

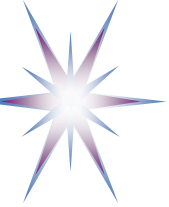
Draft – Document for discussion

Changes will apply without notice

Building FAIR Data Competence Framework for Higher Education
as a way to professionalise Data Stewardship
(FAIRsFAIR T7.3 work)

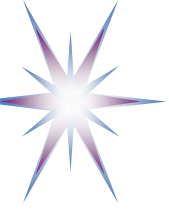
Yuri Demchenko
University of Amsterdam
FAIRsFAIR WP7 Workshop
8-9 October 2020





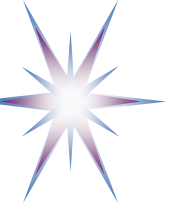
Outline

- Goals and tasks for this workshop
- Landscape overview: Data Management, Data Stewardship, FAIR management
 - Project, initiatives, Frameworks
 - Training, Education, Policies
- Approach for FAIR4HE/DS@HE
 - Evidence based and community driven: Job Market and Landscape
 - Understand landscape and build consensus for Data Stewardship professionalisation
- Mapping between existing Data Stewardship and FAIR competences definition
- Proposed FAIR4HE structure
 - Building on the EDSF and Data Steward professional profile



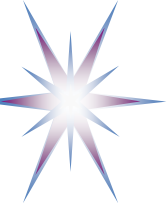
Goals and Tasks – Discussion items

- Review landscape in Data Stewardship and FAIR education and training
- Identify all actors, components of the FAIR data ecosystem
- Propose workable approach to implement FAIR and Data Stewardship in university curricula
- Interconnect and link existing frameworks
- Propose consistent vision for universities
- Skills vs Competences vs Professional Profiles vs Knowledge vs Curriculum elements



Suggested Approach (May 2020)

- Review and integrate existing frameworks
- Apply proven methodologies
 - EDISON EDSF
 - EOSCpilot, ELIXIR, DeIC
- Match/comply to GO FAIR initiative
 - FAIR principles and Maturity Framework
- Verify with main use cases (e.g. as defined in GO FAIR, ELIXIR, or DeIC)
- Investigate job market demand for competences and skills (e.g. LinkedIn Jobs, IEEE Jobs, Indeed.com, Monsterboard, Glassdoor)
- Link to industrial Data Management and Quality Assurance Frameworks
 - Data as part of business processes and industry digitalisation
- Match to main academic curricula:
 - Main courses vs professional issues
 - Computer Science vs Interdisciplinary
- Train teachers (use NL BKO as a pilot, find other champion countries)



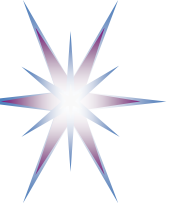
FAIR4HE Design principles – Leverage EDSF

- P1. Higher Education scope: Bachelor and Master programs, Learning Outcomes, mastery levels
 - Empower graduates high freedom in career selection and development
- P2. Linked to educational methodology
 - Master/Bachelor
 - Professional and PhD training
 - Continuous education and VET
- P3. Address Job market demand in Data Steward and FAIR data management
 - Job market analysis as evidence and advice
- P4. Data Stewards, Data Managers, Data Scientists and others are part of the Data Science professional Family (as defined in EDSF)
- P5. Verified with the standards (in particular DMBOK and ESCO)
- P6. Consensus with the existing frameworks
 - Trust them and take best of them



FAIR adoption and Ecosystem Sustainability Elements

- FAIR must be accepted by all roles in organisational data management and governance process
 - FAIR must be endorsed by top management C-level
 - Roles and responsibilities to be defined and staffed
 - Inter-role functions as factor for modern agile organisations
- FAIR must be adopted for the whole Research/industrial Data lifecycle
- FAIR must be practiced by all participants along data lifecycle and specifically started from the data producers i.e. researchers or facility operators or sale agents
- FAIR must be supported by infrastructure and tools
- FAIR must be embedded into applications development
- Organisational capability and capacity management
- **Education and training – To enable them all**
 - **Basic academic and professional education + continuous education**



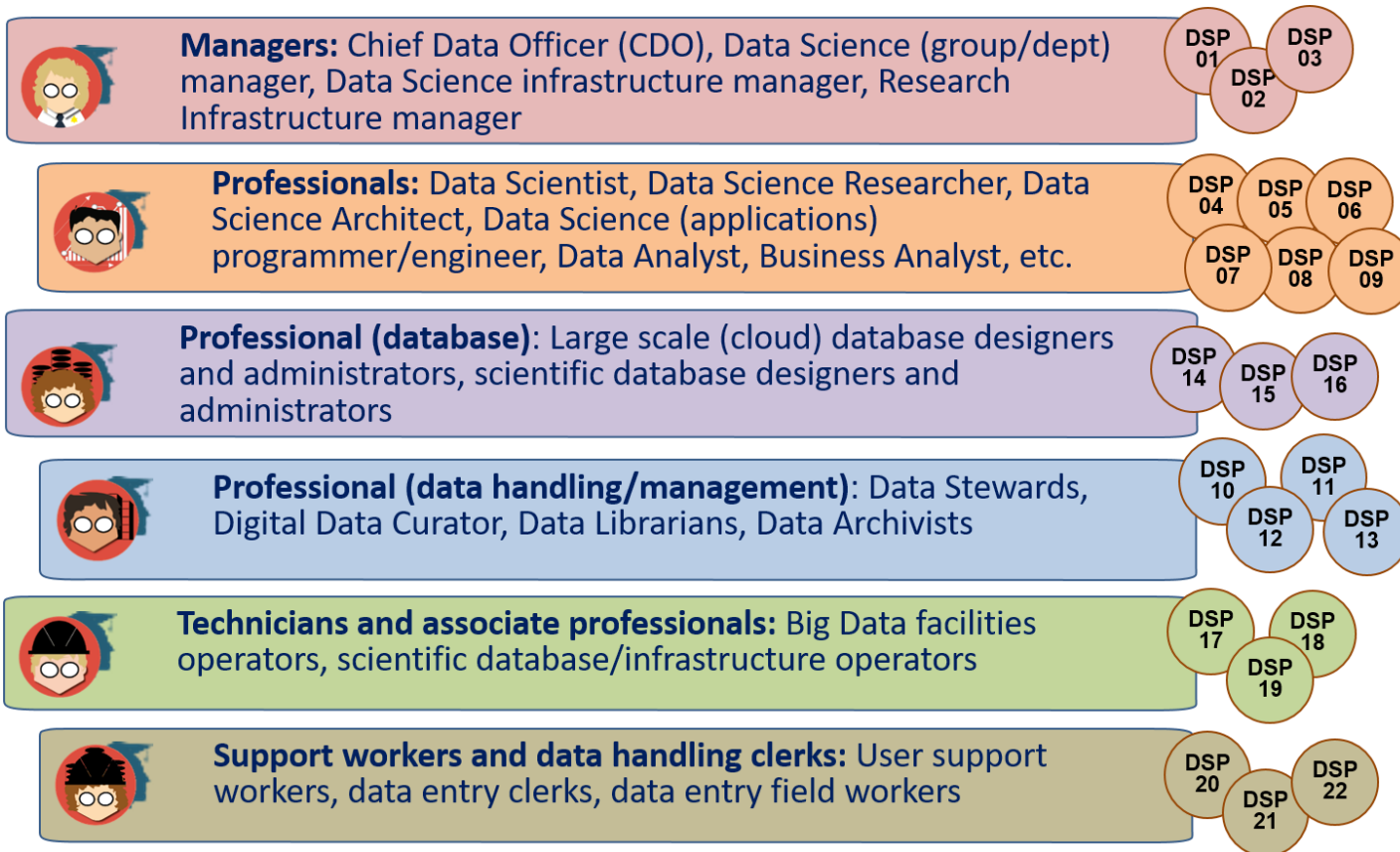
FAIR from the technical point of view

- Findable
 - Metadata and PDI – infrastructure and tools
 - Registries and handles resolution, API
 - Policies and SLA
- Accessible
 - Repositories and data storage: infrastructure and management
 - Policy and access control: infrastructure and API management
 - Data access protocols
 - Usage Policy and Sovereignty
 - Data protection, compliance, privacy and GDPR
- Interoperable
 - Standard data formats
 - Metadata and API
 - FAIR maturity level and certification
- Reusable
 - Data provenance and lineage
 - Preservation
 - Metadata, PID and API – linked or embedded into datasets

This motivates Data Stewards interaction with both **Data Analytics and Applications developers** roles and **Data Infrastructure** roles

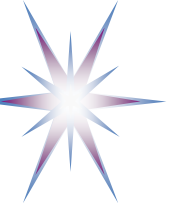
- Consequently related competences from Data Stewards are needed

Why we need to view the whole Data Science Professional Family?



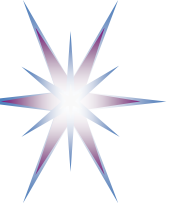
- Career path
- Team composition
- Agile Team working
- Education programs alignment
- HR capacity building and management
- Employability/mobility for job seekers
- Easy to market

EDISON Data Science Professional Profiles family EDSF, Part 4

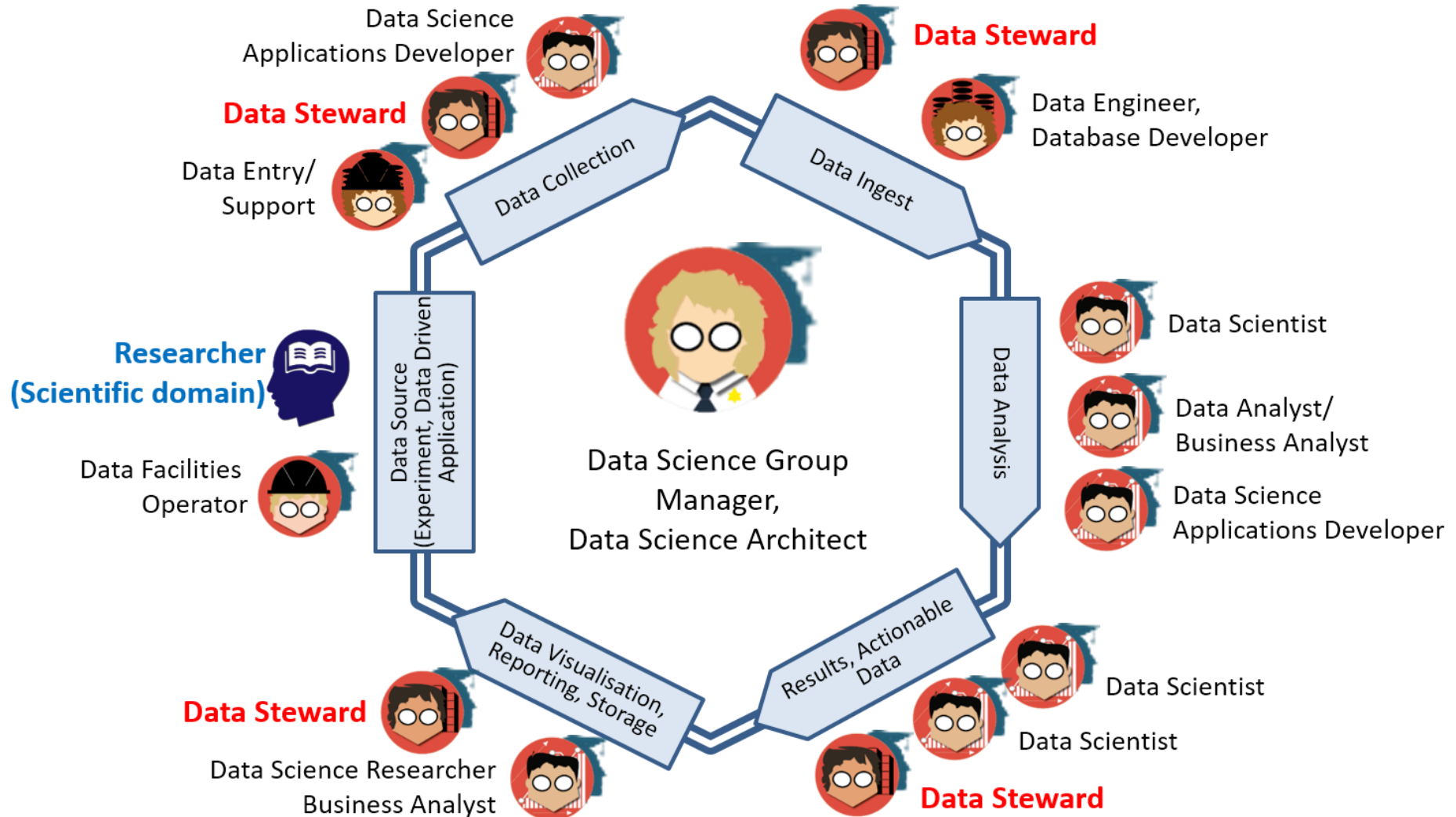


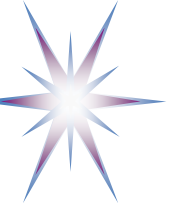
FAIR Data Management and Organisational Roles

- FAIR data principles to be adopted cross organisation for the whole data lifecycle
- Data Governance and Data management
 - Data Stewards and CDO
 - Data policy and data delivery agreements
- Data Infrastructure and tools for data storage and handling
 - Storage, database engineers/managers
 - Metadata and PID services, Master data and Reference data
- Data Analysis
 - Data Scientists, Data Architects, Application developers
- Data collection
 - Researchers, Data Engineers, data entry workers
- Data preservation and curation
 - Data curators, Data Custodians/Archivists
- Data publication, sharing access
 - Data Stewards, Data Curators



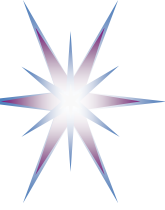
Building Data Science Team – Variety of Roles (Example)





FAIRsFAIR: FAIR Data Competence Framework – Pillars and Cooperation

- EOSCpilot FAIR4S Data Stewardship Competence Framework
- ELIXIR Data Stewardship Competence Framework
- DeIC and DM Forum: Report on National Coordination of Data Steward Education in Denmark
- GO FAIR Data Principles and Maturity Framework
- Belmont Forum Curriculum recommendations
- DAMA BoK (2007) – DAMAI Data Management Body of Knowledge
- EDISON Data Science Framework (EDSF) and EDISON Community Initiative

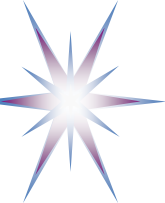


EOSCpilot FAIR4DS Data Steward Competences

- Organisational capabilities for sustaining FAIR data across projects
- Stewardship skills to deliver FAIR data from projects
- Data Stewardship Roles and **Shared responsibility**:
 - Data Stewards and researchers
- 59 competences grouped in
 - 3 general groups
 - Govern and assess
 - Scope and resource
 - Advise and enable
 - 6 Data (curation) lifecycle process stages
 - Plan and design
 - Capture and process
 - Integrate and analyse
 - Appraise and preserve
 - Publish and release
 - Expose and discover



[ref] EOSCpilot D7.5 Strategy for sustainable development of skills and capabilities



ELIXIR - Data Stewardship Competency framework (courtesy ELIXIR Project)

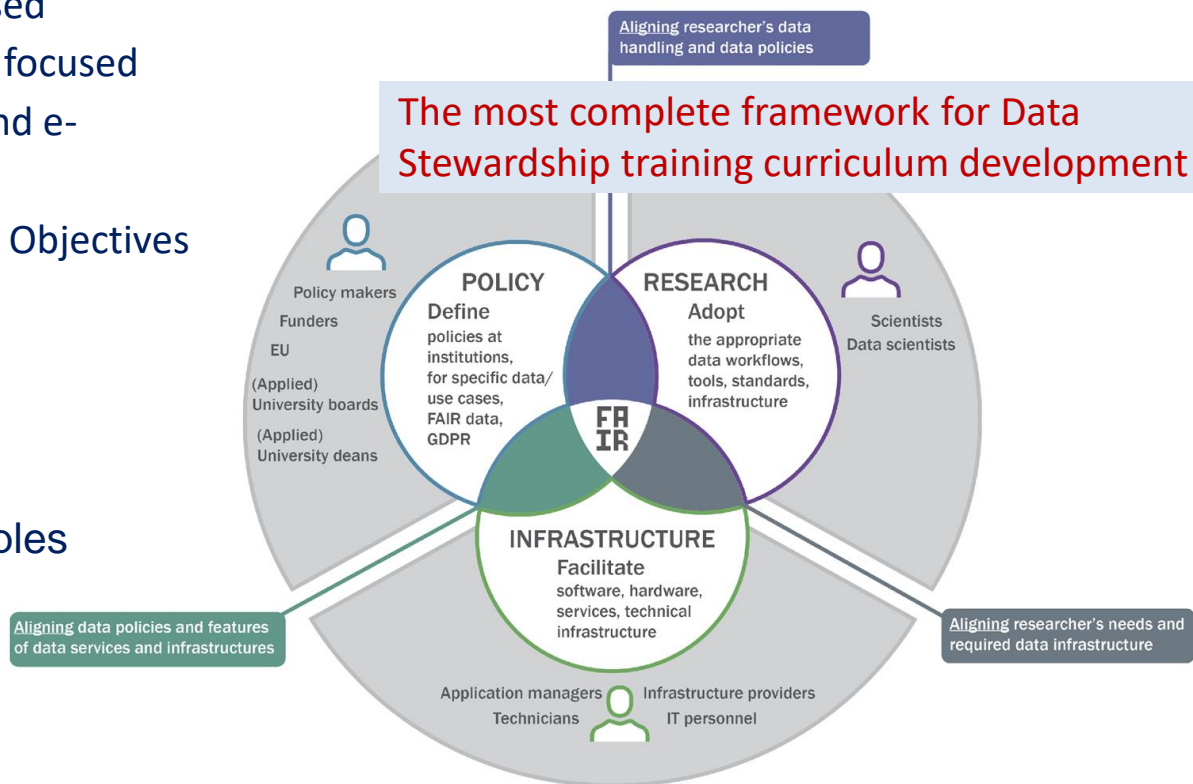
Data Steward Roles and Competence Profiles

- **Policy:** institute and policy focused
- **Research:** project and research focused
- **Infrastructure:** data handling and e-infrastructure focused
- Activities – Knowledge – Learning Objectives

Competence groups

- 1) Policy/Strategy
- 2) Compliance
- 3) Alignment with FAIR data principles
- 4) Services
- 5) Infrastructure
- 6) Knowledge Management
- 7) Network
- 8) Data sharing

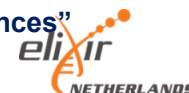
Implementation areas for data stewardship

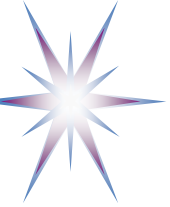


ZonMw & ELIXIR-NL funded project “Towards FAIR Data Steward as profession for the Life Sciences”

○All project output: <https://zenodo.org/communities/nl-ds-pd-ls/>

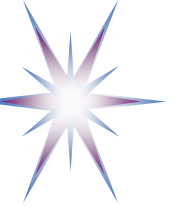
○Final report (Oct 3, 2019): <https://doi.org/10.5281/zenodo.3471707>





DeIC 2020 and DM Forum: Report “National Coordination of Data Steward Education in Denmark”

- Report focuses on the future education of Data Stewards
https://www.deic.dk/sites/default/files/Data%20Steward%20Education%20in%20Denmark_0.pdf
- Analysis based on
 - LinkedIn profiles analysis – 74 profiles – March 2019
 - Job vacancies database in Denmark – 119 vacancies of Data Scientist – March-April 2019
 - Questionnaire September 2019 – 86 complete responses (42 partial)



DeiC Data Stewardship Curriculum principles

- Four roles for Data Stewards

- Administrator
- Analyst
- Developer
- Agent of change

- Competences defined

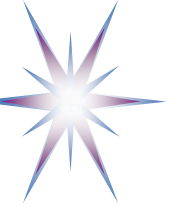
- 6 competence groups, 22 competences

- Three modes for Data Stewards education (using personas mode)

- Student with Bachelor degree
- Student with PhD and equivalent
- Continuing and professional education

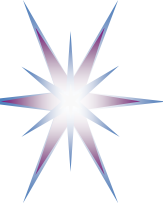
- Education/Training Format

- Master – 1-2 yrs
- PhD/Practitioner
 - Professional training
 - Continuous education
 - Competence Centers
 - MOOC/Self education



Belmont Forum

- Study: Skills Gap Analysis e-Infrastructures and Data Management in Global Change Research by The Belmont Forum e-Infrastructure and Data Management (e-IDM) group
- Recommendations for Multi-, inter- and trans-disciplinary curriculum design
- Main subject groups (A = first year PhD; B = final year PhD; C = postdoc, D = mid-career, Principal Investigators)
 - i. Programming for data intensive research (for those who already code) (A)
 - ii. Environmental data: expectations and limitations (A)
 - iii. Introduction to visualising environmental data (A and B)
 - iv. Data management (A and B)
 - v. Interdisciplinary data exchange (B, C and D, mixed classes engineers, social and environmental scientists)
- OPTIONAL
 - vi. Software development ideas for scientific coding (C and D)
 - vii. Object orientated programming (C and D)
 - viii. Introductory data science topics (C and D)
 - ix. Data organisation (A to D)



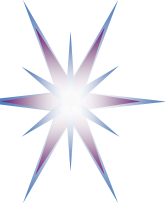
Data Science Curriculum Structure by German Gesellschaft für Informatik (GFI)

14 fields of competences, including

- basic and advanced mathematics
- informatics
- cryptography and security
- data ethics and data privacy
- data governance
- data integration
- data visualisation
- data mining
- machine learning
- Business intelligence
- domain-specific applications and communicating with experts
- implementing data science within an organisation

Mastery levels applied to
curriculum courses for Master
MSc programs:

- (1) Understanding
- (2) Using
- (3) analysing

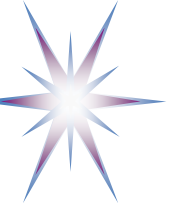


Data Stewardship in Research and FAIR Principles – GO FAIR and GO TRAIN

- FAIR Initiative by Dutch Techcentre for Life Science (DTLS) – Prof. Barend Mons
 - Supported by Germany, France, Spain, UK, USA
 - Part of Horizon 2020 Programme
- FAIR Principles for research data:
Findable – Accessible – Interoperable - Reusable
- Data Stewards as a key bridging role between Data Scientists as (hard)core data experts and scientific domain researchers (HLEG EOSC report)
- Current definition of the Data Steward (part of Data Science Professional profiles)
 - Data Steward is a **data handling and management professional** whose responsibilities include planning, implementing and managing (research) data input, storage, search, and presentation.
 - Data Steward creates data model for **domain specific data**, support and advice domain scientists/ researchers during the whole research cycle and data management lifecycle.

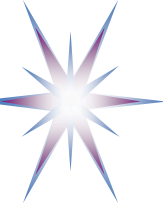


HLEG report on
European Open
Science Cloud
(October 2016)



GO FAIR and GO TRAIN

- GO TRAIN <https://www.go-fair.org/go-fair-initiative/go-train/>
- The Objectives, Scope and Activities of a Possible GO TRAIN Implementation Network – 7 Feb 2018
 - Skills needed + Training materials certification/endorsement + Coordination
- GO TRAIN workshop, 25 November 2019, ZBW Hamburg: Report from World Cafe Sessions at the
 - Incentives
 - FAIR training materials
 - Data Management and Stewardship as organisational role

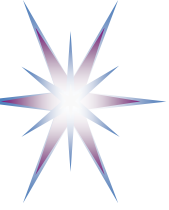


Existing courses to use and recommend – Estimate potential curriculum fitness

- RDA
 - Research Data Management – Catalog to be created
 - Old catalogs from 2016-2018 (still useful)
- Domain specific courses by EOSC and ESFRI projects - many
 - ELIXIR
 - ENVRI
 - DARIAH
 - PARTHENOS
 - Others
- Universities
 - See Deliverable D7.2
- Projects and Associations
 - MATES – Digital and Data literacy courses
 - Belmont Forum?

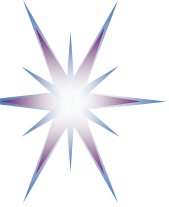
Learning Units (LU) and Learning Outcomes (LO)

- Match Body of Knowledge and Model Curriculum
- Data Steward Professional Profile -> Competences -> Learning Outcomes -> Learning Units -> BoK Knowledge Units



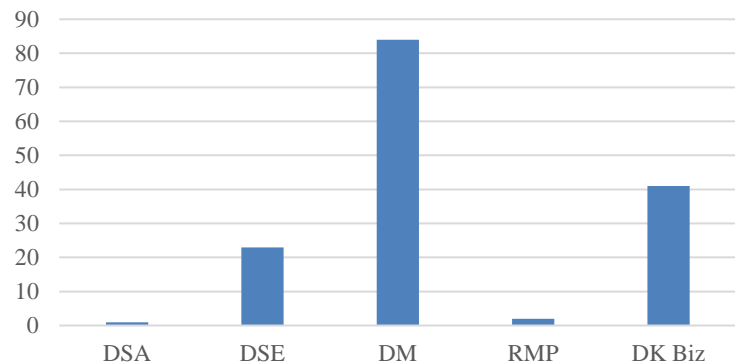
Data Stewards – Job market review

- Date 30 August – 1 September 2020
- Indeed.com – NL, UK, DE, USA
- Days open: >50% more than 30 days
- Data Steward and related vacancies
 - NL – 51, UK – 30+, DE ~20, US – 300+
 - Key skills snapshot
- Sample vacancies detailed analysis
 - NL, UK – 12, US - 6

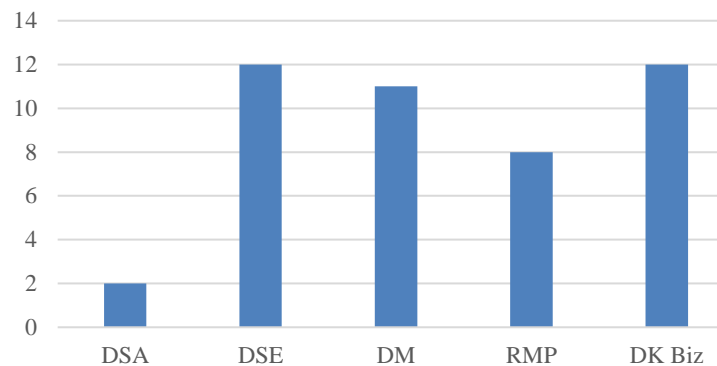


Vacancies profile – By Data Science Competence Groups

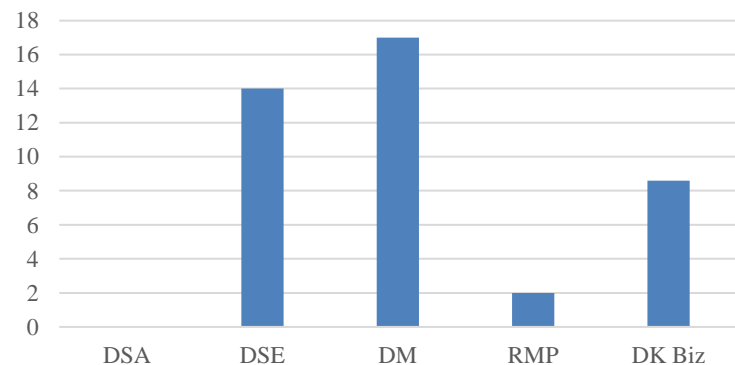
Functions/Abilities - Competences



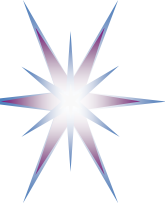
Knowledge topics



Domain related Experience/skills

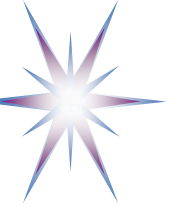


- DSA – Data Science and Analytics
- DSE – Data Science Engineering
- DM – Data Management and Governance
- RMP – Research Methods and Project Management
- DK Biz – Domain Knowledge, particular Business domain



Important Knowledge Items extracted from Job vacancies (indeed.com – NL, DE, UK, US, Sept 2020)

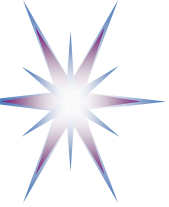
- Data Management techniques
- FAIR data principles
- Data Management and Data Governance principles
- Data integrity
- Metadata, PID and linked data
- Ontology and Semantics
- FAIR metrics and Maturity framework, FAIR certification
- Research methods
- Project management
- Data compliance regulations and standards
- Data privacy law
- GDPR
- Business process management
- Marketing
- Banking financial services and data management
- Multilevel Bill of Materials
- Ethics
- Data Warehouses
- Version control system
- Master Data Management (MDM) and Reference Data
- Visual Basic for Applications (VBA) and interface design
- WebAPI use for data access, collection and publishing
- DevOps, Agile, Scrum methods and technologies
- Data formats, standards
- Data analysis and visualisation tools
- Data lifecycle, lineage, provenance
- Data modeling (SQL and EDBMS, NoSQL)
- Modern data infrastructure: Data registries, Data Factories, Semantic storage, SQL/NoSQL



FAIR Curriculum Design: Notes on important topics from practitioners

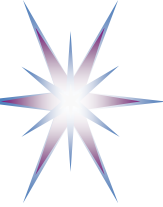
– **Extracted from forums and discussions (May 2020)**

- Linked data, provenance
 - Research data and scientific publications
- Ontology and semantic data interoperability
 - Ontology engineering
- Multi-disciplinary data semantics and Interoperability
 - How to make semantics FAIR
 - Open Knowledge Graph?
- FAIR for different types of data: Observational data, experimental, social, life science
- FAIR metrics and Maturity Framework
- PID for data
- Use of standards
- LERU Roadmap and LEARN project: RDM at universities



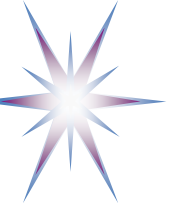
Methodology: EDISON Data Science Framework (EDSF) - Overview

- EDISON Data Science Framework (EDSF) and community based maintenance
- Data Science Programmes at UvA
- Data Management and Governance curriculum design
 - EDSF Data Management Body of Knowledge (DSDM-BoK)
 - DAMA Body of Knowledge (DAMA BoK)
 - Data Stewardship – Existing Frameworks
 - Research Data Management (RDM) best practices
- Methodology: Evidence based and Community driven
 - Job market analysis
 - Existing initiatives and frameworks

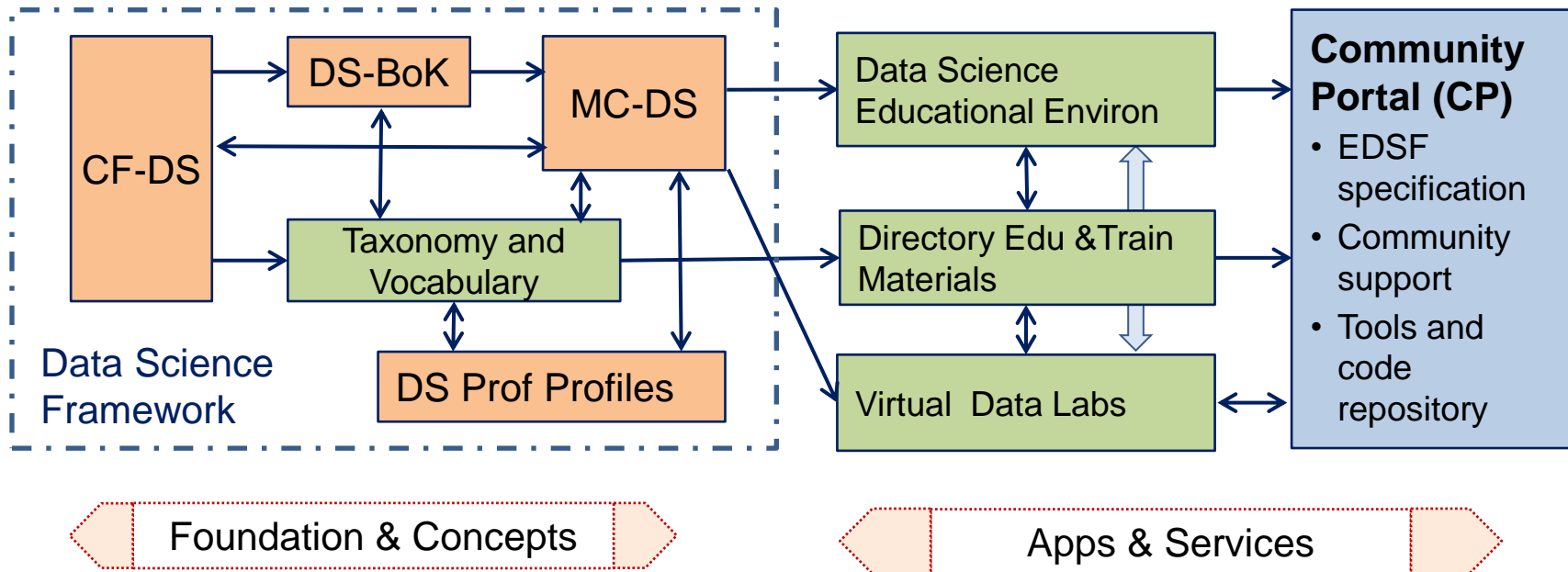


EDISON Project (2015-2017) and Follow on

- EDISON project website - <http://edison-project.net/>
(old domain *edison-project.eu* expired)
- EDISON Data Science Framework (EDSF) – main outcome of the project
- Currently maintained by EDISON Community Initiative, coordinated by UvA
- EDSF Release 3 published in 2018
- EDSF Release 4 Design Workshop – 20 Nov 2019, UvA
 - EDSF Release 4 (EDSF2020) to be published by the end of 2020 (initially planned January 2020)



EDISON Data Science Framework (EDSF) – Core components and community maintained services



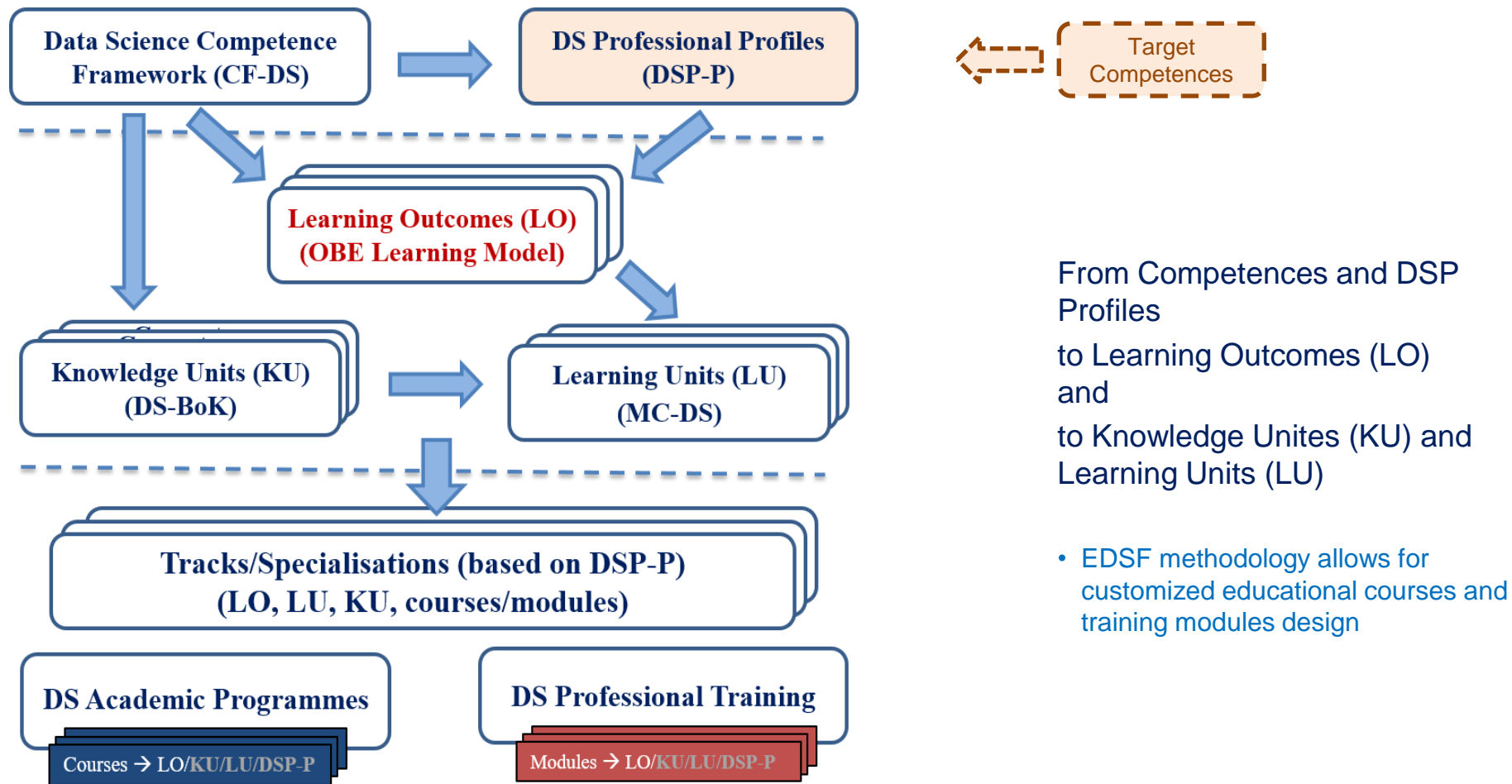
EDISON Framework core components and documents

- 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)
- Data Science Taxonomies and Scientific Disciplines Classification

Applications and Services

- Virtual Data Science Labs
- Data Science Educational Environment
- Directory of edu & train resources
- Community Portal – currently github

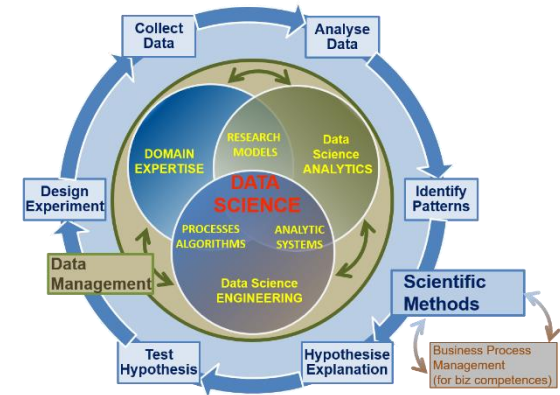
Outcome Based Educations and Training Model: Addressing target competences for the profession

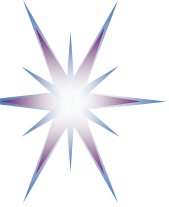


Data Science Body of Knowledge (DS-BoK)

DS-BoK Knowledge Area Groups (KAG)

- KAG1-DSE: Data Analytics group including Machine Learning, statistical methods, and Business Analytics
- KAG2-DSE: Data Science Engineering group including Software and infrastructure engineering
- KAG3-DSDM: *Data Management group including data curation, preservation and data infrastructure*
- KAG4-DSRM: *Research Methods and Project Management group*
- KAG5-DSBA: Business Analytics and Business Intelligence
- **KAG* - DSDK**: *Data Science domain knowledge to be defined by related expert groups*





KAG3-DSDM: *Data Management group: data curation, preservation and data infrastructure*

DM-BoK version 2 “Guide for performing data management”

– 11 Knowledge Areas

(1) Data Governance

(2) Data Architecture

(3) Data Modelling and Design

(4) Data Storage and Operations

(5) *Data Security*

(5a) *Data compliance, Data Privacy, GDPR*

(6) Data Integration and Interoperability

(7) *Documents and Content*

(8) Reference and Master Data

(9) Data Warehousing and Business Intelligence

(10) *Metadata*

(11) Data Quality

Other Knowledge Areas motivated by *RDA, European Open Data initiatives, European Open Data Cloud*

(12) *PID, linked data, data registries*

(13) *Data Management Plan*

(14) *Open Science, Open Data, Open Access, ORCID*

(15) *Responsible data use, Ethics*

(16)* *Data Sovereignty (and Indigenous data protection)*

- Highlighted in red: Considered (Research) Data Management literacy (minimum required knowledge)

Data Governance and Stewardship

Definition: The exercise of authority, control, and shared decision-making (planning, monitoring, and enforcement) over the management of data assets.

Goals:

1. Enable an organization to manage its data as an asset.
2. Define, approve, communicate, and implement principles, policies, procedures, metrics, tools, and responsibilities for data management.
3. Monitor and guide policy compliance, data usage, and management activities.

Business
Drivers

Inputs:

- Business Strategies & Goals
- IT Strategies & Goals
- Data Management and Data Strategies
- Organization Policies & Standards
- Business Culture Assessment
- Data Maturity Assessment
- IT Practices
- Regulatory Requirements

Activities:

1. Define Data Governance for the Organization (P)
 1. Develop Data Governance Strategy
 2. Perform Readiness Assessment
 3. Perform Discovery and Business Alignment
 4. Develop Organizational Touchpoints
2. Define the Data Governance Strategy (P)
 1. Define the Data Governance Operating Framework
 2. Develop Goals, Principles, and Policies
 3. Underwrite Data Management Projects
 4. Engage Change Management
 5. Engage in Issue Management
 6. Assess Regulatory Compliance Requirements
3. Implement Data Governance (O)
 1. Sponsor Data Standards and Procedures
 2. Develop a Business Glossary
 3. Co-ordinate with Architecture Groups
 4. Sponsor Data Asset Valuation
4. Embed Data Governance (C,O)

Deliverables:

- Data Governance Strategy
- Data Strategy
- Business / Data Governance Strategy Roadmap
- Data Principles, Data Governance Policies, Processes
- Operating Framework
- Roadmap and Implementation Strategy
- Operations Plan
- Business Glossary
- Data Governance Scorecard
- Data Governance Website
- Communications Plan
- Recognized Data Value
- Maturing Data Management Practices

Suppliers:

- Business Executives
- Data Stewards
- Data Owners
- Subject Matter Experts
- Maturity Assessors
- Regulators
- Enterprise Architects

Participants:

- Steering Committees
- CIO
- CDO / Chief Data Stewards
- Executive Data Stewards
- Coordinating Data Stewards
- Business Data Stewards
- Data Governance Bodies
- Compliance Team
- DM Executives
- Change Managers
- Enterprise Data Architects
- Project Management Office
- Governance Bodies
- Audit
- Data Professionals

Consumers:

- Data Governance Bodies
- Project Managers
- Compliance Team
- DM Communities of Interest
- DM Team
- Business Management
- Architecture Groups
- Partner Organizations

Technical
Drivers

Techniques:

- Concise Messaging
- Contact List
- Logo

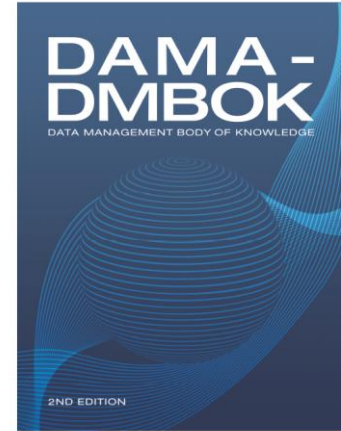
Tools:

- Websites
- Business Glossary Tools
- Workflow Tools
- Document Management Tools
- Data Governance Scorecards

Metrics:

- Compliance to regulatory and internal data policies.
- Value
- Effectiveness
- Sustainability

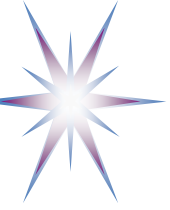
DMBOK: Data Governance and Stewardship



Technics Publications
BASKING RIDGE, NEW JERSEY

Scope of a Data Governance Programme

- Strategy
- Policy
- Standards and quality
- Oversight
- Compliance
- Issue management
- Data management projects
- Data asset valuation



DMBOK: Data Management Principles

DATA MANAGEMENT PRINCIPLES

Effective data management requires leadership commitment

Data is valuable

- Data is an asset with unique properties
- The value of data can and should be expressed in economic terms

Data Management Requirements are Business Requirements

- Managing data means managing the quality of data
- It takes Metadata to manage data
- It takes planning to manage data
- Data management requirements must drive Information Technology decisions


Data Management depends on diverse skills

- Data management is cross-functional
- Data management requires an enterprise perspective
- Data management must account for a range of perspectives

Data Management is lifecycle management

- Different types of data have different lifecycle characteristics
- Managing data includes managing the risks associated with data

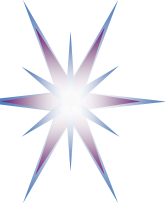
- Data is an asset with unique properties
- The value of data can and should be expressed in economic terms
- Managing data means managing the quality of data
- It takes Metadata to manage data
- It takes planning to manage data
- Data management requirements must drive Information Technology decisions
- Data management is cross-functional; it requires a range of skills and expertise
- Data management requires an enterprise perspective
- Data management must account for a range of perspectives
- Data management is lifecycle management
- Different types of data have different lifecycle characteristics
- Managing data includes managing the risks associated with data
- Effective data management requires leadership commitment



Data Stewardship (according to DM-BOK)

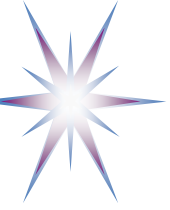
- **Creating and managing core Metadata:** Definition and management of business terminology, valid data values, and other critical Metadata.
- **Documenting rules and standards:** Definition/documentation of business rules, data standards, and data quality rules.
 - High quality data are often formulated in terms of rules rooted in the business processes that create or consume data.
 - Stewards help surface these rules and ensure their consistent use.
- **Managing data quality issues:** Stewards are often involved with the identification and resolution of data related issues or in facilitating the process of resolution.
- **Executing operational data governance activities:** Stewards are responsible for ensuring that, day-today and project-by-project, data governance policies and initiatives are adhered to. They should influence decisions to ensure that data is managed in ways that support the overall goals of the organization.

“Best Data Steward is not made but found” DMBOK1 (2009)

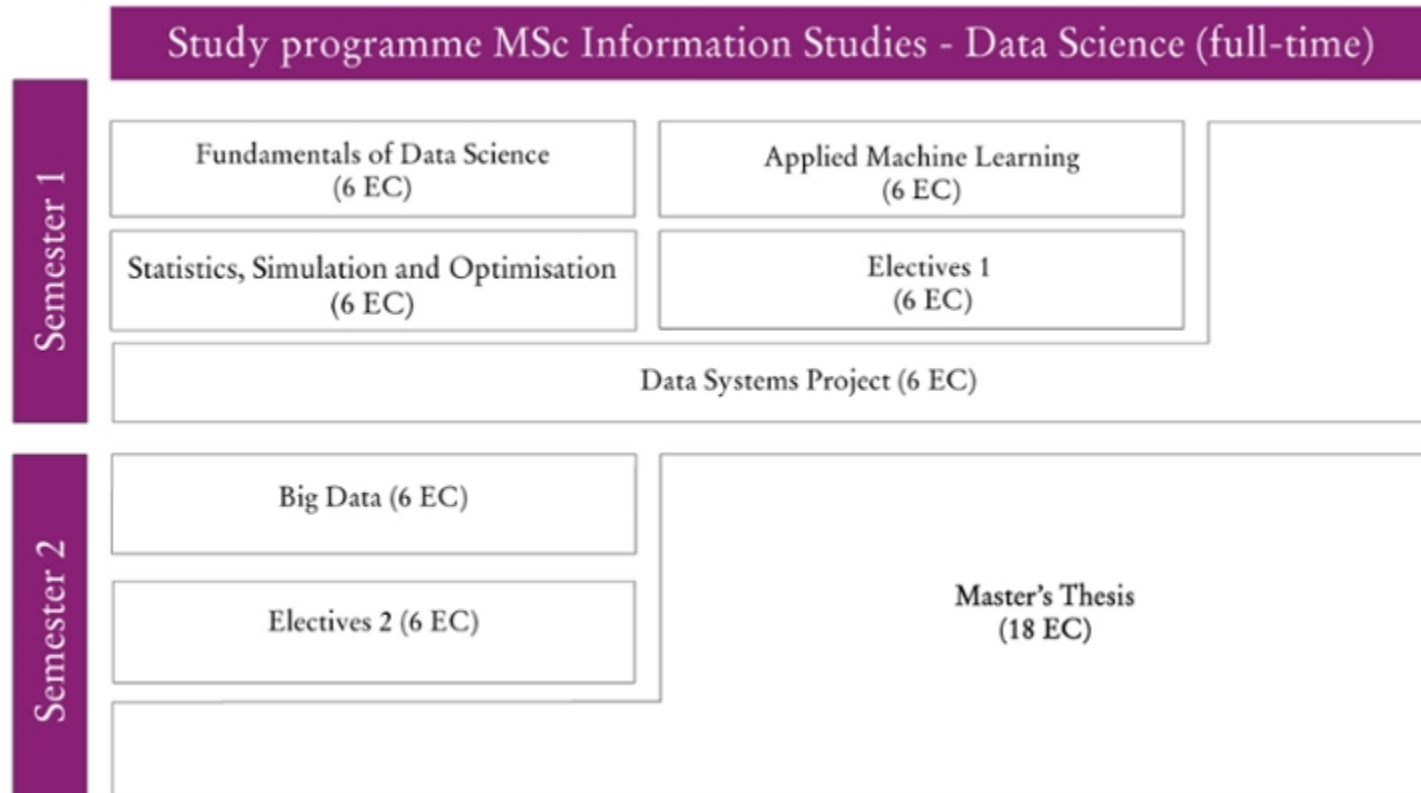


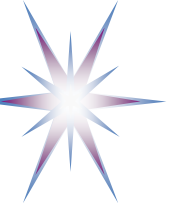
Case Study: Data Science related Master Programmes at UvA

- Big Data Engineering (joint UvA and VU) - since 2016
- Data Science at AI department (stopped 2018)
- **Data Science at the Institute of Interdisciplinary Studies – since 2017**
 - <https://www.uva.nl/en/programmes/masters/information-studies-data-science/data-science.html>
- Data Science in Business Administration (MBA) at Business School Amsterdam – part time, evening, since 2017
- **Data Science in Business Analytics (MBA) at Business School Amsterdam, full time starting 2020**



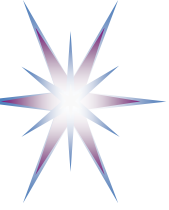
Study Programme MSc IS – Data Science (1yr)



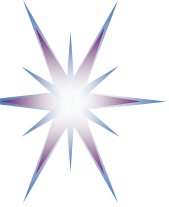


Case study: BKO program for university teachers in Netherlands

- Rich program of basic and advanced instructional and didactic methodologies
- Still lacks of Research Data Management (RDM) courses
 - Currently in contact with the BKO program coordinator at UvA

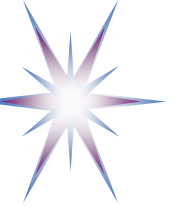


Discussion - TBD



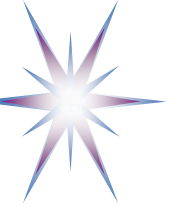
Additional information

- FAIRsFAIR WP7 Task T7.3
- Outcome and recommendations from FAIRsFAIR seminar 8 May 2020

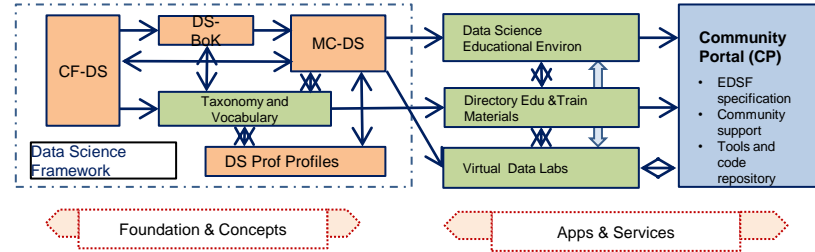


Task 7.3: FAIR data competence framework

- Objectives:
 - Develop a FAIR data competence framework complementary to or as extension to existing and adopted data science and other competence frameworks (e.g. the ESCO-compliant EDISON Data Science Framework).
 - The activities will focus on data science but also address other disciplines, where identified as relevant, to enhance the impact and **sustainability of the project in fostering a FAIR data culture** throughout different scientific disciplines, research communities and professions.
- The FAIR data competence framework will include
 - FAIR data competences which can be acquired through higher education, e.g. in data science programmes in other data-intensive and data driven disciplines
 - FAIR data competences for graduates continuing to work as professionals in FAIR data management (e.g. data stewards, research infrastructure managers).
- Reporting
 - D7.3 FAIR Competence Framework for Higher Education [M24; UvA]
 - M7.5 First Draft of competence framework [M20] [first draft finalised]



T7.3 Actions plan and status



1. Analyse existing FAIR data, Data Management and Data Stewardship framework

- 1.1. Inventory of identified and proposed competences and practices, academic programmes
- 1.2. Analyse available training and academic programme on Data Management, Stewardship, FAIR
- 1.3. Revisit EUA Survey on DM/FAIR at universities (D7.1) – [PP02](#)

Timeframe: January - June 2020

2. FAIR data competence framework development

- 2.1. Use EDISON methodology to create FAIR competence framework (including competences, skills and knowledge topics definitions)
- 2.2. Community discussion: forum, feedback form, survey

[PP01: Liaise with T7.2/D7.2 \(August 2020\)](#)

[AW/PP: EOSC Training Group by Natalia Manola](#)

May - October 2020

3. Design Workshop to discuss the proposed Competence Framework

- 3.1. Workshop preparation meeting during FAIRsFAIR project meeting in April 2020
- 3.2. Design workshop

Goal: to discuss the proposed Competence Framework, review and validate

October 2020 in Amsterdam

- 3.3. Community workshops and/or tutorials

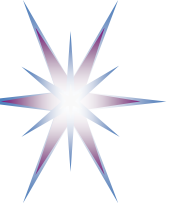
[PP03: Stakeholder Workshop](#)

October - November 2020



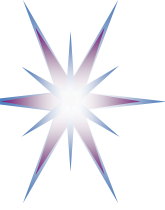
Overview existing Projects and Initiatives

- Open Science Skills and Digital Literacy in Danish Research Libraries
<https://zenodo.org/communities/os-skills-dk>
 - “This project was funded by Denmark’s Electronic Research Library (DEFF) with the aim to investigate, model, and deploy Open Science skills for Danish Research Librarians.
- The project provides an overview of Danish digital collections, with a focus on openness and applicability.
 - 1) access and harvest data
 - 2) identify and guide in tools to build data corpus and
 - 3) analyze data, were considered.



Courses related to RDM: MANTRA

- MANTRA is a free online course for those who manage digital data as part of their research project (Univ of Edinburgh)
<https://mantra.edina.ac.uk/>
 - Research data in context
 - Data management plans
 - Organising data
 - File formats & transformation
 - Documentation, metadata, citation
 - Storage & security
 - Data protection, rights & access
 - Sharing, preservation & licensing
 - Data handling tutorials
- Also offered via EDSA - <https://courses.edsa-project.eu/mod/checklist/view.php?id=505>



Used materials and teaching models

The tutorial used a number of materials produced by EU and national funded projects which are released as Open Access and/or under Creative Common Attribution 4.0 International License

- Data Management guidelines and template are provided by Mariëtte van Selm from University of Amsterdam Library

The following materials used (April-May 2016)

- Directory of efforts and resource related to research data management education by RDA/UA and Amy Nurnberger
https://docs.google.com/spreadsheets/d/10RTW-nZk0x_mpQw2VAltcc656MV9EeCaDe2IM4umb4/edit#gid=0
- FOSTER Project: Open Science training materials
<https://www.fosteropenscience.eu/resources>
- UK Data Service training materials <https://www.ukdataservice.ac.uk/use-data/advice>
- Digital Curation Center, UK: Training materials: Support DM, DC 101,
<http://www.dcc.ac.uk/training/training-and-reference-materials>
- RDMRose Learning Materials http://rdmrose.group.shef.ac.uk/?page_id=10
- TraD project supportDM course <https://www.uel.ac.uk/trad/outputs/resources/>
- Data Management Short Course for Scientists by ESIP Federation
<http://commons.esipfed.org/datamanagementshortcourse>



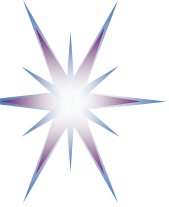
EDSF and EDISON Community Initiative

- 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
- EDISON project website - <http://edison-project.net/>
(old domain *edison-project.eu* expired)



Course: Data Management and Governance (DMG) in Enterprise

- Introduction. Big Data Infrastructure and Data Management and Governance.
- Data Management concepts. Data management frameworks: DAMA Data Management framework, the Amsterdam Information Model. Extensions for Big Data and Data Science.
- Enterprise Data Architecture. Data Lifecycle Management and Service Delivery Model. Data management and data governance activities and roles.
- Data Science Professional profiles and organisational roles, Skills management and capacity building.
- Data Architecture, Data Modelling and Design. Data types and data models. Data modeling. Metadata. SQL and NoSQL databases overview. Distributed systems: CAP theorem, ACID and BASE properties.
- Enterprise Big Data infrastructure and integration with enterprise IT infrastructure. Data Warehouses. Distributed file systems and data storage.
- Big Data storage and platforms. Cloud based data storage services: data object storage, data blob storage, Data Lakes (services by AWS, Azure, GCP).
 - Trusted storage, blockchain enabled data provenance.
- FAIR data principles and Data Stewardship, Data Quality assessment and maturity model. Data repositories, Open Data services, public services.
- Big Data Security and Compliance. Data Privacy and GDPR. Data security and data protection. Security of outsourced data storage. Cloud security and compliance standards and cloud provider services assessment.



Course: Research Data Management and Stewardship (RDMS) (1)

A. Use cases for data management and stewardship

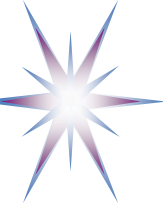
- Preserving the Scientific Record
- Data Lifecycle and Provenance

B. Data Management elements (organisational and individual)

- Goals and motivation for managing your data
- Data formats, Metadata, related standards
- Creating documentation and metadata, metadata for discovery
- Using data portals and metadata registries
- Tracking Data Usage, data provenance, linked data
- Handling sensitive data
- Backing up data, backup tools and services
- Data Management Plan (DMP)

C. Responsible Data Use (Citation, Copyright, Data Restrictions)

- Data privacy and GDPR compliance
- Ethical issues



Course: Research Data Management and Stewardship (RDMS) (2)

D. FAIR principles in Research Data Management, supporting tools, maturity model and compliance

E. Data Stewardship and organisational data management

- Responsibilities and competences
- DMP management and data quality assurance

F. Open Science and Open Data (Definition, Standards, Open Data use and reuse, open government data)

- Research data and open access
- Repository and self- archiving services
- RDA products and recommendations: PID, data types, data type registries, others
- ORCID identifier for data and authors
- Stakeholders and roles: engineer, librarian, researcher
- Open Data services: ORCID.org, Altmetric Doughnut, Zenodo

G. Hands on practice topics: DMP, Metadata, Data Formats, Data publishing, etc