Defining Data Science Professions Family

Yuri Demchenko, Adam Belloum, Wouter Los, University of Amsterdam, Netherlands
Steve Brewer, University of Southampton, UK

Abstract

The increasing use of data intensive technologies in modern research requires new competences and skills and demands for new professions that should support all stages of the research data lifecycle from data production and input to data processing, storing, and obtained scientific results publishing and dissemination. This poster presents the results of the EDISON project that proposes the Data Science Professions (DSP) family definition based on the analysis of research and industry demand and in accordance to existing European standards and practices for professional profiles definition. The competences and skills required for different professions are defined in accordance with the Data Science Competence Framework (CF-DS) proposed in the project.

Introduction

Modern research requires new types of specialists that are capable to support all stages of the research data lifecycle from data production and input to data processing, storing, and scientific results publishing and dissemination, which can jointly defined as the Data Science Professions (DSP) family. The paper refers to the Data Science Competence Framework (CF-DS) that provides a basis for defining the Data Science Professions family that includes different occupation groups that are typically present in modern research organisations and industry. The project proposes extensions to the recently introduced European Skills, Competences, Qualification and Occupations (ESCO) taxonomy as described below.

Data Science Competence Framework (CF-DS)

The Data Science Competences Framework (CF-DS) is a cornerstone of the EDISON Data Science Framework and used for defining such components as Data Science Body of Knowledge (DS-BoK) and Data Science Model Curriculum (MC-DS).

The CF-DS includes the following competences groups

- Data Analytics including statistical methods, Machine Learning and Business Analytics
- Engineering: software and infrastructure
- Data Management, Curation, Preservation
- Subject/Scientific Domain competences and knowledge
- Scientific or Research Methods (for research professions) and Business Process Management (for business related professions)

Knowledge of the scientific research methods and techniques makes the Data Scientist profession different from all previous professions.

Data Science Professional Profiles Definition

The proposed Data Science professional profiles definition is based on the analysis of the research and industry demand in data related professions as well as current companies practices (IBM-CDO, 2016). The identified professional profiles are classified using ESCO taxonomy, and necessary extensions are proposed to support the following hierarchy of the data handling related occupations:

1) Managers that include but limited the following occupations Data Science manager, Data Science infrastructure manager, Research Infrastructure manager
2) Professionals: Data Scientist, Data Science Researcher, Data Science Architect, Data Analyst, etc.
3) Professional (database): large scale (cloud based) scientific database designers and administrators;
4) Professional and clerical (data handling/management): Data Stewards, Digital data curator, Digital Librarians, Data Archivists;
5) Technicians and associate professionals: Big Data facilities operators, scientific database/infrastructure operators.

The competences and skills required for different professions are defined in accordance with the Data Science Competence Framework (CF-DS) proposed in the project. The poster will provide an example of mapping CF-DS competences to identified data handling related occupations.
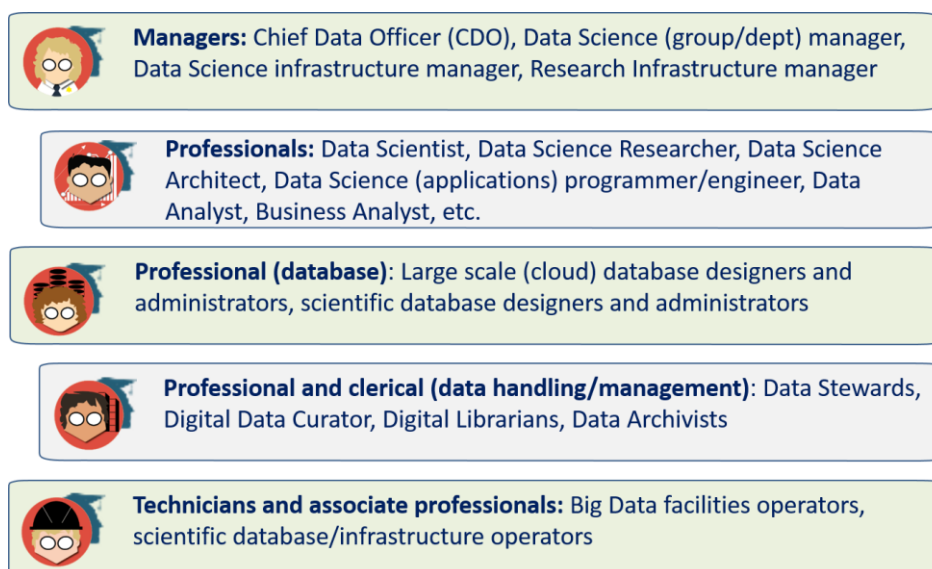

Figure 1. Data Science Profession family groups

Further Developments

The proposed DSP taxonomy will be validated with the community survey that will be conducted by the EDISON in September 2016. The initial version of the DSP definition was proposed in the project deliverable D2.2 and the first public version is available for community comments and contribution (DSP, 2016). The project will undertake necessary steps to achieve formal standardisation of DSP as a part of the ESCO taxonomy.

Acknowledgements

Competing Interests

The authors declare that they have no competing interests.

References [alphabetically by surname, then year]

EDISON Project: Building Data Science Profession [online] http://www.edison-project.eu/

Andrea Manieri, et al, 2015, Data Science Professional uncovered: How the EDISON Project will contribute to a widely accepted profile for Data Scientists, Proc. The 7th IEEE International Conference and Workshops on Cloud Computing Technology and Science (CloudCom2015), 30 November - 3 December 2015, Vancouver, Canada

ESCO, 2016, ESCO (European Skills, Competences, Qualifications and Occupations) framework [online] https://ec.europa.eu/esco/portal/#modal-one

CF-DS, 2016, Data Science Competence Framework (CF-DS). EDISON draft V0.6, 10 March 2016 [online] http://www.edison-project.eu/data-science-competence-framework-cf-ds

DS-BoK, 2016, Data Science Body of Knowledge (DS-BoK). EDISON draft V0.1, 20 March 2016 [online] http://www.edison-project.eu/data-science-body-knowledge-ds-bok

DSP, 2016, Data Science Professional profiles definition (CF-DS). EDISON draft v0.1, 11 July 2016 [online] http://www.edison-project.eu/data-science-professional-profiles-dsp

IBM-CDO, 2016, Cortnie Abercrombie, What CEOs want from CDOs and how to deliver on it [online] http://www.slideshare.net/IBMBDA/what-ceos-want-from-cdos-and-how-to-deliver-on-it