Developing a Data steward curriculum: the Danish experience

Lorna Wildgaard, PhD

Copenhagen University Library,

Research Support: lowi@kb.dk

Haakon Lund

University of Copenhagen, Department of Communication hl@hum.ku.dk

UNIVERSITY OF COPENHAGEN





1: Review of existing DS educations & certifications



2: Job profiles of DS in Denmark



3: Analysis of job vacancies





Collaboration with academia & industry = relevant educations

> Course components should support eward roles

oft skills need to be taught alongside

Standardized, well defined concepts

Experiential pedagogies & best practices





THE ADMINISTRATOR

- · Team player
- · Can-do attitude · Implement solutions and educating
- end-users about them · Focus on execution in strategic
- · Structured and analytical mindset
- · Passion for policy and IT security
- · Establish good practices in compliance & data privacy
- · Risk assessments Fast learner



THE ANALYST

- Programmer
- · Statistical & data analyst
- · Fast learner and innovative on building custom software and databases
- · Ensure data quality · Enthusiasm in cloud
- · Seek challenges, have positive attitude towards
- business intelligence Innovative · Develop procedures &



· Process optimization via

FAIR principles advisor

Data planner

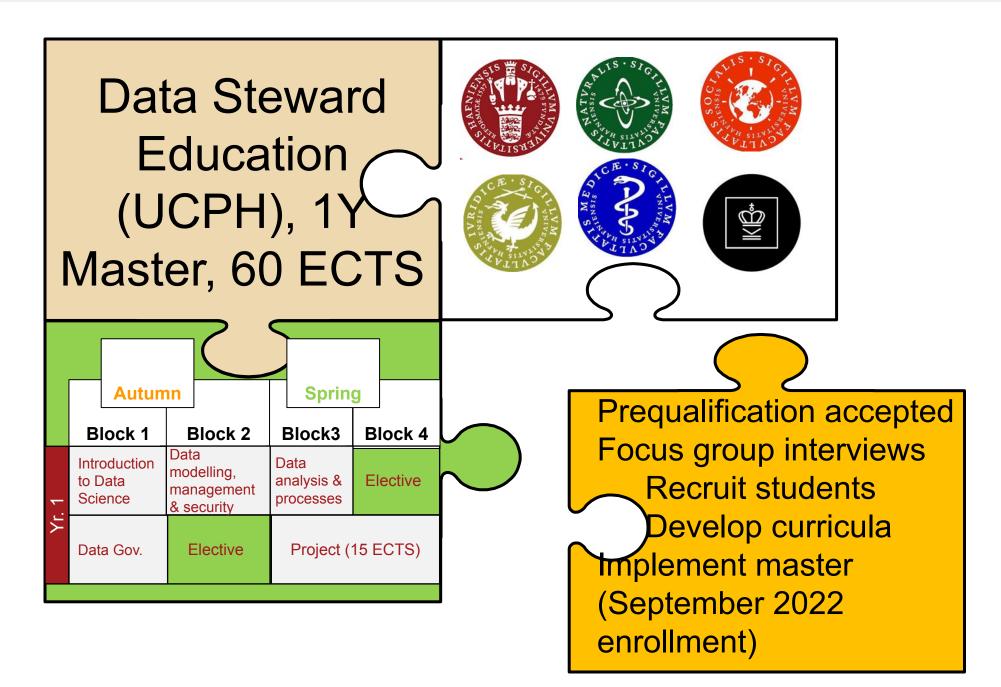
good project management

· Focus on collaboration and knowledge sharing to raise

THE AGENT OF CHANGE

- · Client & customer oriented
- · Mediate processes &
- · Passionate to implement solutions via project & change
- User empathy
- · Focus on execution of policy and strategy awareness
- Agile mindset
- Enthusiastic





Process

- Developed by a cross-university work group
- Document analysis (background), Informal hearing with Danish E-Infrastructure Consortium, dialog with advisory boards (major players in the private and public sector)
- Major input from report "National Coordination of Data Steward Education in Denmark"
- The proposed curriculum is now beeing evaluated by the Ministry of Higher Education and Science expected approval by the end November 2021

Relevance

- Increase the number of people with specialized skills in IT and data -> there is currently a lack of data scientist, ML, IA, computer scientists, software developers
- The increased societal focus on digitalization demands knowledge about how to manged and activate stored data -> support for business and process development
- Data stewardship is currently undertaken by professionals from related disciplines i.e. data scientist to law
- Data stewardship combine digital expertise with domain specific knowledge



Number of data stewards needed

- UCPH estimates a need of at least 50 masters in data stewardship yearly reports have estimated a need for at least 20.000 skilled staff before 2030
- Advisory board members representing 5 private and public sector organiszations (medical industry, industrial production and diary sector) estimated a need for 150-200 data stewards during the next 5 years alone
- 115 questionnaire respondents from private and public sector did express a need for data stewards in their organizations
- We estimate a yearly need for 100 200 with a master in data stewardship



Structure

	Autumn		Spring	
	Bloc 1	Bloc 2	Bloc 3	Bloc 4
Year 1	Python for data science	Data modelling, management and security	Data analysis & processes	Elective course : methodologies
	Data governance	Elective course : Domaine specific	Project (15 ECTS)	

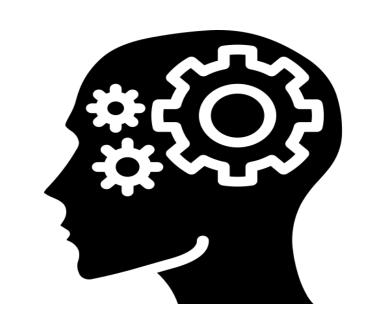
Elective course :
methodologies
Machine learning & AI
Open data & FAIR
Project management / change management
Citizen science

Background reports – needs analysis

- Background reports:
 - Iris Group and HBS Economics, 'Mismatch in the Danish labor market for IT graduates in 2030', HBS Economics, Jun. 2021. https://hbseconomics.com/publications/mismatch-inthe-danish-labor-market-for-it-graduates -in-2030/
 - McKinsey & Company and Innovationsfonden, 'An Al Nation? Harnessing the opportunity of artificial intelligence in Denmark', 2019. [Online]. Available: https://innovationsfonden.dk/sites/default/files/2019-09/an-ai-nation-harnessing-theopportun ity-of-ai-in-denmark.pdf
 - L. Wildgaard, E. Vlachos, L. Nondal, A. V. Larsen, and M. Svendsen, 'National Coordination of Data Steward Education in Denmark: Final report to the National Forum for Research Data Management (DM Forum)', Zenodo, Jan. 2020. doi:10.5281/zenodo.3609516.
 - + a number of reports in danish

Reference Slides





Coordination of Data Stewardship Education in Denmark

April 16 - 2020 | NPOS, Project F | NL Michael Svendsen, Head of Research Support, Copenhagen University Library, Royal Danish Library

@tullemich

https:/bit.ly/2ToEnga



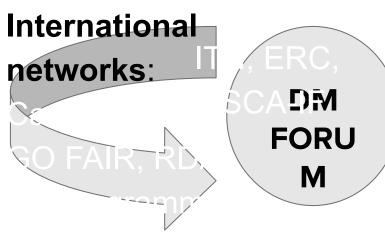
- The DS project in brief: commissioned study
- Outcome: level of demand and key compentencies
- Prequalification work: new master programme in Data Stewardship
- Future focus: Network and training activities





A National Data Management Forum (DeiC) funded project

Researchers:



Univ, public & private sector: sities in



Educational programmes::



Partcipating Institution

AAU - Aalborg University

AUL/KB - Aarhus University Library

CBS - Copenhagen Business School

ITU - IT University (observer)

KP - University College Copnehagen (observer)

KUB - Copehagen University Library (project lead)

SA - The National Archives

RUC - Roskilde University

SDU - University of Southern Denmark

Project period (2019):



- 1. Is a national Data Steward education a feasible solution?
- 2. Is there a rationale in terms of benefits for a national coordinated effort?
- 3. Should the development of Data
 Steward education in Denmark be in
 collaboration or competition between
 universities and businesses?
- 4. How to ensure recommendations in the study address the demand side both in the private and public sector?

DATA STEWARDS

competition or coordination?





Methods & Data



1: Review of existing DS educations & certifications





2: Job profiles of DS in Denmark



3: Analysis of job vacancies

4: National survey & interviews

GROUP 1, the Administrator			
Job Vacancies	LinkedIn Hard Skills	LinkedIn Soft Skills	
business	data	project	
management	system	communication	
global	sap	innovative	
team	business	team	
develop	customer	professional	
information	marketing	solutions	
organization	data steward	positive	
reporting	user	player	
system	sales	business	
security	product	colleagues	
support	analysis	performance	
solutions	implementation	importance	
implementation	reporting	analytical	
compliance	data quality	developing	
relevant	database	making	
customer	sql	great	
risk	architecture	extra	
requirements	information	data	
financial	management	execution	
		•	

Topic modeling of text corpus from the LinkedIn and job vacancies vocabulary

as

ce

→ to discover hidden thematic structures of our text corpora



G	ROUP 1, the Administra	ator
Job Vacancies	LinkedIn Hard Skills	LinkedIn Soft Skills
business	data	project
management	system	communication
global	sap	innovative
team	business	team
develop	customer	professional
information	marketing	solutions
organization	data steward	positive
reporting	user	player
system	sales	business
security	product	colleagues
support	analysis	performance
solutions	implementation	importance
implementation	reporting	analytical
compliance	data quality	developing
relevant	database	making
customer	sql	great
risk	architecture	extra
requirements	information	data
financial	management	execution

as ce n d n g 0



GROUP 2, the Analyst			
Job Vacancies	LinkedIn Hard Skills	LinkedIn Soft Skills	
data	processes	communication	
data quality	sap	management	
business	management	training	
team	customer	customer	
support	develop	structured	
customer	business	tool	
data steward	marketing	important	
description	masterdata	collaboration	
system	product	outgoing	
knowledge	reporting	solutions	
international	data quality	professional	
monitoring	user	simple	
analytical	maintenance	various	
reporting	governance	tasks	
ensuring	architecture	focus	
projects	change	understanding	
finance	ensuring	procedures	
tools	reports	extra	
activities	tableau	user	
compliance	microsoft	innovative	

as ce n d n g 0



GROUP 3, the Developer			
Job Vacancies	LinkedIn Hard Skills	LinkedIn Soft Skills	
data	data	project	
business	develop	team	
data quality	system	tasks	
team	masterdata	management	
customer	management	important	
data steward	super (user)	close	
requirements	processes	structured	
quality	customer	training	
management	data steward	player	
support	sales	procedures	
information	support	life	
system	project	time	
standards	manager	performance	
develop	product	human	
data_management	implementation	challenge	
processes	analysis	games	
issues	global	create	
knowledge	data management	seek	
analysis	data quality	knowledge	

as ce n d n g 0



Method:

August 2019, survey validated & sent to universities, libraries, companies and NGOs.

Snowballing approach.

Software: SurveyXact

Survey closed: September 19th

Respondents:

329 opened survey

62 partial answers

84 complete answers

86% from public organisations

14% private organisations

56% from employed at universities

SURVEY

for developing a Data Stewardship education in Denmark





Employed VS. non-employed at University

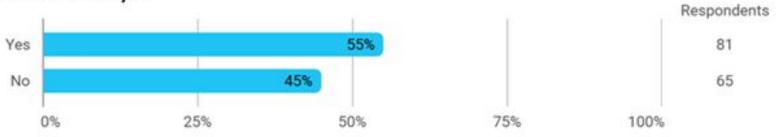


for developing a Data Stewardship education in Denmark





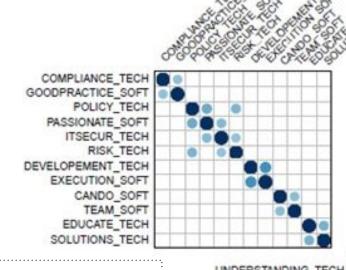
Are you employed at a university?





Not employed at a University

3 confirmed roles + 1 new identified to survey questionnarie and interviews



Agent of Change?

Client oriented and

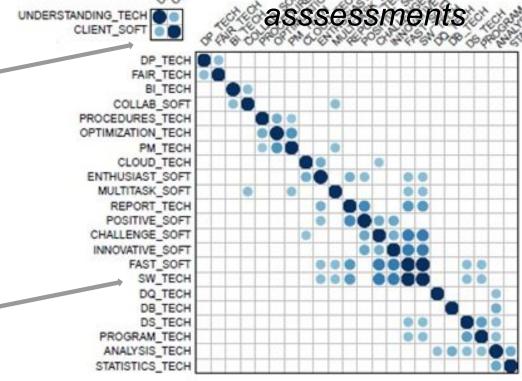
Combination of Analyst and of processes

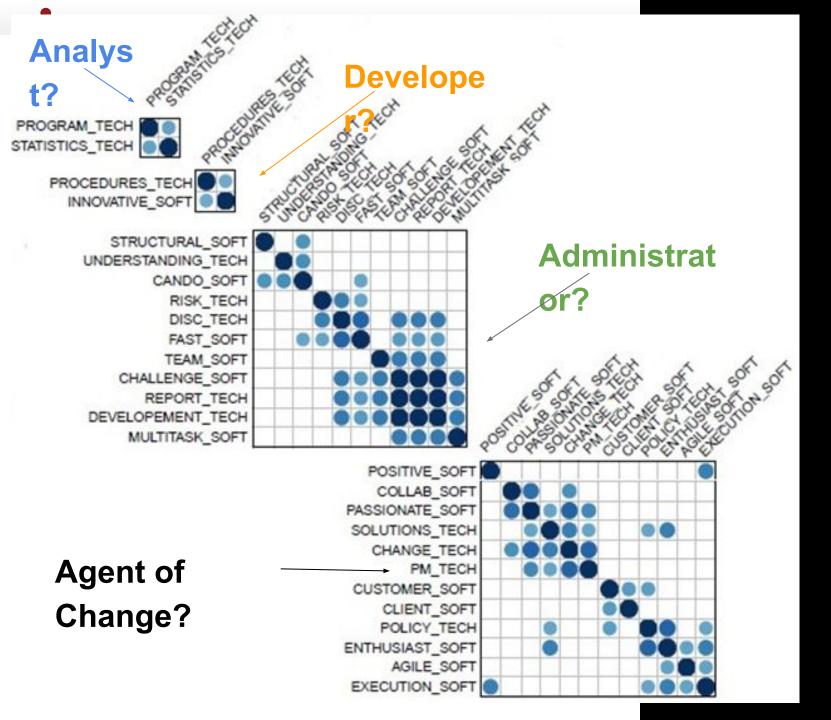
Developer

and operations Fast learner, innovative who Builds custom SW, seeks Challenges with

Administrator:

Execution of strategic development, passionate about policy w. strategy awareness & risk





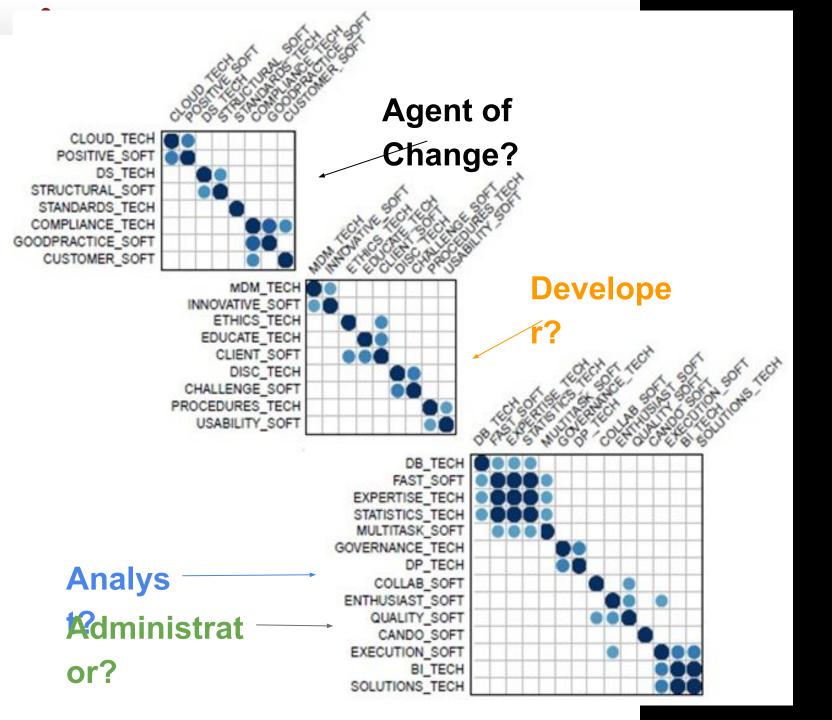
Employed at a University

0 standard deviation (no correlation):

Expertise and Business Intelligence

NAT-SCI TECH HEALTH





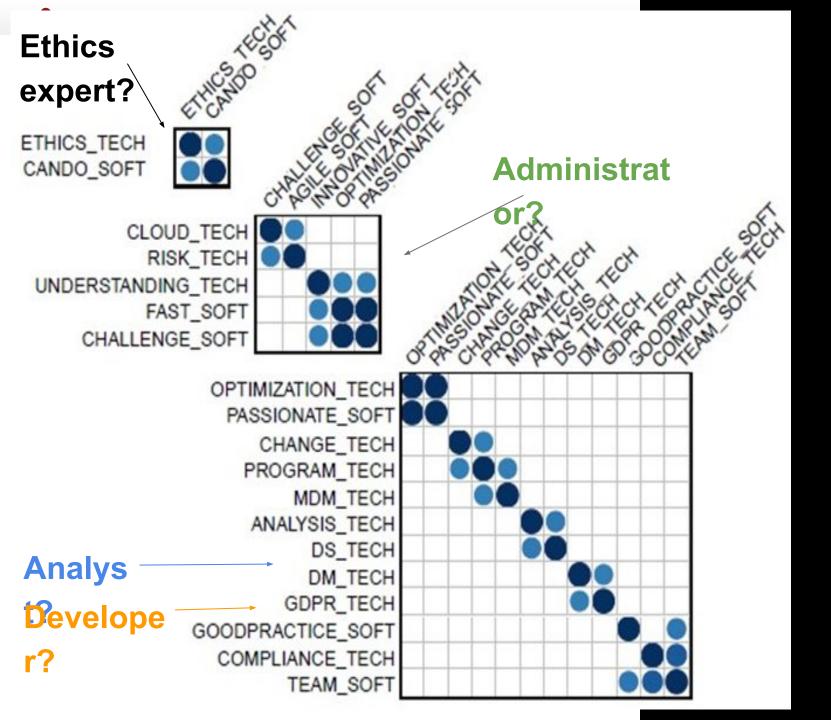
Employed at a University

0 standard deviation (no correlation):

Project Management,
Strategic Development,
Optimization of Processes,
Change Management,
SW development

HUMANITIES SOC-SCI





Employed at a University

0 standard deviation (no correlation):

Focus on Execution, Project
Management, Strategic
Development,
Statistics, Multitasking,
Enthusiasm,
Positive Attitude, Reporting,
Implementing Solutions SW
development

4 competency groups identified in job pro LinkedIn (hard+soft skills) + job vacancies combined the administrator, focused on business

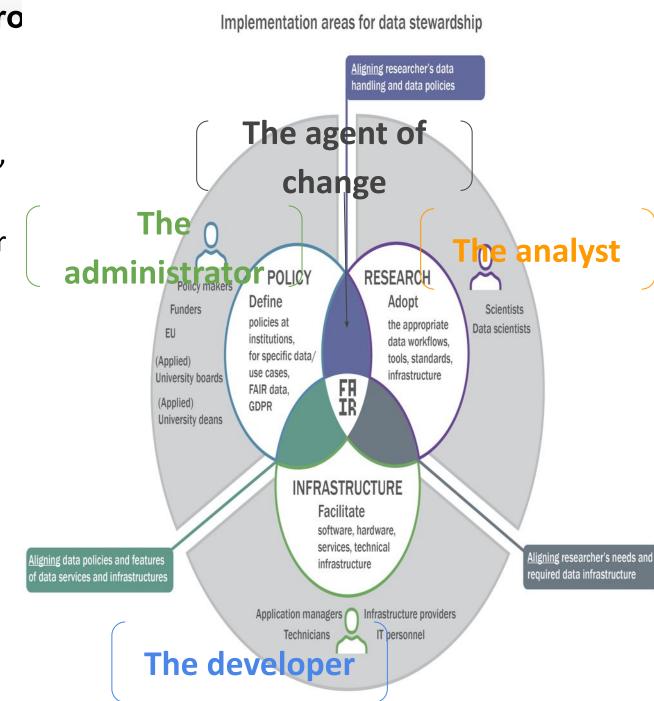
intelligence solutions and data-service support, providing management with KPI's

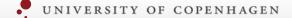
The analyst, focused on data quality, customer

The analyst, focused on data quality, customer support and crunching data for the industry or doing research for the university.

The developer, project oriented, emphasizing on data management, procedures and testing with technical and analytical skills

The agent of change, client, customer oriented, process & operations mediator, implementation of solutions via project + change management, execution of policy and strategy





The 4 roles of a Data Steward

The 4 roles of a Data Steward



THE ADMINISTRATOR

- Team player
- Can-do attitude
- Implement solutions and educating end-users about them
- Focus on execution in strategic development
- Structured and analytical mindset
- · Passion for policy and IT security
- Establish good practices in compliance & data privacy
- Risk assessments
- Fast learner



THE ANALYST

- Programmer
- Statistical & data analyst
- Fast learner and innovative on building custom software and databases
- Ensure data quality
- Enthusiasm in cloud solutions
- Seek challenges, have positive attitude towards reporting



THE DEVELOPER

- Process optimization via good project management
- FAIR principles advisor
- Data planner
- Focus on collaboration and knowledge sharing to raise business intelligence
- Innovative
- Develop procedures & guidelines



THE AGENT OF CHANGE

- Client & customer oriented
- Mediate processes & operations
- Passionate to implement solutions via project & change management
- User empathy
- Focus on execution of policy and strategy awareness
- Agile mindset
- Enthusiastic

47% know term Data Steward 52% unfamiliar

"so many of the terms are still completely interchangeable or incomprehensible to me.

I actually think this might be your biggest challenge to get this project flying....Finding a common language."

"We use the term 'Data Steward' to refer to the lead of University academics responsible for generating the research OM FORURI, Project Lead, First Supervisor in the case of PhD students."

SO, WHAT IS A DATA

STI

Preferred term	n
Don't know	37
Data manager	29
DPO	13
Data Steward	10
Data librarian	6
Information Management Specialist	4
Arkivar / Digitaliseringskonsulent	3
Data scientist	2
Dataansvarlig	2
IT specialists	2
Records manager	2
Consultant	2
Colleague	2
Administrative medarbejdere - IT	2
Project managers/ dept. coordinators	2
Business intelligence analyst	1
Data Engineers	1
Data handler	1
Data koordinator (coordinator)	1
Data specialists	1
Data/system owners	1
Datadisciple	1
IT Risk Management	1
IT-koordinator	1
Systemansvarlig or digitaliseringskonsulent	1
Projektmedarbejder	1
Research Data Officer	1
Research supporter	1

Most frequent DS concepts and skills

University

Don't know (28%)

Data Manager (25%)

Data Steward (10%)

Not a position, teamwork (10%)

Data Protection Officer (6%)

Not University

DPO (20%)

Don't know (15%)

Data Manager (14%)

IT, Administrive (13%))

Librarian/information spec (10%)

Tech skills: DM, GDPR, FAIR

Soft skills: Usability, user oriented,

good practices

Tech skills: DM, GDPR, Compliance & privacy

Soft skills: Good practices, Quality, Customer/user support

DIVI FURUIV

Danish National Forum far

Research Data Management

DATA STEWARDS competition or coordination?

- 1. <u>Collaboration between academia and industry</u> result in relevant educations and improve the student's employability.
- 2. Course design should support the competencies of a DS as an:

Administrator, Analyst and/or Developer, Agent of Change?

- 3. Data Stewards work in a multitude of positions, therefore:
 - skills and competencies <u>must not be limited to</u> <u>technical</u> proficiency and hard skills alone
 - must be supported with "soft skills" like the ability to assess:

the value of DS to communicate, teach, have disciplinary knowledge and the ability to teamwork and project manage are requirements for curricula development.





4. Education & competence development support 4 DS roles:

Administrator, Analyst, Developer, Agent of Change

including the Dutch recommendations highlighting focus areas of: **Policy**, **research & infrastructure** with functions of Data Stewards as: **Embedded**, **Generic**, **Policy**

- 5. Clearly defined concepts
- 6. Develop experiential pedagogies: New and innovative educational structures in collaboration with local/national stakeholders
- 7. Align a common national approach to University education
- 8. Form a national Data steward network in order to coordinate DS training & knowledge sharing





Recommendations for DS education pregalification

The 3 models for Data Stewardship Education

Students with BACHELOR degree

Directed corporate employment

- One-year master programme
 For students who fulfill requirements to basic programming skills, study skills, subject knowledge and academic language level.
- Two year master programme
 For students who do not meet the requirement for programming and study skills follow a premaster's year before the master's and then continue with the one-year master.
- Two-year candidatus
 Traditional university candidatus combining theory, methods and internships. A

 dissertation in the area of DS is required.

Students with a PH.D. or equivalent

Por PhDs from any field at university faculties or part of research teams in industry. Educational programmes are a collaborative endeavor between faculties, library or knowledge centers, Centers for Information Security, Data Steward Community, Research coordinators, Project Pl's, System Developers, Communication and Teaching teams, the Graduate School (for PhD training) & Human Research Ethics Committees.

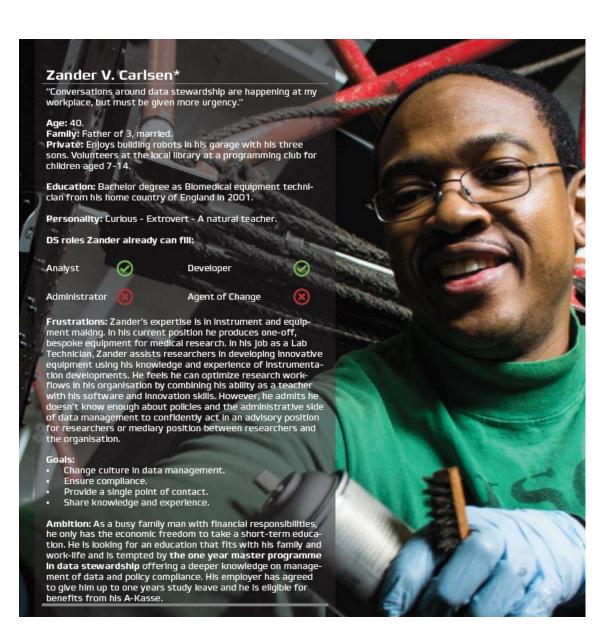
 Requirements: Short courses, workshops, mentorships, online modules, summer school programmes & MOOCs, that immerse the student in the DS Community.

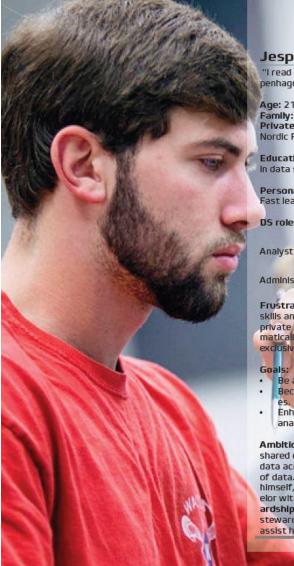
VET-students (Vocational Education & Training)

Directed corporate or academic employment
Flexible master programme

- For students who have professional experience and wish to improve their DS skills but keep working full-time, or already have a master's degree in any field.
- The programme is a part-time vocational education with a prescribed period of study, that the student individually plans, and a total of 60 ECTS-points.
- The student combines elements from established programmes, thus it is a requirement that DS courses, such as the one year master and two year candidatus, are available through the Open University.

Made-up personas: illustration of potential candidates





Jesper L. Andersen*

"I read an article that all the major companies in Greater Copenhagen have a need for data stewards."

Age: 21.

Family: Single.

Private: Plays handball in a local team and has run his first Nordic Race (obstacle course racing).

Education: Attending his fourth semester at a bachelor (BSc) in data science at IT University of Copenhagen.

Personality: Structured - Outgoing - Keeps things simple -Fast learner.

DS roles Jesper already can fill:

Developer

Administrator



Agent of Change



Frustrations: Even though Jesper's degree equips him with skills and knowledge that are in high demand across both the private and the public sector, the competition is increasing dramatically. Jesper wants to differentiate his profile and focus exclusively on business intelligence in the financial sector.

- Be able to ensure data quality.
- Become skilled on building custom software and databas-
- Enhance his programming skills in statistical and data

Ambition: Jesper wants to be able to assess the quality of shared enterprise datasets, ensure accurate and consistent data across the enterprise and understand the business value of data. He would like to grab that chance and become a DS himself, so he is already considering to build on top of his bachelor with a two year full-time master degree in data stewardship (Candidatus). Currently, he is trying to get a data stewardship internship at one of these companies that would assist him greatly in clarifying his future steps.

UNIVERSITY OF COPENHAGEN 22.10.2021 33

Made-up personas: illustration of potential candidates



- Collaboration between academia and industry result in relevant educations that have a stronger recognition in stakeholder communities, including potential employers, thus improving the student's employability.
- Course design based on experiential learning creates motivation and presents the student with real-life requirements to data stewardship.
- DS work in a multitude of contexts, primarily functioning as bridge between different stakeholders.
- Skills and competencies must not be limited to technical proficiency alone. Hard skills such as coding, software and system knowledge must be supported with Soft skills like the ability to communicate, teach, and assess the value of DS.
- Disciplinary knowledge, teamwork and project management are requirements for curricula.
- Certification should be renewed regularly and further education should be offered to ensure skill maintenance.

DATA STEWARDS

competition or coordination?



Royal Danish Library have joined working group at UCPH: Planning & designing a Data Steward Education - through pre-qualification (launch Autumn 2021)

Drafting a **library specific training module** in the DS programme: 7,5 ECTS

Continuation of the project:

Carpentries involvement (2020-); Participation in GO-FAIR IN: DSCC; RDA WG + IG; EOSC WG: Training & Skills; EOSC Nordic (H2020), DeIC/NeIC nordic collaboration + networking and training events and exploration of potential exchange programmes (European level)



Projects, training & networking activities





Data Steward Education (UCPH), 1Y Master, 60 ECTS Suggested distribution

- Data modeling and management inc. HPC, metadata, archiving etc (2-3 courses) 7,5 - 15 ECTS
- Legal and ethical issues with DM, FAIR, GDPR & Research Integrity (2-3 courses) 7,5 - 15 ECTS
- Characteristics of domain specific data (1-2 courses) 7,5 - 15 ECTS
- Final project 15 ECTS



Prequalification in cross facultary working group organised by the Department of **Computer Science** https://di.ku.dk/english/



Data Science Training for Librarians (DST4L)

3-day workshops (2015-2016)

Bash, Shell (command line operations)
OpenRefine (cleaning of data)
Git, Github (versioning control,
collaboration)

R & Python, Jupyter Notebooks (simple coding and Markdown)

Tableau, Plotly (DataViz)

Networking and learning together (community building)







FAIR data stewardship: Supporting FAIR data interoperability

Welcome to the second Nordic course on "FAIR data stewardship". This course is organised and subsidised jointly by the Danish National Forum for Research Data Management and NeIC.

This is a fully fledged 5-day training event which will provide the much needed found, fonal skills for competent data stewards and data managers in the Nordic countries with knowledge of the FAIR principles and their application. The course will be held by two traine troo. FAII and provides a broad but fairly technical introduction to data stewardsh. Go HEL for more info about the course content. Registration will be limited to max 35 person and the course is subsidised by NeIC and DM-Forum for Nordic + EE participants. Farly registrations (until 1 oct) are priced NOK 6750,-, while standard registration is NOK 8500 (a) b for Nordic+Estonian participants). Non-Nordic participants are not subsidised and p. No. 3750,-



View entries

Prices

Farly registration - NOK 6 750,00

Standard registration - NOK 8 500,00

Schedule

Start date: 18.11.2019 09:00 End date: 22.11.2019 17:00 Register by: 09.11.2019 00:00

Add to calendar

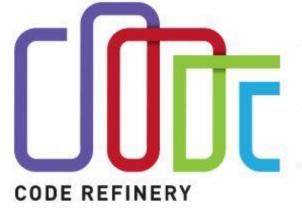
Place

Nordunet as Kastruplundgade 22 D-2770 Kastrup

Contact person for the event

Andreas O Jaunsen +47 99598800 https://www.deic.dk/ da/fair-data-steward ship-kursus





Training and e-Infrastructure for Research Software Development &

We are working with students, researchers, Research Software Engineers from all disciplines and national e-infrastructure partners to advance FAIRness of Software management and development practices so that research groups can collaboratively develop, review, discuss, test, share and reuse their codes.

Check out our upcoming events and workshops



Training opportunities

We offer training opportunities to researchers from Nordic research groups and projects to learn basic-to-advanced research computing skills and become confident in using state-of-theart tools and practices from modern collaborative software engineering.





Lesson materials

We develop and maintain training material on software best practices for researchers that already write code. Our material addresses all academic disciplines and tries to be as programming language-independent as possible.





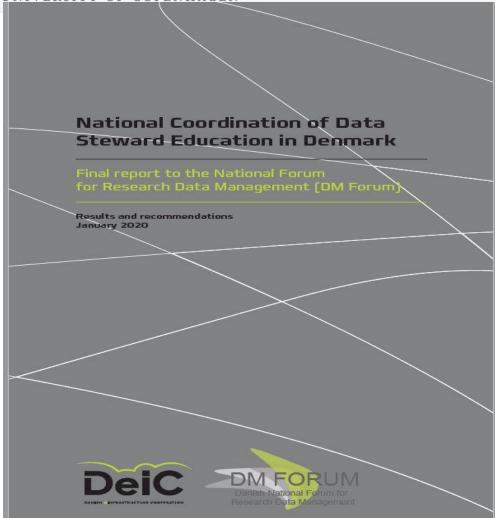
Code Repository Hosting

Our code repository hosting service is open and free for all researchers based in Nordic universities and research institutes. Please contact us if you would like to use these services.









Wildgaard, LE, Vlachos, E, Nondal, L, Larsen, AV & Svendsen, M 2020, National Coordination of Data Steward Education in Denmark: Final report to the National Forum for Research Data Management (DM Forum). 1 udg, Danish e-Infrastructure Cooperation. https://doi.org/10.5281/zenodo.3609516

https://www.zenodo.org/communities/ds-edu-dk











Data Steward Education: Coordination or competition?

<u>Data Stewards to the rescue, please!</u>

A data what? Preliminary investigations into the role and qualifications of Data Stewards

National Coordination of Data Steward Education in Denmark

https://doi.org/10.5281/zenodo.3629191

