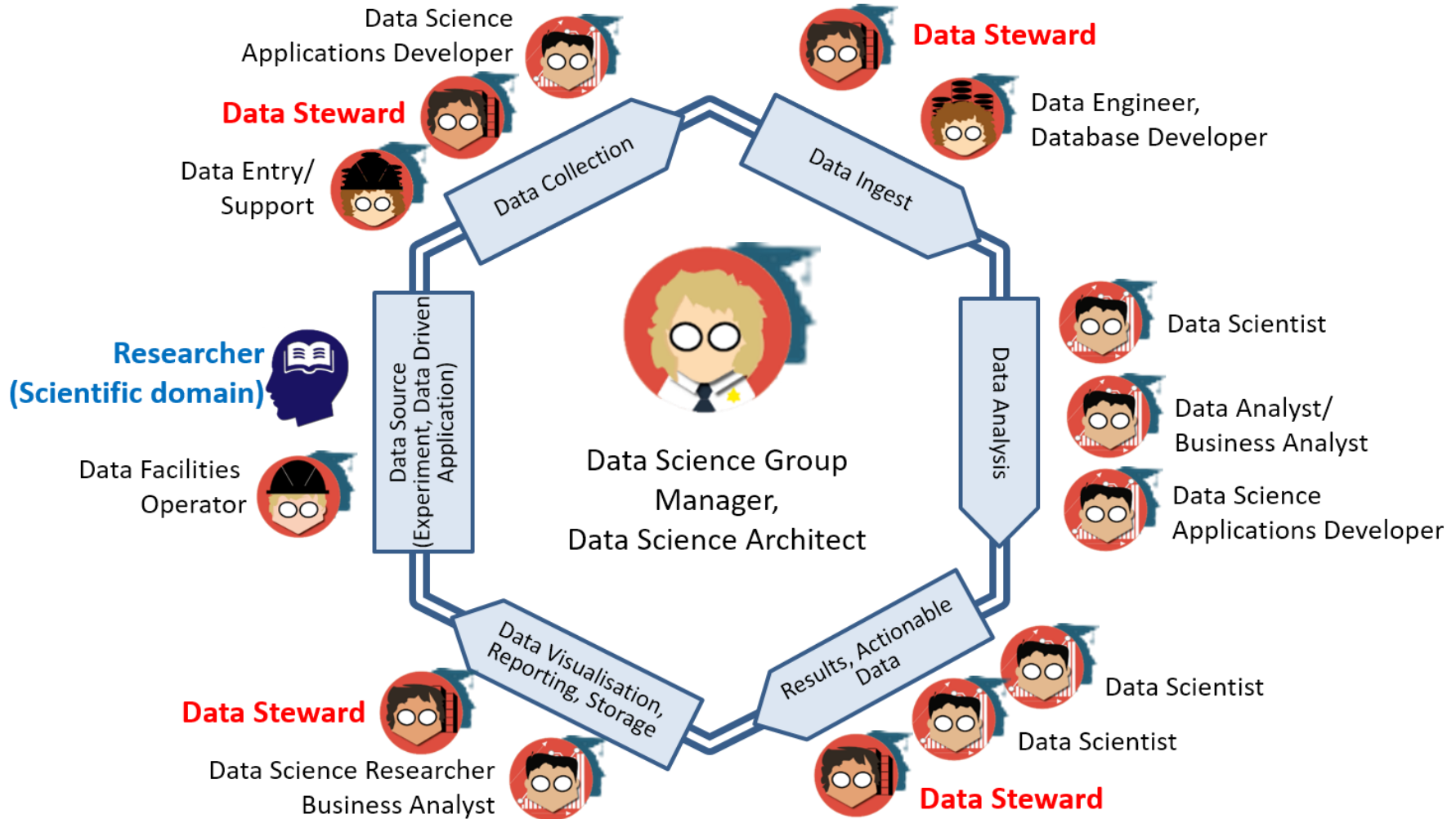
## Individual Education/Training Path based on Competence benchmarking

- Red polygon indicates the chosen professional profile: Data Scientist (general)
- Green polygon indicates the candidate or practitioner competences/skills profile
- Insufficient competences (gaps) are highlighted in *red*
  - DSDA01 – DSDA06 Data Science Analytics
  - DSRM01 – DSRM05 Data Science Research Methods
- Can be use for team skills match marking and organisational skills management

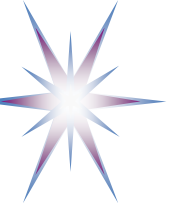
[ref] For DSP Profiles definition and for enumerated competences refer to EDSF documents CF-DS and DSP Profiles.



# Building a Data Science Team



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



# Statements by Panel members

---

- Christoph Best (Google)
- Joshua Ryan-Saha (The Data Lab)
- Robert Monné (The Analytics Academy, ORTEC Data Science)





# Questions to discuss and to pick up by the Panel or audience

- *How to tackle digital skills gaps, how to ensure effective collaboration across stakeholders, and within organisation/company*
- Who are the key players and how should these cooperate? Look for win-win scenarios.
- What do we have to do together to address digital skills gaps?
- *Do you see a role for central/EU level policy or standardisation to facilitate the digital and data skills development?*
- Data Analytics vs Data Management and Governance in companies – are they connected?
- Training: in-house v. consultants v. certified outsiders?
- Data Science, Data Skills – What will Artificial Intelligence advent will change?
- How to secure continued engagement of employer sectors with the developing universities network?



# Questions and discussion

## Links to 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>



# Other related links

- Amsterdam School of Data Science
  - <https://www.schoolofdatascience.amsterdam/>
  - <https://www.schoolofdatascience.amsterdam/education/>
- 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>