Grant 675419 (INFRASUPP-4-2015: CSA)

EDISON
Coordinating the establishment of a new profession of Data Scientist for European Research and Industry

Andrea Manieri
Engineering Ingegneria Informatica S.p.A

EGI Community Forum 2015
12 November 2015, Bari

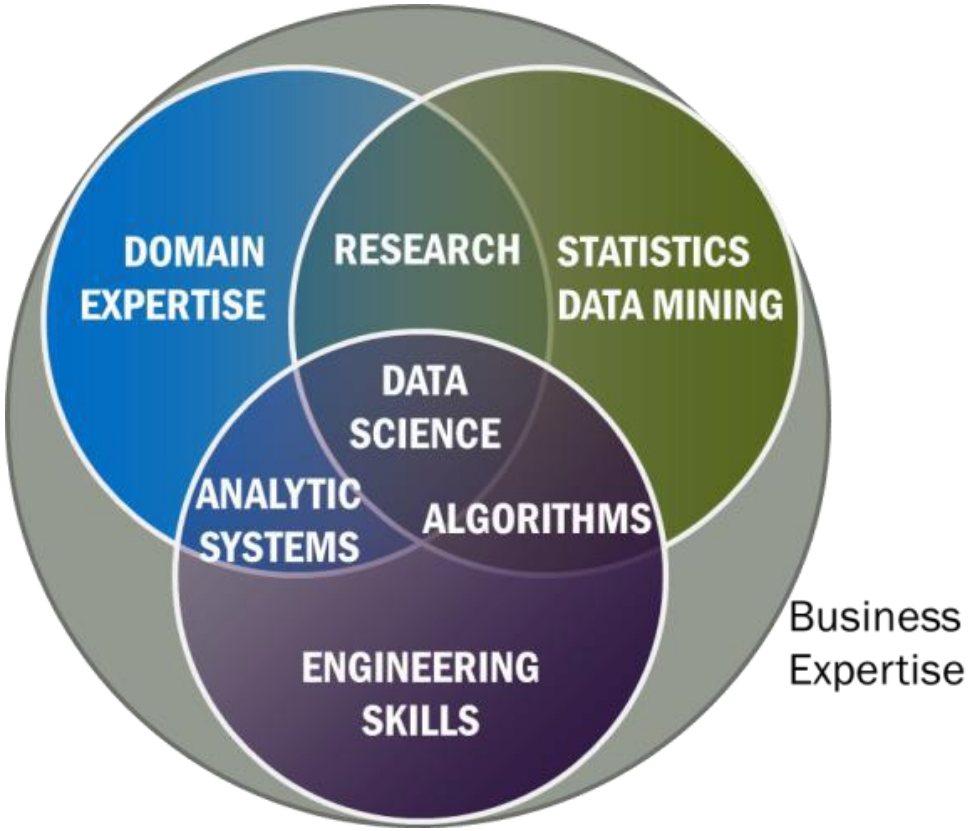
Session Objectives

- To discuss and to describe new outlines of professions in the field of data scientist for academic and industrial purpose.
- EDISON promotes an eCF v.30 compliant profile for Data Scientist and related Body of Knowledge.
- Participants are invited to
  - To contribute to the EDISON inventory and related taxonomy by providing an overview of existing curricula, training programmes and related educational resources
  - To assess the Body of Knowledge for Data Science,
  - To verify the proposed Data Science Model Curriculum
  - To support the formalization of the Data Scientist profession

EDISON Objectives

- **Data Science Curricula Foundation**
  - *Competence Framework for Data Science (CF-DS)*
  - *Data Science Body of Knowledge (DS-BoK)*
  - *Model Curriculum for Data Science (MC-DS)*
- **Education and Training Environment**
  - *Piloting Carrier development best practices (VET and HEI)*
  - Edison On-line Education Environment
- **Sustainability Model**
  - *Certification-based sustainable model*
  - *Long-term roadmap for*
  - *Community of practice - EDISON Liaison Group(s).*

Data Scientist - a mix of competences



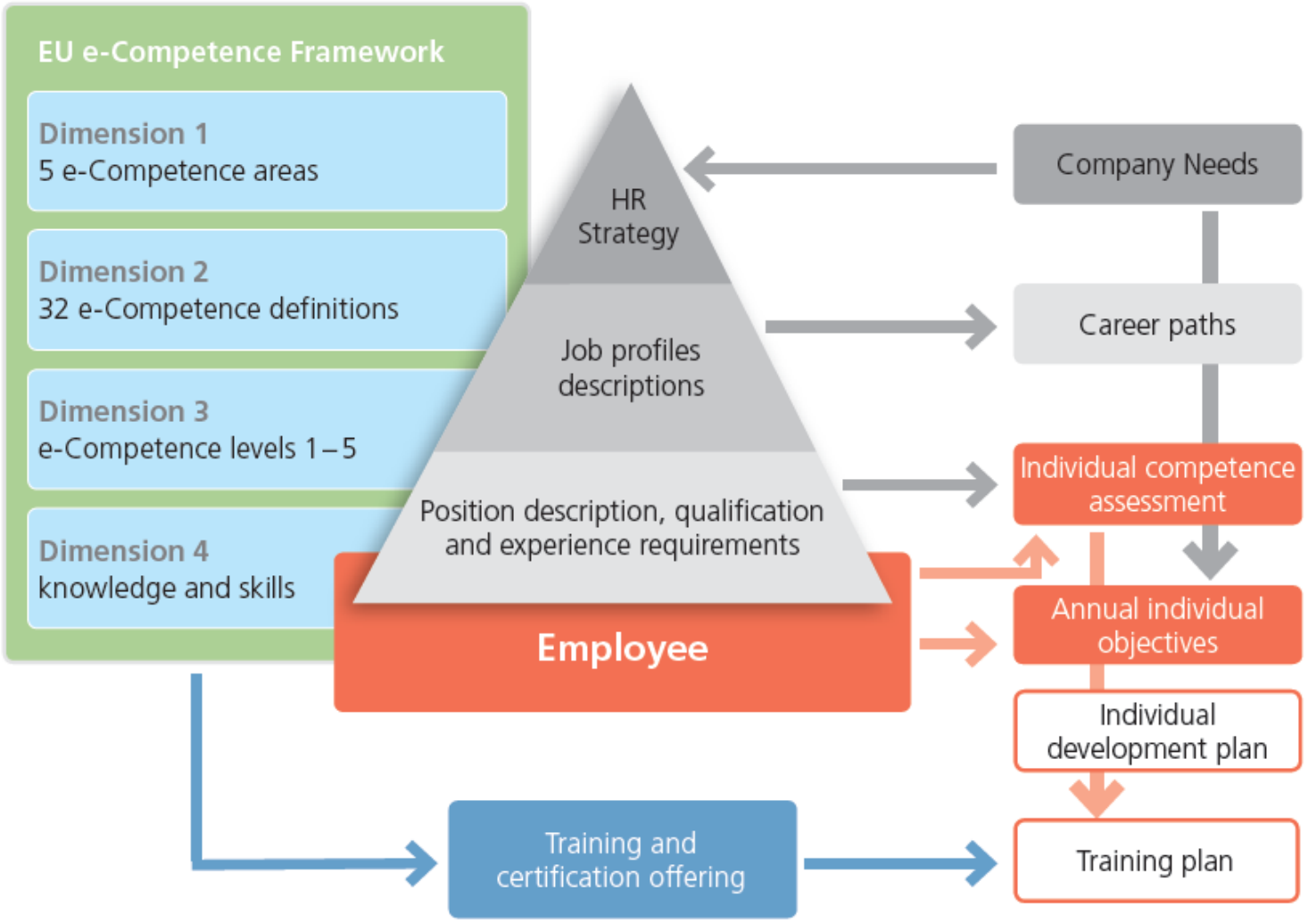
Definition by NIST Big Data WG (2014-2015)

*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**.*

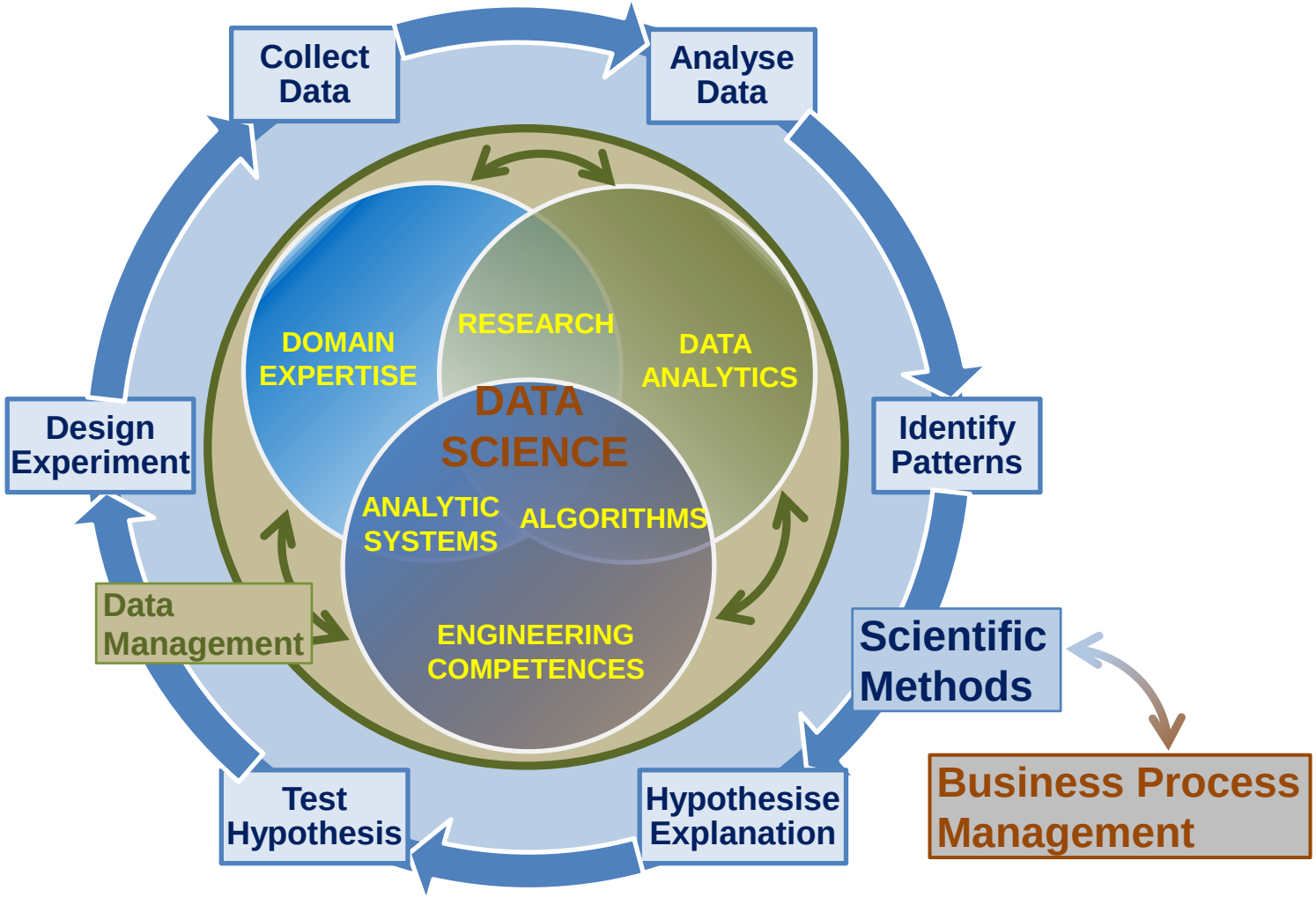
The EDISON Approach

- To provide an operational definition of Data Scientist
  - By analysing the demand side (employers)
  - By letting the supply side (trainers) to bridge the gap
- Mixing qualitative approach with quantitative analysis
  - By ensuring research disciplines and market sectors coverage
  - By gain consensus and engaging stakeholders
- To pave the way for long-term results
  - By demonstrating the soundness of the attempt
  - By providing concrete and useful results

Competence Framework - Data Scientist



Legacy: Data Science Body of Knowledge



Grant 675419 (INFRASUPP-4-2015: CSA)

Process Groups

Knowledge Areas	Data Identificazione And creation	Data Access and Retrieval	Data Cleansing	Data Processing	Data Preservation And curation	Data Analysis	Data Visualisation And communication
Data Mngt							
Algoritmics And coding							
Engineering (ICT infra And tools)							
Maths And Stats							

Process Groups

Knowledge Areas	Data Identificazione And creation	Data Access and Retrieval	Data Clearning	Data Processing	Data Preservation And curation	Data Analitycs	Data Visualisation
Business/ researdh needs							
Team And Stk mgnt							
Project Mgnt							
Domain specific							

Legacy: Model Curricula – Data Scientist

Programming & Digital Technologies skills

- Computational System (concurrency and distributed systems)
- Programming Languages and Paradigm (C, R, OpenMPI, Python)
- Tools (Big Data, Cloud Platforms, Databases, Sensors,…)

Domain-specific & Analysis skills

- Data sources (Open and Linked Data), preservation and curation
- Standards and Certification for the domain
- Interpretation skills (Knowledge extraction)

Maths and Statistic competences

- Statistics and Probability; Algebra and Calculus
- Machine Learning
- Data Mining And Business Intelligence

Business-orientation & Communication

- Marketing and Market Analysis (Innovation leadership)
- Legal and Ethical elements
- Data Visualisation and Communication Skills

Legacy: Model Curricula – Data Scientist

1 BASIC COMPETENCES

At University grade (1° or 2° level) in Computer Science, Engineering, Science, Physics, Maths, Statistic Economics

2 ADVANCED SKILLS

University Master or PhD or similar (eg. Research Project) or Working experience and certified training.

3 HANDS-ON PRACTICE

On-the-job, either in Industry or Research field

4 DATA SCIENCE PROFESSIONAL

Certified by third-party independent entity by certified experience and examination

Legacy: Model Curricula – Data Scientist

1 BASIC COMPETENCES

Basics in Maths, Stats, Physics, (distributed and Parallel) Computing and Elettronics (STEM) + Arts and Innovation Leadership

2 ADVANCED SKILLS

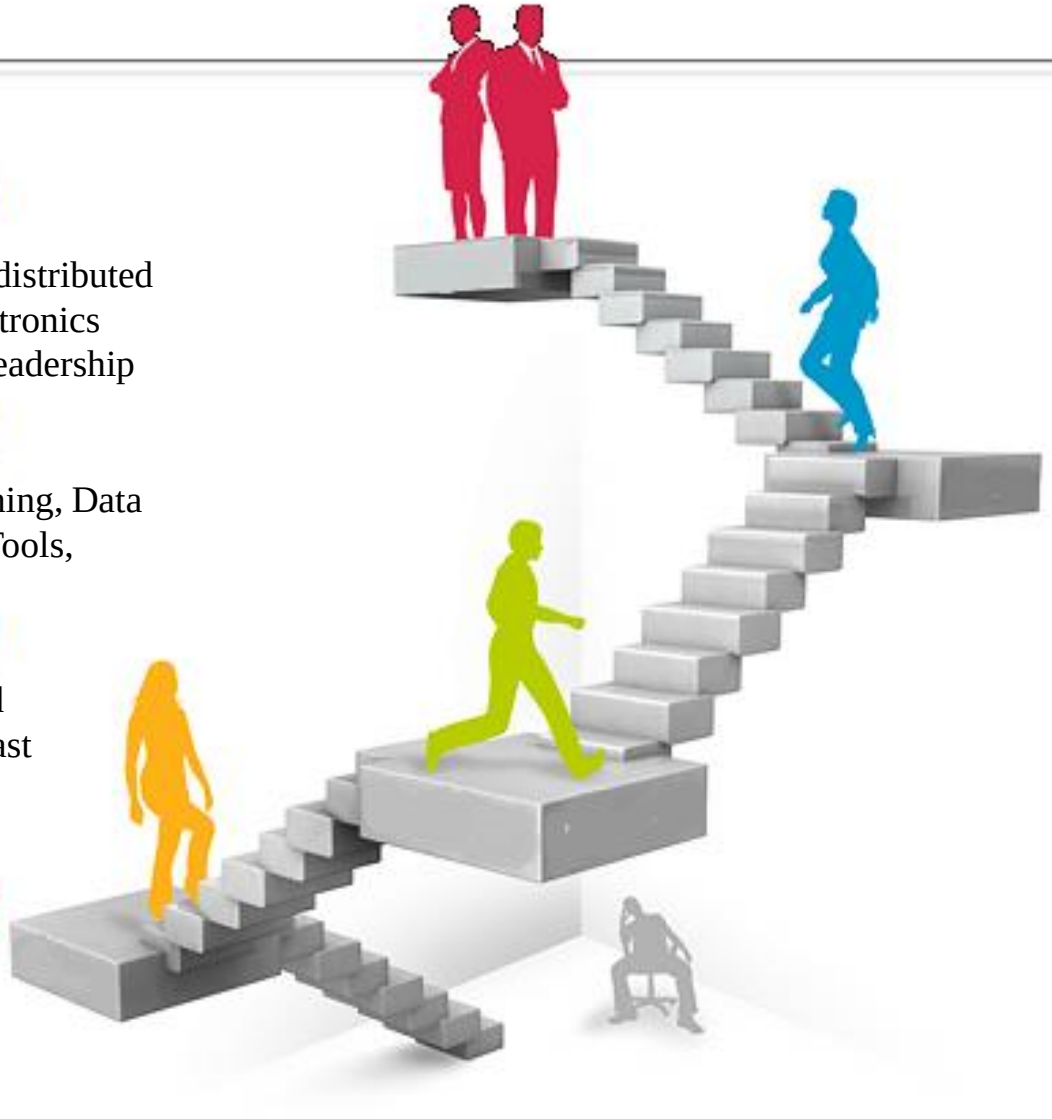
Specialisation in Machine Learning, Data Analytics and Cloud/Big Data Tools, Sensors, Market Analysis

3 HANDS-ON PRACTICE

Domain specific knowledge and hands-on projects (include at least two years experience in the selected domain

4 DATA SCIENCE PROFESSIONAL

Certification by third party of the acquired competences



Thanks for your attention