EDISON Data Science Framework (EDSF) for Data Science Professional Education and Training

Session “Collaboration of Certification and Education on Data Science”
IFIP 60th Anniversary “Future of Information Processing”
27 September 2021

Yuri Demchenko, EDISON Project
University of Amsterdam
EDISON Community Initiative

Building the Data Science Profession

Maintaining the EDISON project legacy

- Community Initiative supported by University of Amsterdam since 2017
- Further development in EU projects MATES, FAIRsFAIR

EDISON Products for Data Science Skills Management and Curriculum Design

- **EDISON Data Science Framework (EDSF)**
  - Release 3 components CF-DS, DS-BoK, MC-DS, DSPP
  - Compliant with EU standards on competences and professional occupations e-CFv3.0, ESCO
- Skills development and career management for Core Data Experts and related data handling professions
- Academic programmes and professional training courses (self) assessment and design
- Individual competences benchmarking and Data Science team design
- Contribution to European and international bodies defining competence, Body of Knowledge, curricula: ITPE, CEN, ABAC, DFN, Japan IFIP
EDISON Data Science Framework (EDSF) – Core components and community maintained services

- **CF-DS** - Data Science Competence Framework (Part 1)
- **DS-BoK** - Data Science Body of Knowledge (Part 2)
- **MC-DS** - Data Science Model Curriculum (Part 3)
- **DSPP** - Data Science Professional profiles (Part 4)
- **DS-UCG** - EDSF Use cases and Applications (Part 5)

**CF-DS Competence Groups**
- **DSDA** - Data Science Analytics
- **DSENG** - Data Science Engineering
- **DSDM** - Data Management and Governance
- **DSRMP** - Research Methods and Project Management
- **DSDK** - Domain related Knowledge

**DS-BoK Knowledge Area Groups**
- **KAG1-DSA**
- **KAG2-DSE**
- **KAG3-DSDM**
- **KAG4-DSRM**
- **KAG5-DSBA/DSDK**
EDSF for Education and Training

• Customised curriculum design for target professional or learners group

• Variety of Data Science programs – Inventory/list still to be created

• Data Science related courses (based on DS-BoK and MC-DS)
  – Big Data Infrastructure Technologies for Data Analytics (BDIT4DA)
    • Offered at UvA, ABS, NTUU “KPI”
  – Data Management and Governance and FAIR data principles
    • Offered as part of university curricula and professional training
  – Data Science Analytics Foundation (DSAF)
    • Offered as part of professional training
  – Research Methods and Data Science Project Management (DSRMP) – in development
  – Data Science Professional Issues – in development
From Individual Competences Benchmarking to Career Path Advise and Data Science Team Composition

Individual Education/Training Path based on Competence benchmarking:

- Red polygon indicates the chosen professional profile: Data Scientist (general) – DSPP04
- 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
Demand for Data Steward is growing in research and industry
- Indeed.com snapshot Sept 2020 and May 2021 – 60+ vacancies in NL, UK and 400+ in USA
- Competence are focused on data management and domain specific knowledge (including Data Warehouses)

DSP-CF defined based on EDISON Data Science Framework
Data Stewardship additional competences added
DSP-BoK: Body of Knowledge revised and extended with the specific Data Management and Governance competences

FAIR (Findable – Accessible – Interoperable – Reusable) data principle addressed in CF-DSP and DSP-BoK
MATES Project (2018-2022): Digital and Data Skills for Maritime Industry

• Strong demand for digital and data skills in maritime industry
• Training curricula defined based on EDSF and European Digital Competence Framework DigComp 2.0 (2016)
• Set of training courses covering topics
  – Big Data Technologies for Data Analytics and Industry Automation (BDIT4DA)
  – Industrial Data Spaces and Data Management and Governance (DMG)
  – Data Science Analytics Foundation (DSAF)
Call for contribution EDSF Release 4

- Call for EDSF Updates and Part 4 Use Cases and Applications
  [https://github.com/EDISONcommunity/EDSF/wiki/EDSFhome](https://github.com/EDISONcommunity/EDSF/wiki/EDSFhome)
  - Deadline first stage 30 November 2021
Topic 1: Current collaboration with other organizations related to DS education and/or DS certification

- Regular collaboration in the framework of existing projects and initiatives
  - FAIRsFAIR project on Data Stewardship Professional Competence Framework
  - MATES Project on Digital and data skills for maritime industry
  - EOSC (European Open Science Cloud) on skills management and career path for European Research Area (ERA) and EOSC

- Communities
  - Research Data Alliance (RDA) Interest Group on Data Stewardship Professionalisation and IG on Education and Training on Handling Research Data
  - Big Data Value Association on Data Scientist profiles and Badges
Topic 2: Future plans related to DS education and DS certifications

- Certification: Major commercial IT certification companies are primarily focused on updating and extending their current certification programs, no new certification development is planned.

- EDSF Release 4 planned for December 2021
  - New Part 5: Use cases and Applications
    - Including Data Stewardship Professional Competence Framework and Body of Knowledge
  - Updated DS-BoK

- Extending use of EDSF for common tasks for competences and skills management, curriculum design, career path management
  - Extending to Artificial Intelligence Competence Framework
    - Primarily based on job market analysis and existing AI programs analysis