

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

## Contributing to EDSF: Workplace and Soft Skills



**EDISON**  
building the data  
science profession

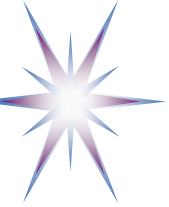
EDISON – Education for **D**ata Intensive  
**S**cience to **O**pen **N**ew science frontiers

H2020 INFRASUPP Grant 675419 (2015-2017)

Yuri Demchenko, EDISON Project  
University of Amsterdam

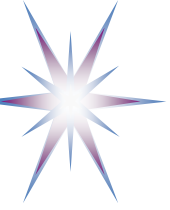
31 May 2018, Amsterdam





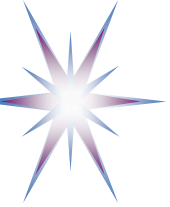
# Discussion, definition, rating: Soft skills and Workplace skills

- Data Science professional skills: Thinking and Acting like Data Scientist
- 21st Century Skills: Personal, inter-personal communication, team work, professional network
- Data Scientist and Subject Domain Specialist



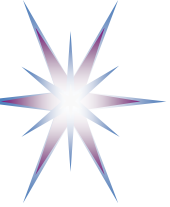
# Data Science Professional Skills: Thinking and Acting like Data Scientist

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)



# Data Science Professional Skills: Thinking and Acting like Data Scientist (1)

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



# Data Science Professional Skills: Thinking and Acting like Data Scientist (2)

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



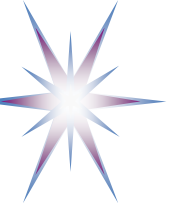
# Data Scientist and Subject Domain Specialist

- **Subject domain components**

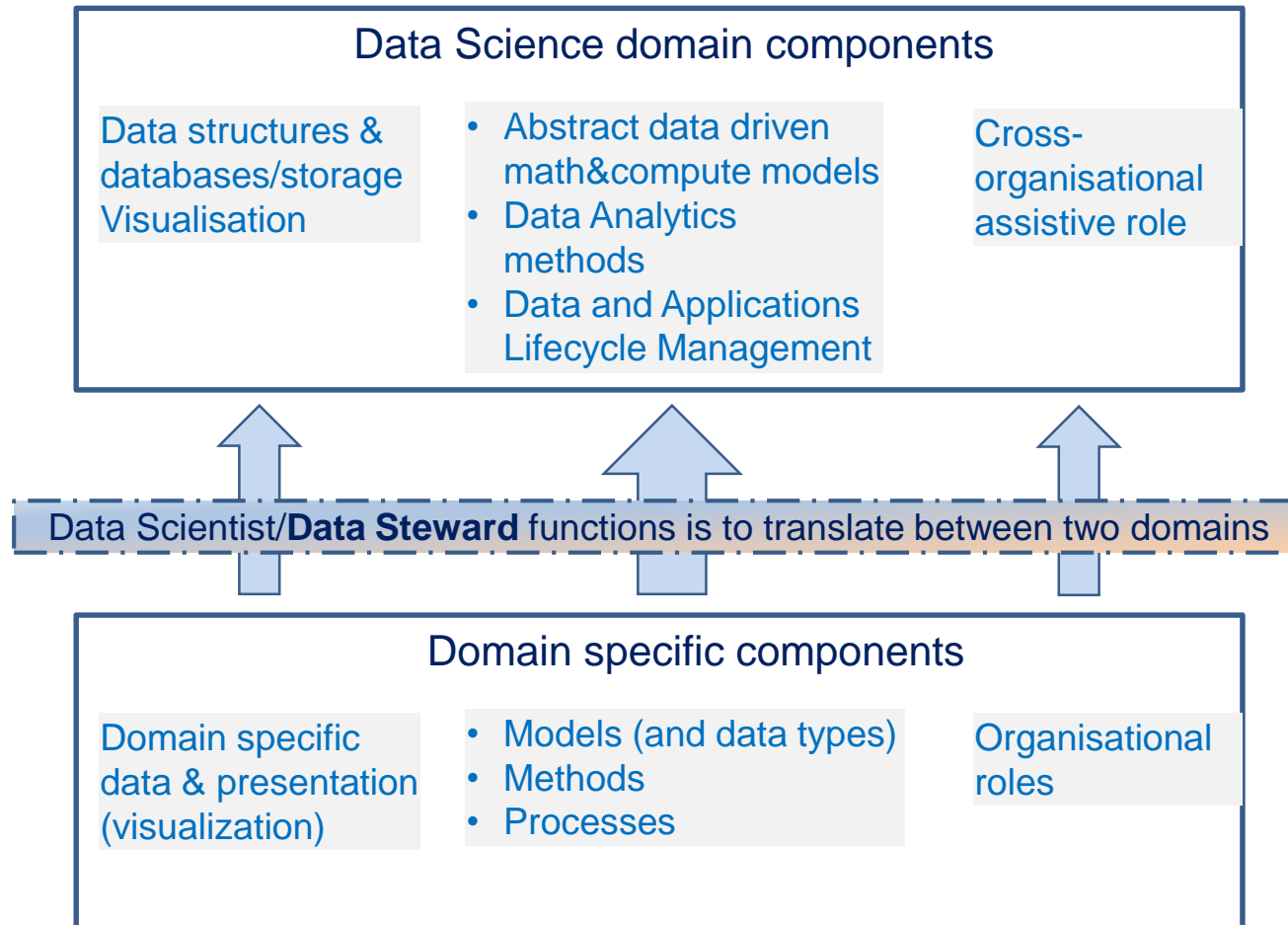
- Model (and data types)
- Methods
- Processes
- Domain specific data and presentation/visualization methods
- Organisational roles and relations

- **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

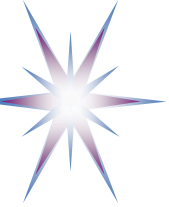


# Data Science and Subject Domains



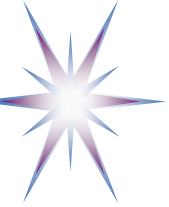
- Data Scientist role is to maintain the Data Value Chain (domain specific):**
- Data Integration => Organisation/Process/Business Optimisation => Innovation





# 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 (still active) <http://edison-project.eu/>
  - EDISON Data Science Framework Release 2 (EDSF), 3 July 2017  
<http://edison-project.eu/edison-data-science-framework-edsf>
- Data Science Community Portal (<http://datasciencepro.eu>)
  - To host future EDISON Community initiatives



# Links to Workshop Resources

- EDISON Workshop home  
<https://github.com/EDISONcommunity/EDSF/tree/master/EDSFcurriculaDesign>
- EDISON Data Science Framework Release 2 (EDSF)  
<https://github.com/EDISONcommunity/EDSF>
- (old <http://edison-project.eu/edison-data-science-framework-edsf>)

## Component documents

CF-DS – Data Science Competence Framework

[https://github.com/EDISONcommunity/EDSF/blob/master/EDISON\\_CF-DS-release2-v08.pdf](https://github.com/EDISONcommunity/EDSF/blob/master/EDISON_CF-DS-release2-v08.pdf)

DS-BoK – Data Science Body of Knowledge

[https://github.com/EDISONcommunity/EDSF/blob/master/EDISON\\_DS-BoK-release2-v04.pdf](https://github.com/EDISONcommunity/EDSF/blob/master/EDISON_DS-BoK-release2-v04.pdf)

MC-DS – Data Science Model Curriculum

[https://github.com/EDISONcommunity/EDSF/blob/master/EDISON\\_MC-DS-release2-v03.pdf](https://github.com/EDISONcommunity/EDSF/blob/master/EDISON_MC-DS-release2-v03.pdf)

DSPP – Data Science Professional profiles

[https://github.com/EDISONcommunity/EDSF/blob/master/EDISON\\_DSPP-release2-v05.pdf](https://github.com/EDISONcommunity/EDSF/blob/master/EDISON_DSPP-release2-v05.pdf)