Reframing Data Stewardship educations in Denmark and abroad

Lorna Wildgaard, 0000-0002-3900-5058

KUB Research Support, Copenhagen University Library, Royal Danish Library, Copenhagen, Denmark

Abstract

Whether employed in the corporate or in the university realm, a Data Steward's role is to ensure that the data management practises of the organisation follow established governance policies, standards or best practises. The Data Steward is in a position of trust, having responsibility for the governance, management and exploitation possibilities of a researcher's or enterprise's data. However, the professionalisation of Data Stewardship can only move forward and become stronger with improved data steward education opportunities. In this paper we explore the differences and commonalities in academic and corporate approaches to educating data stewards. The aim is to process a set of recommendations that can guide the development of novel data steward educations in Denmark. A content and thematic analysis of 24 current data stewardship educations, 8 academic and 16 corporate, was conducted resulting in identification of 5 recommendations. These are:

- (1) 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
- (2) employability
- (3) Course design based on experiential learning creates motivation and presents the student with real-life requirements to data stewardship.
- (4) Skills and competencies must not be limited to technical proficiency. Being able to assess the value of Data Stewardship, communicate, teach, soft skills, have disciplinary knowledge and the ability to project manage are, amongst, requirements for curricula development.
- (5) Association certification is recommended. Certification should be renewed, and further education opportunities should be offered to indicate commitment to skill maintenance.

KEYWORDS: data stewardship; education; further education; corporate education; enterprise education; academic education;

Data Files (Nvivo-file and Coding scheme) can be found at the Zenodo Community

This document is part of the results from the final report: *National Coordination of Data Steward Education in Denmark*, Jan. 2020 available at: https://doi.org/10.5282/zenodo.3609516

1. Introduction

Research data management in all of its facets can be a complicated and time consuming enterprise and with the emergence of ever bigger, accessible data sets and demands to methodological transparency, the pressure for expert data management is growing. In the corporate world it is becoming clear that the value of data is increasing, the amount of available data is growing and awareness of how using collated data in enterprise increases opportunities for gaining a competitive edge in a crowded marketplace. The General Data Protection Regulation (GDPR) has contributed to both raise the bar of responsible handling of personal data, but also to making exchange of data across the EU possible and legal, in for example connection with trade (Albrecht, 2016). This means that there is a demand for employees with the necessary skills to act as stewards and handle and analyse these data appropriately. Likewise in the university world, researchers are facing increasing demands to make their data collection, analysis and archiving practices transparent, accountable, reproducible (Jensen et al., 2018) and make their research, publication and data processes open, shareable, finable, accessible, interoperable (Ayris et al., 2018; (European Commission Expert Group on FAIR Data, 2018). Thus requiring expert guidance from researcher-support personal with education in data stewardship. In 2016, the European Commission estimated that an average 5% of all research expenditure should be spent on properly managing and stewarding data (European Commission & Directorate-General for Research and Innovation, 2016). In the same year, the chair of the European Open Science Cloud, Barend Mons, estimated that 500.000 Data Stewards will be needed in Europe alone to ensure effective data management (Teperek et al, 2018; Versweyveld, 2016). New job profiles need to be defined and education programmes put in place to train the large cohort of data stewards required.

Whether employed in the corporate or in the university realm, a Data Steward's role is to ensure that the data management practises of the organisation follow established governance policies, standards or best practises. In some instances the Data Steward may be defined as a main stakeholder in developing such data governance strategies or policies, while in other instances the Data Steward is the person who implements these rules into daily practise. Yet, a recent study by Teng & McGrail found that organizations varied greatly in terms of their skills and expertise regarding the rules and procedures around data governance and management. Twenty-four interviewed data stewards and data professionals noted that they experienced differences in interpreting legislation regarding data access requests, resulting in disagreements when they were working with other data stewards. Nearly all noted that there is little guidance provided for the job of "data steward" and either no or very little training when taking on these positions (Teng & McGrail, 2018). The Data Steward is in a unique position of trust. They handle data on behalf of the researcher or enterprise, yet the researcher or enterprise retains the overall responsibility for the research or corporate project during which the data were collected. The question is, without accredited education for data stewards, will researchers and enterprises trust data stewards as a source of expertise? (Terperek, 2018).

Educational possibilities as a data steward in Denmark are few. Corporate bodies offer on-the-job training to fulfill data steward functions relevant for the specific enterprise and product developers offer courses in data stewardship using their product(s). At the university level, data stewardship typically fills a single module on data scientist education, or as in-house training for Phd-graduates who wish to function as research support and at the same time put their own research experience to good use (Kurepati, 2019). The pace in which the field of data stewardship is developing and how international open science initiatives are being implemented at universities and by funders has presented us with the acute need to unite data management roles that were previously split between the researcher, the department, the university, the

library and IT specialists. Consequently, there is an opportunity to re-frame the profession of data steward and develop accredited educational opportunities for data stewards.

The first step offered in this paper is to consider how universities could design an education, possibly nationally coordinated, for data stewards. To investigate practices and innovations in data stewardship education, this paper investigates the differences and commonalities between the corporate approach and the university approach to Data Stewardship education. The aim is to produce a set of recommendations, that should be considered in the development of Data Steward education at Danish Universities.

2. Methodology

Data stewardship educations were identified and collected in Google to ensure international coverage of both corporate and academic educations. Only educational programmes where data stewardship comprises the full course content were included, therefore educations that reduced data stewardship to one module of a larger course were excluded. Included programmes also had to be established and not suggestions or draft recommendations for data stewardship training. Google was chosen because the database covers websites and educational catalogues from institutions of higher education, research councils, companies, EU projects and materials produced by the data steward community and other public and private bodies offering educations in data stewardship. Google covers a vast range of sources describing data stewardship educations, i.a. webpages, articles, slide shows, promotional materials, curricula and advertisements from a vast variety of information providers.

We searched Google on the 3rd of April 2019 using combinations of the following search terms: *education, uddannelse, course, courses, kursus, kurser, training, "data stewardship", "data steward"* (Appendix 1). As Google search algorithm weights the importance of the terms with regards to their order in the search string and inserts an automatic "AND", Google was searched multiple times using a few terms at a time. The search produced in total 710 hits, which were sorted manually for educational programmes dedicated to providing qualifications to become a data steward. The search was limited to English, Danish, Swedish and Norwegian sources as these are the languages in which the authors of this review are fluent. The searches produced a lot of noise but it proved a quick process to exclude job advertisements, discussions, posts and blogs expressing the needs for data stewards.

In total 35 Google snippets describing data steward educations were collected and organized in Excel spreadsheets. One snippet contained links to a further 13 educations and these were included in the spreadsheet, bringing the total to 48. After duplicates were removed, detailed information on 44 possible educations were collected. Information collected included: the name of the education, the link to the education, if the education was corporate or academic, admission requirements, intended audience, certification details, discipline, duration, ECTS, format, name of institution offering the education, job titles the education could lead to, the outreach of the programme, skills and competencies described in the curricula, aims of the education, learning aims, and the type of education (vocational, up-grading, university, etc).

At this stage 20 posts were excluded, as they proved on closer inspection to be policy documents (n=1), curricula development paper (n=1), deadlinks (n=1), user profile/user account required (n=1), duplicate (n=2), service to find education (n=1), not data steward education (n=8), no information (n=1), best practice guidelines (n=4).

The resulting data set consists of 24 educations. Throughout the remainder of this review, in both analysis and discussions, the educations will be referred to by their abbreviations, Appendix 2. The abbreviations will be written in italics.

2.1 Content analysis

There were 2 levels to the content analysis. A content analysis intends for themes to emerge in an inductive way from the text under analysis and creates thematic structures and patterns for more detailed analysis. This process has its weaknesses, as described in (Hsieh & Shannon, 2005), as it is dependent on the quality and completeness of the written text, but it is useful for identifying recurring concepts, patterns and themes across the data set.

The first level was carried out manually and explored the descriptions of each education to identify broad conceptualizations of programme structure: outreach, audience, domain, admission requirements, course format, duration of course, job titles the education could lead to, certification and certificate renewal. The second level of analysis was conducted in Nvivo, where a more detailed reviewing and coding of the data could take place. The reviewed data described the aim of the education, skills and competences, learning outcomes and certification requirements with the purpose to identify themes in educational content and the different facets of concepts in respectively academic and corporate data steward educations. The approach for level 2 was a four stage process:

1. **Coding**: identifying features of the data that may inform the research question. The entire dataset was coded. Codes and relevant text extracts were collocated for thematic analysis.
2. **Thematic analysis**: the codes and collocated text extracts were examined to identify significant broader patterns (candidate themes), commonalities and differences across the data.
3. **Reviewing themes**: Candidate themes were checked against the dataset, to determine representation of the concepts of data steward education and ensure we have identified the information necessary to answer the question. Themes were adjusted, refined, combined and discarded accordingly.
4. **Defining and naming themes**: The scope and focus of each theme is defined. Each theme is named.

3. Results

3.1 Level 1: program structure

The search identified 8 educational programmes offered by academic institutions and 16 corporate educations. The first level of the conceptual analysis, Table 1, identified nine common themes in data stewardship educations.

Table 1: first level of conceptual analysis

	Academic education, n=8	**Corporate education, n=16**
Outreach	International, national and departmental programmes	Global or national product users
Audience	Undergrad & graduate, further education for information professionals	Product users, customers or employees at an enterprise company, data professionals, IT engineers, and administrators
Domain	Life & health sciences, diverse domains of human endeavor, library and information science, engineering, computer science	Business, earth science, marketing, enterprise, industry, R&D companies
Admission	General and specific education requirements, expert level knowledge, disciplinary knowledge	Data management professional expertise, knowledge of specific tools/operating systems, employed at organization, open to all (on payment)
Course format	Short workshops, seminars, university course (weekly lectures, practical sessions including internship)	Webinars, online learning (self-paced), conferences, events, onsite training & tailor-made solutions, virtual instructor-led courses, boot-camps
Duration	1 day-2 years	2 hours to 5 days
Job titles	Data Steward, FAIR data steward, librarian, network engineer, records manager, archivist, information architect, data Scientist, data analyst, business intelligence-developer, data consultant, project managers, data manager, data engineer, data architect, data strategist	data steward, information server, new collar jobs that do not always require a traditional degree, claims analyst, lead underwriter, senior financial analyst with additional responsibilities of a data steward, chief data steward, data governor, data manager, platform administrator, data architect, IS auditor, consultant, educator, IS security professional, risk professional, internal auditor
Certificate	University degree, certification of attendance from professional community	CDS certification; skills badges; use-expertise certificates, professional certifications e.g., PMP, CDMP, CBIP, CAMP, EMC
Certificate renewal	No	Yes. Regular skills update required, typically every 3 years

3.2 Level 2: thematic analysis review

In order to highlight the commonalities and differences between the corporate and academic approach to data steward education key concepts in the education were identified. These provided the "macro-themes" that helped us understand the different approaches to data steward education. The first macro theme is the general environment in which the education takes place, and the input from the surrounding context within which the education occurs. The second is the infrastructure supporting the participants of the education. The individual themes within an education are coded to reflect the approach and expectations of the education. They include the learning aims, skills and competencies, aim of the education, demands and expectations to the participant during and after the education, and certification which together summarize the structure and content of the programme. These can be described as which qualifications the educational institution is investing in the participants, what the education is trying to achieve and the perception of data stewardship.

Theme	Word Count	Sentences coded	Percentage of coded data
Academic context and input from the surrounding context (AC)	1996	114	14
Corporate context and input from the surrounding context (CC)	2615	122	18
Learning aims (LA)	2,128	301	16
Skills and competencies (SC)	1,540	111	10
Aim of the education (AE)	2,609	92	19
Demands and expectations to participant (DE)	548	33	4
Certification (C)	857	73	6
Employability (E)	1,788	103	13

Table 2: thematic analysis, macro themes

Candidate theme	Second level	Word count	Sentences coded	Percentage of coded theme
AC	Market-relevant skills	703	48	38
AC	Business awareness	440	15	24
AC	Collaboration with industry	435	13	24
AC	Feedback	264	9	14
CC	Position the role of Data Steward	1,000	46	39
CC	Career development	997	44	39
CC	Input from industry	358	23	15
CC	Expertise in a software	175	13	7
LA	Data Management	395	35	29
LA	Product expertise	271	34	18
LA	Technological expertise	266	16	18
LA	Disciplinary knowledge	183	17	12
LA	Ethics and morals	89	3	6
LA	Communication skills	81	6	5

LA	FAIR	76	5	5
LA	Policy and strategy awareness	56	4	4
LA	Value for end-users	49	12	3
LA	Didactics & pedagogy	6	1	1
SC	Data management & planning	274	32	23
SC	Expertise in specific software or product	236	26	18
SC	Domain & industry knowledge	207	15	17
SC	Experimental design & methodology	161	13	13
SC	Ethics and responsible conduct	133	12	11
SC	FAIR data stewardship	118	5	10
SC	Information science	31	3	3
SC	Librarianship	31	3	3
SC	Machine learning	12	2	1
SC	Data analysis, visualization and mining	8	2	1
SC	No information	1	5	<0
Soft Skills (SS)		326	27	
SS	Liaison	99	8	30
SS	Communication	91	7	28
SS	Collaboration	48	3	15
SS	Project management	38	4	12
SS	Mediation of policies & legislation	33	2	10
SS	Teaching	8	2	2
SS	Career & education	9	1	3
AE	Establish good practices	1,744	57	66
AE	Proficiency in a software/platform	338	18	13
AE	Industry readiness	281	15	11
AE	Excellence i partner institutions	265	14	10
DE	Undergraduate qualification or other qualifying exam	412	15	54
DE	Works with data on a professional level	195	11	26
DE	Disciplinary knowledge and skills	80	2	11
DE	Works at level of researcher	47	3	6
DE	Programming skills	13	1	2
DE	No information about demands or expectations	6	3	1
C	Association accreditation	162	15	66
C	Renewal requirement	58	3	25
C	Certification of attendance	7	2	3
C	Badges	6	1	2
C	University certification	6	2	2
C	No information	3	2	1
C	None	3	3	1
E	Industry readiness_Academic approach	988	64	55
E	Industry readiness_Corporate approach	800	39	45

Table 3: thematic analysis, micro themes

4. Discussion

4.1 Level 1: content analysis of programme structure

In this first part of the discussion we apply a content analysis to gain an overview of how the 24 educational programmes in our corpus are structured and marketed to potential course participants. The 8 common themes are summarized in Table 1. Below is a comparison of the commonalities and differences in the approach to educational programme design and composition offered by respectively academic and corporate educational centres.

4.1.2. outreach

Outreach to the surrounding data stewardship community, disciplinary actors and other stakeholders is a functional design element of the curricula of the academic educations. Common strategies include the use of industry and disciplinary experts as instructors create alignment between the education and the job market (TUD, DTL, HA), making the course attractive to a diverse audience with wide range of experiences, working with national and international networks to share and inform best practice, and, using teaching formats such as seminars and workshops with built in time for networking. Feedback is also an important form of outreach. Upon completion of its' course in data analysis and stewardship in Spring 2019, HA will critically assess the course to determine where improvement is possible. To ensure industry relevance, a questionnaire will be sent to companies to inquire about their specific training needs. In addition, they are actively reaching out to companies for interviews about their challenges and training needs and for an inventory of possible company trainers. Feedback will be continuously incorporated in subsequent versions of the courses to be sure they continue to address the evolving needs of companies in their area.

Both HA and DTL focuses on the development and coordination of appropriate training, internships and traineeships in the Life Sciences and Health sector by matching companies' job requirements and profiles to curricula and develops new training programs where needed. DTL partner with over 50 private and public institutions in their training platform, where HA is one of them. On a smaller level, USD and UP encourage the participant to engage with local companies and work on real-life projects in collaboration with the company.

Outreach in corporate education is similar to the academic approaches, typically achieved through credit transference to recognised partner organisations (DMU, CU, ISACA and ANA), providing tailor-made, onsite teaching and including industry experts in workshops and webinars (ISACA). An exception is ESIP. ESIP has a philanthropic approach in that it helps researchers and associated groups that produce, interpret and develop applications data. ESIP helps this community facilitate connections, collaborations and innovation across the science, technology and applications communities.

4.1.3 audience

The academic educations are aimed towards graduate or doctoral students with a strong mathematical, data management or computational backgrounds, and primarily with disciplinary expertise in the life sciences (*CC, DTL, NeIC, TUD, UP, USD, LERU*). *HA* places itself in the intersection between the academic and the corporate world, targeting researchers and professionals in universities and local companies with the aim to bring together centers that produce data in the life sciences for industry.

The audience for the corporate educations' fall into two categories, both aimed at skill maintenance and development. First, educations clearly marketed to specific product users or company employees who are working with data quality violations, data governance, marketing, defining data policy for their businesses, and monitoring policy compliance (IBM, DV, HP, QU, DELL, *DMU, DV, CU, IA, ANA*). The aim is to attract an

audience of key people who have the strategic perspective and the requisite knowledge to help an organization successfully leverage information for maximum value using the promoted software and platforms.

Second, professionals already established as data stewards or IT professionals, including information specialists, researchers, Post docs, and PhD students in data-driven fields who want to become data stewards or who are working with data stewards and want to learn more about how data stewardship can give a company strategic advantages. reduce costs and mitigate the risks associated with using and governing data (PC, ISACA, CDS, ESIP, ARMA).

4.1.4 domain

Application domains in the academic educations are research data produced in the life and health sciences (HA, DTL, LERU), all fields of human endeavour, including business management, science, sports, health-care management, (CC, NeIC, UP, USD), geo-sciences and aerospace engineering, engineering, applied sciences in mechanical, maritime and materials engineering (TUD).

In the corporate educations the application domain is data created and collected in enterprise companies (IBM, IA, DV, CU, DELL, ISACA, HP) for example in the fields of finance, insurance, oil, telecoms, software development, consulting, industrial government (CDS), or earth science (ESIP). The application is geared towards business intelligence and market activities, data governance, gaining customer insights, data integration and security and, importantly protection of information assets (DMU, ANA, QU, PU, ARMA).

4.1.5 admission requirements

Admission requirements for academic educations are undergraduates with an academic background from an accredited educational institution (*CC, UP, USD)*. If the education is also open as a vocational program or further education for information professionals, it is expected the candidate works with data at the level of expert for example as a technician, and preferably with substantial knowledge of the scientific domain he or she operates in (*TUD, DTL, LERU*). For the TU Delft Data Stewardship programme (*TUD*), it is further expected the applicant has disciplinary knowledge at the level of a researcher. The data science skills necessary to function as a data steward are taught through an extended programme. Apart from the candidate's academic abilities, professional experience and accomplishments are considered (*UP*). Programming skills for master level data science specialisations that lead to data steward positions are essential (*USD*). Admission requirements were not found for *NeIC*.

Generally in the corporate world, admission is restricted to employees at a company, interested in working with Data Stewardship (CDS, DMU) or customers using a platform or software (DELL, IBM, QU, CU), where progressive courses from beginner to expert are offered as part of career path modules. Training courses are available only to the noted audiences. Payment can be required for participation in the course, certification exams and application processing fees (averaging 200 dollars each). Prerequisites, pertinent technologies covered and additional specifics for each certification track and knowledge level are described as part of product packages. The ISACA training modules for example require the participant to have full time work experience in relevant job practice areas and a current Certified Information Systems Auditor (CISA) exam as a prerequisite, whereas ANA develop courses specifically with the ANA marketer in mind using the CAMP marketing ontology.

No admission requirements were found for PC, though they aim courses at librarians, researchers and data experts at universities and other researcher intensive companies.

4.1.6 course format

The academic courses aimed at graduate students maintain the classic university format of full-time on campus in-class teaching, lectures and practical lab sessions, assignments and short term projects with local business (AC, USD). For professionals the format is more flexible consisting of short face to face workshops and seminars (DTL, LERU, NeIC), offered only by AC as online modules. TUD have an innovative approach combining general training in data and researcher support using external experts from industry and research with in-house training from local subject experts where the disciplinary specifics of data management are taught. Only HA did not provide any information about the format.

The corporate educations are designed to fit in to a busy working day. Online courses are offered through interactive webinars (CDS, ESIP), or as self-paced, on-demand virtual classroom training with supplementary materials such as PDF, books and tutorials (IBM, ANA, CU, DELL, ARMA, ISACA, HP). Seminars and bootcamps requiring physical presence are short and intense (IA, QU, DV, CU, PC ISACA) while some educators offer on-site, tailor-made client focused classes ((DMU, ANA, CU, ISACA)

4.1.7 duration

The duration of the academic educations and courses range from one and a half days focusing on competence development and upskilling (LERU) up to full master courses, full-time for 2 years (USD, UP). Similar to LERU, NeIC offer multiple intense 4 day upskilling workshops. CC offer course modules on a part-time basis, spread over an academic term and, worth 3 credits out of a required 4.5 credits students at Canisius College need per term. DTL prefer a strategy of short workshops and seminars interspersed throughout the academic year. No detailed information about the actual duration of the training courses was found for HA and TUD.

The corporate courses are noticeably shorter. CU offer a suite of training possibilities designed to fit in the working day. From 2 hour webinars to 4 day bootcamps, and full week in-person training. Once enrolled, applicants have 30 days to successfully complete the exam and are granted five attempts. If the applicant is unsuccessful after five attempts they must wait 30 days before they can re-enroll in the exam course. There are no extensions or exceptions. All certifications expire 2 years after issue. Likewise, ISACA refers to the CISA CPE policy, that requires the attainment of “CPE hours” over an annual and three-year certification period while the CDS course package is more flexible and offers 3 courses and 3 exams comprising of 15 hours of education that can be completed at anytime within 24 months to pass.

One day workshops are offered by QU, IBM, DV which all aim at data stewardship using a specific suite of software and tools. ANA, HP, PU and DMU offer 3-5 day workshops that support career path development, marketed specific job-profiles, application areas and expertise.

Information regarding the duration of the training could only be gained by contacting ESIP, IA, DELL, ARMA after creating user profile, which the author of this rapport chose not to do.

4.1.8 job titles

In academia it is clearly stated in the description of the education, that data steward (HA, CC, DTL, TUD, USD, UP) or FAIR data steward (NeIC, LERU) is the job title gained from completing the course. The educations emphasize that data stewardship skills are sought after competences and can beneficially be

applied in areas such as data scientist, data analyst, business intelligence developer, data consultant, data manager, data engineer, data architect, data strategist(USD); web-designer, software engineer, librarian, records manager, archivist, information architect(UP).

In corporate educations data stewardship is marketed as one skill in a suite of skills important for professional growth of the person and of strategic importance for the growth of the company. A prime example is how the title of data steward is presented as a tool for marketing, in the use and collection of behavioural data in online advertising (ANA). Whereas PC educate solely "FAIR data stewards", a new profession working with the FAIR principles and their application in data management, a different breed of Data Steward is found at CDS, IBM, IA, DMU, ANA, QU, CU, DELL, ARMA, HP. Through these educations the job title "data steward" prepares the graduate for positions such as an information server and "other new collar jobs that do not require a traditional degree" (IBM), data manager (ESIP), claims analyst, lead underwriter, financial analyst with additional responsibilities of a data steward (DMU), business data steward, chief data officer (DV), data governance, data manager, platform administrator, data architect (CU, DELL), and other professional IT related positions such as data auditor, consultant, educator, IS security professional, risk professional, chief information officer and internal auditor (ISACA).

4.1.9 certification

No information about certification was available for HA. It is unclear if a certification programme is offered by TUD and DTL. Traditional university certification is offered by CC, UP and USD at the completion of the education.

The corporate approach is to reward the participant with certification dependent on level of expertise (CDS, IA and CU). Similarly IBM and HP add badges to the participants profile, that showcase their skill sets at a glance and enable the employee to design career pathways and tailor courses to gain certain badges. DMU, ARMA and ISACA give credits for each completed module or course, and these credits can be transferred to appropriate certifications from sister institutes and thereby are recognized as standardized qualifications within the community. ANA and DELL certification require renewal after one year and the participant is obliged to maintain and develop their knowledge through active participation in networks and training events. Only PC appear to reward with a certificate where attendance is the only requirement, however how ESIP, QU and DV certify their courses is unclear.

Further details about the accreditation scheme and renewal policies are described in the thematic analysis, section 4.2.7.

4.2 level 2: thematic analysis of programme content

In this second section of the discussion we conduct a thematic analysis of the descriptions of the academic and corporate educations in our corpus. Eight macro themes and 55 micro-themes were identified (Table 2) and are used to structure the following discussion of the commonalities and differences between the mission of the education, learning aims and expected outcomes of the academic and corporate educations.

4.2.1 Academic and corporate context, input from surrounding environment

The academic educations use their collaborations with industry as input in course content and to inform the relevancy of the learning outcomes and the production of competent data stewards from the programmes. Further, traces of business awareness can be observed in the motivation and structure of the programme (CC, DTL, NiEC, TUD, HA), operationalised in course modules where an understanding of the Data Steward occupation, activities, skills and attributes needed to be successful at these activities, understanding the business environment, how an organisation, company or research discipline operates, and its context are course requirements. As the educations are situated in different disciplines (Health, Life Sciences, Business and enterprise and Data Science), candidates from these courses will to some extent be subject-specific experts and therefore bring different skills and knowledge to the job market. This also means that the candidates will have have different degrees of understanding about the overall trends and expectations in the field of research data management (DTL, TUD). Therefore, in order to ensure that candidates deliver a general toolbox of appropriate skills and expertise, training programmes, case-studies and internships are developed in collaboration with industry (HA, CC, DTL, TUD, USD). Consequently, while data analytic programmes are rapidly being developed and marketed to data stewards (USD), other data stewardship programmes prioritise the following market-relevant skills: the ability to work in and lead multidisciplinary teams (CC, DTL, TUD, HA), to communicate with different audiences (CC, DTL, TUD, HA, UP) and have a solid grounding in ethical and legal data stewardship (CC, NeIC, DTL, TUD, HA, LERU). These skills are described as "(essentials) to enable a data supporter to take the first steps towards supporting researchers in storing, managing, archiving and sharing their research data" (TUD). Finally, feedback from both industry and course participants are used to inform the academic profile of the programmes (USD, HA, CC, TUD, DTL, UP). These interactive initiatives include direct influence from the student to the professor in the design of the speed and the content of lectures (USD), using local companies as internships, cases and guest speakers, and surveying local industries to find out their specific training needs and the challenges they face in data stewardship (HA, CC, DTL, TUD).

The academic educations are marketed towards information professionals and graduate students, likewise the corporate educations are rationalised to fit a person with an already existing understanding of the functional area and data flows of the company in which they are employed. Accordingly, Data Steward roles in the corporate context are expected to be filled by someone who has been in post for some time or where Data Stewardship will be a part-time incremental task to their primary job (DV).

In many ways Data Stewardship is a classic matrix management role (DV. CDS, DMU, ANA, PC, ARMA, IBM). This could explain why the educational programmes in the corporate context are situated using overlapping themes of career development and the strategic positioning the Data Steward within the company.

Career development is supported through cost-effective and essentially fast-track ways to learn and gain certification, expertise and experience that are recognised by employers and by the surrounding industry (CDS, DV, QU, CU, DELL, IBM, IA, DMU, ANA, ISACA, ARMA, HP, ESIP, PC). Just as corporate programmes are marketed towards individuals, they are equally targeted companies to create awareness of the need to have and benefits of having a data steward employed at the company. In a minority of cases buying a specific data management platform to enable good data governance practices is required (IA, QU, DV, CU). More interestingly, is the requirement of the identification of the organizational structures and roles that have to be in place within the context of a company for an enterprise-wide data stewardship initiative to be successful (DMU, ANA, CU). Further, learning through trusted industry experts is a seal of trust, that helps drive personal and organizational success, growth and innovation (ISACA) while at the same time avoiding the common pitfalls that derail typical governance initiatives (DMU). The programmes designed with and/or certified by industry partners (CDS, IBM, ESIP, IA, DMU,ANA, CU, ARMA, ISACA, HP) thus use input from industry and partners as a validation tool for the education, the learning aims and the ability of the candidate to apply the skills learnt on the course in practice(CDS) and importantly apply concrete results of these practices in the development of the enterprise company where they are employed. Thus giving the company a competitive advantage and adding value to data in compliance with data governance standards and practices (PC, CDS, IBM, DMU, ANA, DELL, ARMA, ISACA, HP).

4.2.2 Aims of programme

The primary aims of the academic educations and the majority of the corporate educations is to establish good data stewardship practices and enable the candidate to be resourceful in the solution of the challenges connected to data stewardship in industry and other fields where data is used to inform business and research decisions. The practical aspects of the educations aim to reinforce the learning experience and provide the candidates with the necessary skills to implement their knowledge as soon as they enter the position of Data Steward in an organisation.

Good data stewardship practices are operationalised as learning aims, skills and competences and are described in the following sections, 4.2.3 - 4.2.4. In the academic and corporate educations the principles of good practice are managing data in ethical, secure and effective manner, technical ability to work with data, maintain and add quality in processes (CC, NeIC, TUD, USD, HA, CC, DTL, UP, CDS, DMU, PC, ESIP, IBM) and drive the establishment of data stewardship as a profession (NeIC, LERU, TUD, DTL, CDS, DMU, DELL, HP). Further, as some of the educational programmes are continuously developed in collaboration with partner institutions, both private and public, these programmes have the obligation to support professional development and excellence in these institutions (DTL, TUD, ISACA, HP) and accordingly the education will progress from teaching good practices to being able to recommend best practice.

The exceptions are IA, QU, DV, CU where the aim is proficiency in their suite of DS tools.

4.2.3 Learning aims

Beginning with the academic educations, the main learning aims are firstly, familiarity of the stages of the research data life cycle and what it takes at each stage to work with research data. Secondly, to learn about and practice ethical and responsible data stewardship. The corporate educations agree with the academic that the mission of the data steward is to formalize the role of data stewardship and to drive recognition of

data steward as a professional designation. Yet it is clear in the corporate educations that the learning aims are strategically beneficial for the participants career. Any increase the participants knowledge will add to their value and employment options in a competitive market.

The learning aims in the corporate educations can be grouped in three perspectives of data stewardship.

1. If the data steward identifies as primarily a data manager, then the learning aims include preserving the scientific record, knowing the elements of a data management plan, knowing what to and what not to archive and how to archive (FAIR) data, understanding how to provide access, share and re-use (FAIR) data, and working actively on the issues of long term archiving (CDS, DMU, PC).
2. If the data steward identifies as a support function, then the learning aims are slightly different. Here the aim is to be able to work strategically with collection and acquisition policies, object identification, assess potential communities, identify data external to the company relevant for the field, assess transfer options, curation, costs and feasibility (CDS, ANA, QU, HP).
3. If the data steward's position in the company is embedded in data governance, then a third set of learning aims become apparent, such as how to implement the governance plan for managing information requirements, how to address the political issues and organizational challenges of data and importantly, defining the roles and responsibilities of the data steward in the enterprise (DMU, CU), with special emphasis on GDPR data protection regulations (DV).

Eight distinct themes were identified in the learning aims across the academic and corporate educations.

Technological expertise
Not surprisingly the largest shared theme across the academic and corporate educations are the technological learning aims. All educations expect the students to have advanced data and technological expertise. Specifically in how to download, clean, and prepare data for future analysis, including how to work with ontologies, and document the process, as well as understanding how seemingly harmless actions can pose threats to the information security of others. Described in more detail the involved processes the student is expected to master are the ability to , analyse the advantages and disadvantages of different problem-solving methods, critically refer to own and others' research results and scientific models, develop new variants of the learned methods when the specific problem or customer requires it and be able to disseminate and discuss professional and scientific issues with both professionals and non-specialists. Project management is a highly praised learning aim, where the student is expected to be able to plan and carry out high-level projects, including managing work and development situations that are complex, unpredictable, and require new solutions and new collaborations.

TUD stands out from the other academic educations by emphasising the importance of the student being able to gather quantitative and qualitative metrics to assess the success of Data Stewardship. A trait seen more commonly in the corporate educations as evaluating the "return on investment".

Value for end-users:
In the previous section TUD was accentuated because of the technological skills needed to asses the value of data stewardship. More commonly in academic educations, the learning aim is to able to communicate (but not assess) the value of FAIR data for end-users. In the corporate educations the ability to assess return on investment is an important learning aim, in regards to agency requirements, return on the company's investment, on the public's investment and on the level of service and other curation activities.

Communication
The students will learn how to communicate the importance of data management. The educations aim to embed data stewards in university faculties and communities who have the skills to facilitate the broader trends in research data management and advise on practical solutions that support the research life cycle. This will establish complementary information needs e.g., format, data descriptions, provenance, reference information, context, fixity of information (TUD, LERU and ESIP). Communication outside of the data stewards immediate community is also an important ability, to be able for example at identify other sources of expertise (ICT, Legal, advice on Tools ...) and connect with potential interessanter through tailor made training.

Data management
Students will be able to assist in planning the collection, management and publication of data in new and ongoing projects. They will have an awareness of the ethic and moral issues that arise in working with large data sets, and understand the steps that need to be taken to protect the rights and privacy of the individuals involved. They will be able to name the purpose of a data management plan and list the most important issues to consider in writing different parts of a data management plan. Further, it is expected that a data management plan meets the requirements of funders, submission agreements and undertake validation checks (AC, LERU, TUD, ESIP).

Disciplinary knowledge
Disciplinary knowledge is key to embedded data stewardship. On completing the educations the student is expected to be familiar with all the stages of the research data life cycle within a field of practice and accordingly be able to identify sources of information within and outside said field on data management and data stewardship (LERU). Connecting data stewardship with the surrounding community entails maintaining regular contact with stakeholders and providing advice, particularly on specific issues related to data and publication practices within the field (TUD). Disciplinary knowledge also encompasses methodological knowledge intended for practice in professional areas (USD) and the the ability to understand on a scientific basis, how knowledge is created, used, stored and other scientific and policy issues related to data (TUD, ESIP).

Ethics, policy and strategic awareness
Ethical data practices is a common theme across the academic educations. Students learn how to download, clean, and prepare data for future analysis, and document the process, as well as understanding how seemingly harmless actions can pose threats to the information security of others (CC). They will be expected to be able to interpret and apply common policies regarding information ownership, security and privacy protection within their field of practice (LERU).

FAIR
Another theme common across the academic educations are the FAIR principles. DTL and LERU explicitly state FAIR data is a point of practice. The student is expected to learn how to evaluate research metadata using the FAIR data principles and know what it requires at every stage to create FAIR research data. NeIC require more from the student. Both technical expertise in how to actionize the FAIR principles when working with semantic Web and linked data, and didactic expertise in advising on a FAIRification processes and the value of fair data for end users.

Product expertise
Expertise in specific software and platforms is a major theme in corporate educations. The learning aims described by IBM, IA, QU, DV and CU expect the student to gain knowledge and skills in data steward components available in their suite of tools. The participants learn how to plan, execute and follow-up end-to-end solutions, and also how to engage with end-users through written, oral and visual modes of communication.

4.2.4 Skills and competencies

In this area, the academic educations are in strong agreement that data management and planning is the most important skill for the Data Steward. Data management and planning includes: handling data in all phases of research, including the ability to describe the research life cycle and the basic options to store, backup, organize and document research data; support researchers in storing, managing, archiving and sharing their research data and making sure it meets the requirements of funders; and finally name the purpose of a data management plan and list the most important issues to consider in writing different parts of a data management plan (DTL, TUD).

The corporate educations also agree that understanding of the fundamentals of data management, data stewardship, data quality, data governance, metadata management, the intersections and relationships of metadata management to data governance and stewardship are essential skills (CDS, DMU). The skills in the corporate educations have a strong enterprise application context, such as the ability to develop the company's roadmap for data governance and stewardship (DMU) and based on this knowledge, initiate the project plan. There also appears to be a stronger focus on the ability of the Data Steward in a managerial position, responsible for defining the data governance and stewardship charter for the organisation or company, and operationalising this charter in concrete record management, mobile communication and information management practices (IGP). In contrast, the academic educations prepare the data steward with skill sets directed towards managing research data in collaboration with researchers.

Soft skills make up 21% of the coded skills and competencies descriptions, therefore it was decided to undertake a third level analysis of this theme. Soft skills include skills such as communication abilities, collaboration skills, instructional skills, and leadership traits. In contrast, "hard"skills – such as technological skills in machine learning or experimental design – are readily measurable and (importantly) easier to train. Primarily, the soft skills taught in Data Steward educations are communication and liaison skills. Liaison skills build on the Data Stewards networking abilities and knowledge and contact to the surrounding community. They are expected to be able to identify sources of information and support within and outside the organisation or company on data management and data stewardship (DTL, TUD). Because the Data Steward understands the relationship other roles and functions in or outside the organization have for data governance, they are expected to act as a liaison between the IT department and their business unit in an organization, showing the link to the classic definition of the word "steward": a person who is responsible for managing something on behalf of someone else (DMU, ANA)

4.2.5 Demands and expectations to participant

Demands and expectations concern the performance of the participant during and after the educational programme. Admission requirements were described in section 4.1.6.

During the programmes participants are expected to combine existing disciplinary knowledge with new knowledge and skills taught on the programmes, that ultimately will lead to a job as a data steward (USD). For the academic programmes that means knowledge about research practice in a specific field, typically from a bachelor (CC, UP), from a master (DTL) or Phd education (TUD, DTL , LERU). For the corporate educations work experience and in depth knowledge of the enterprise including how the business works, customer profiles and the surrounding market (CC, ANA, CDS, ESIP; DMU, ARMA, ISACA, PC). As all areas of research and industry are expected to have an increasing need for professional data stewardship, the academic educations encompass as many disciplinary application areas as possible, which is why the graduate educations in our corpus accept all forms of bachelor and professionals. Accordingly, based on the participant's profile, some introductory courses may be waived on the academic programmes (CC, UP). For example, this might occur for a mature student with an engineering degree, and thus strong computational and mathematical skills, or a finance degree with strong business and mathematical grounding. Applicants for graduate study must have earned a bachelor degree from an accredited college or university with a scholastic average of B (3.0 on a 4.0 scale) or better, corresponding to an average of 10 on the Danish marking scale (CC, HA, and possibly USD but this is unclear as transcripts of bachelor content and subject areas covered are used to assess the applicants suitability to the programme rather than point averages that are only used in cases of uncertainty). There are thus both expectations to the participants academic abilities (DTL, UP, HA, CC, USD, PC,) disciplinary knowledge (TUD, DTL, ARMA, ESIP) professional experience (DTL, TUD, LERU, PC, DMU, ARMA, ANA, DELL, ISACA) and competences as an data-information professional (NeIC, UP, PC, CDS, IBM, DMU, CU, DELL, HP, IA, QU, DV, DELL, HP).

4.2.6 certification

The University of Southern Denmark, University of Pittsburgh and Canisius College offer traditional university certification based on a credit system that the rewards graduates with the title of MsC computer Science (USD) or a masters in Information Culture and Data Stewardship from the American Library Association (UP). Similarly CC offers Data Stewardship as a specialisation supported on the master course, rewarding the student with 3 out of the 4½ credits they need to pass the term. No information on certification could be found from HA, LERU and NeIC, while it is unclear if a certification programme is offered by TUD and DTL.

Even though the DMU courses are corporate, they are certified for credit by several bodies, including the University of Illinois and DAMA International. This collaboration indicates that universities and companies are uniting to develop courses and curricula. Likewise the credits acquired on the DMU courses are transferable to other professional certifications e.g., Project Management Institute PMP; CDMP and CBIP (DAMA International and TDWI), thus standardising expectation to skill level and recognizing the candidates qualifications within the practicing community. Corporate accreditation schemes do not underplay the importance of association accreditation and the strategic benefits of making visible for the candidate and the company (CDS, IA, DMU, ANA, CU, DELL, ARMA, ISACA, HP).

Typically, the corporate courses have different levels of certification and can to a greater or lesser extent be tailored to the career path of the participant. Learning paths are supported by HP and offer certificates that

allow the candidate to follow a recommend path of courses and exams in regard to their career ambitions. He or she can track their training, progress and certificates in the Learning Center. IBM use a similar system that rewards the participant with badges for skill development, that can be displayed on the individuals profile. To earn the CDS Ex designation the participant must demonstrate a combination of great Expertise, Experience, and Excellence, and the certificate indicates the level of the students proficiency in these three focus areas (CDS). A similar approach is found at CU, where 5 levels of proficiency in Collibra products and data stewardship are recognised.

Unlike in academia, corporate certifications require updating periodically to ensure the skills remain relevant (DELL, ANA, ISACA, ARMA). Failure to do so results in the certificate no longer being recognised and hence the person's skill set is out-dated. DELL proven professional certification requires an update every 3 years whereas ANA certification lasts for one year and then has to be renewed and is kept current by attending conferences and training throughout the year, amassing minimum 10 CEUs (continuing education units). The ISACA policy requires the attainment of CPE hours over an annual and three-year certification perio. ARMA has both the prerequisite of eligibility requirements (such as years of work experience) to sit for the certified exam as well as ongoing requirements (such as continuing education, renewal fees), and/or recertification requirements. Certifications usually include passing of an assessment covering a broad area of knowledge and skills and award a title and a designation, such as the ICRM's Certified Records Manager and "CRM" or ARMA's Information Governance Professional and "IGP" that is a recognized standard in the community.

Only PC reward participants in their workshops with a certificate of attendance. How assessment of the learning objectives and outcomes are made is unclear.

Information about how participation and completion of courses offered by QU, IA, ESIP and DV are recognised is unclear or missing.

4.2.7 Employability

The industry-readiness of life scientists (HA, DTL, TUD) and the rapidly developing field of using data to inform business and research decisions (CC, NeIC, UP, LERU, USD) are the primary motivation of the academic educations. The role of data analysis and data stewardship in research practices will only increase and all of these courses offer the chance to become industry-reading by providing hands-on lessons in the four major pillars of data analysis and data stewardship: (FAIR) data stewardship, statistics, data analysis and machine learning, and disciplinary knowledge. The candidates chances of employment are increased in some of the educations, which coordinate training and internships within stakeholder sectors and match companies' job requirements and profiles to curricula and thereafter develop new training programs where needed (HA, DTL, TUD, CC, UP). Hence already during the education the candidate is placed in the position to assist with data management, learning how to connect within and across boundaries to identify synergies and how to build on them, and understanding how policies are implemented and the possibilities to inform them. Consequently, from early in the education, driving the establishment of data stewardship and FAIR data stewardship (LERU, NeIC) as a new profession.

Whereas the academic educations develop their curricula and internships in varying degrees of collaboration with industry, the corporate educations choose rather to employ industry leaders as instructors to ensure both in depth and up-to-date understanding of the skills needed to apply concepts,

techniques, and practices of data stewardship, data quality, data governance, metadata management, and master data management (CDS, ISACA)

However the mission of both the academic and the corporate educations are the same - to establish and formalize the role of data stewardship and to drive recognition of Data Steward as a professional designation. Where they diverge is in how actively the certification of the candidate is employed. In the corporate world, the currency of the certification has a direct impact on the employability of the candidate in their chosen field. Some organizations start their data stewardship programs with IT-focused data stewards and train business unit staff members to assume the role (IBM, HP, DELL, QU, PC); other organizations believe that starting the data steward function in the business areas is the best route (ANA, CDS, DV, DMU). This is a cultural issue, and each education must determine how it will approach data stewardship according to the culture of its affiliated organizations while respecting the validity of the industry best practices and the need for sustainability of the data stewardship function. Accordingly, the majority of the corporate educations require the renewal of skills and that certification is recognized across a stakeholder community made up of industry partners, societies, councils and universities (CDS, IBM, ESIP, ANA). Thus signalling to employers that the employee is dedicated to the data stewardship profession, and additionally the certification is used by employers for level-setting teams of employees (DMU, DV).

5. Recommendations

Based on the analysis in the previous section, we recommend the following five considerations could be beneficial in the future development of Data steward education.

Collaborate

Universities and industry that unite to develop curricula and course modules are able to facilitate a sustainable collaboration with disciplinary actors and other stakeholders, including the data stewardship community. This collaboration creates alignment between the education and the job market, ensure the relevancy of learning outcomes and production of competent (employable) data stewards. Collaboration can take the form of a) an advisory panel to course structure and content or b) as contributors to course content, for example as instructors, cases for short term projects and offer internships. Invitins industry leaders as guest teachers ensures update knowledge on data stewardship practice and skills in how to connect across boundaries. Both approaches are vital for upholding the relevance of the course, keeping the course up-to-date with the evolving needs of data users, identifying challenges and specific training needs and more.

Further, collaboration strengthens the recognition of the education i.e. the relevance of the academic, and practical abilities of the candidate, as well as their disciplinary knowledge in partner organisations. Thus serving as a validation tool for the education, the learning aims and the extent the skills learnt on during the education can be put into practice. Collaboration informs the students academic abilities, disciplinary knowledge, professional experience and competencies as a data-information professional.

Student profile and course content

The majority of the educations analysed in this report have been marketed as post-graduate opportunities. The outset are persons with well-established disciplinary knowledge and experience on a professional level, preferably with existing data and programing skills. Course content is mediated through in-house training as data stewards at other faculties or in local businesses, where the disciplinary effects of data management are taught. As the student is mature, skills, competencies and knowledge exchange is

facilitated are only to a limited extend taught in the traditional classroom, favouring instead online, on demand modules, seminars and boot-camps with the aim to encourage professionalism and self-regulated learning, networking.

To establish an education for undergraduates requires a different approach. A flexible study programme, that frame-sets the individual course within relevant professional perspectives is recommended, with activities that support experiential learning. The teacher (as an instructor and supervisor), course content and employability inform each other and stimulate the undergraduate student while maintaining a structure the younger student can rely-on for guidance and grounding.

The interaction between teacher to teacher who are course responsibles must be extended to include interaction between teacher and professionals. Professional interaction determines the currency of the teachers knowledge be it interaction with research, methods and skills, mediation of developments in the field or engaging with professionals at the relevant workplaces (Troelsen and Tofteskov, 2013).

The form the content is presented to the students in should encourage interaction and motivate the student's own perspectives in a problem field. Typical forms could be case-teaching, problemsolving makerspace, internships or poster sessions. Students need to experience, not just learn about, how the presented theories, methods and skills are relatable to current workplace practices and challenges.

The variable of employability is an essential consideration in course development. Employability encourages the student's motivation, and how organised and prepared they are for class and the maturity in which they will interact with learning activities. Overall, the term employability is seen as the inducement for the degree of motivation, organisedness and prepardness.

Skills and competencies

Choice is a major factor for student motivation to learn data steawardship and yet we have to carefully consider how and when choice is presented to the student (Evans and Boucher, 2015) to ensure that all participants acquire a basic skill set on which they can build on according to their interests. The student is encouraged to explore data requirements and potentials across a range of users, hence the teacher must have the knowledge and trust to combine the strong model with the weaker model, allowing the student to take the lead in developing their interests and learning and ultimately their professional identity. Professional identity steers the content of data stewardship education in the corporate context (data steward as a data manager, as a support function or as data governor, p. 13). However the following mandatory skills were identified across academic and corporate educations in the analysis in section 4:

- Technological expertise: programming, analytics, data integration.

- Value for end users: business intelligence, customer insights, protection of information assets, policy and strategy development & implementation, research support & research data lifecycle, ability to leverage data for maximum value.

- Data management, FAIR data management, data governance, data security, ethical and legal data stewardship. Understanding of Data Stewardship as a profession, the function of data stewards, ability to assess the effect of data stewardship within and outside an organization.

- Communication, teaching, liaison skills, knowledge and contact with surrounding community.

- Disciplinary knowledge: specialization in specific area of data stewardship. This could be as broad as choosing a research or corporate direction or investigating the research and corporate requirements to stewardship within a specific field, fx the life sciences.

- Project management and the ability to work in multidisciplinary teams.

Certification
The analysis emphasised the importance of peer recognition of the certification. By collaborating with industry partners the university education has the potential to receive the seal of approval in the surrounding community through association accreditation and consequently increase the visibility of candidate for potential employers.

In the innovative and fast moving field of data stewardship a one time certification in not sufficient to ensure the skill set and knowledge of the candidate is up-to-date. Renewal of skills is recommended, and an obligation to the candidate to maintain and develop their knowledge through active participation in networks and training events. Commitment to skill maintenance also is an opportunity to network, facilitate connections, and define levels of proficiency (as because candidates will have different degrees of knowledge about different disciplines, and acquire different skill sets throughout the education). Thus the institution is advised to consider further education, as workshiops, online training programmes, or short courses in collaboration with local partners from industry, professional and trade unions.

Employability
The candidate's chances of employment are increased in educations that coordinate training and internships with stakeholder sectors and match job and policy requirements to curricula. The administrative challenge in course and curricula development is the speed in which the teacher can implement changes and their (economical) freedom to include guest teachers and innovative teaching and learning pedagogies.

6. Conclusions

This paper reviewed 24 data steward educations, eight from academia and sixteen from the corporate world. The differences and commonalities between the academic and corporate Data Stewardship educations were discussed. Mission of academic and corporate educations proved to be the same - to establish and formalize the role of data stewardship and to drive recognition of data steward as a professional designation. Where the educations diverge is how actively the certification (skill set of the candidate) is monitored, maintained and updated. Accordingly, there is also a need to supplement data steward education and work actively with developing adult learning, further education, and the future role of data management, security and enterprise, thus pledging dedication to continued support to the data stewardship profession.

The conceptual analysis illustrates that the structure and content of such educations have a lot to do with how data stewardship is perceived within and across academia, research fields, industry and enterprise. The majority of the academic data stewardship educations were on a post graduate level and bound to

research support and the research data lifecycle. The application of data stewardship in the corporate educations was more diverse and the learning aims can be grouped into three perspectives of data stewardship: data manager, data support or data governance. Each perspective includes different learning aims and skills and the ultimate role of the data steward in the enterprise, thus informing the content of the education and the availability of the education to students. Students wishing to learn data stewardship, both on the academic and corporate tracks, require at the present time certain skills, professional roles and disciplinary knowledge to apply for an education. As a result, the existing educations are somewhat closed to a wider audience.

The lack of educational opportunities for undergraduates present a golden opportunity for the development of innovative and relevant educations that could educate dedicated data steward professions. Learning from the Netherlands, we can see that strong academic educations can be developed in collaboration with industry. First, to broaden the academic view of the applications and the role of the data steward, second to expand course content to include more than research data management, third, to be inspired from corporate educations and support data stewardship as a managerial role, where analytics, policy awareness and business intelligence are essential, fourth support experiential learning to motivate students, and finally, to increase the employability of the graduate.

7. Limitations

The conceptual analysis is dependent on the textual quality of the description of the education. The analysis does not allow us to conclude if the appropriate skills are being taught on the education or if the programme is successfully embedded in the wider context of the educational institution or governmental policies on data stewardship education. The conceptual analysis does however give us the possibility to systematically understand the current status of available data steward educations and the diverse approaches to what such educations contain.

8. References

Albrecht, J. (2016). How the gdpr will change the world. European Data Protection Law Review (EDPL), 2(3), 287-289.

Ayris, P., Román, A. L. de S., Maes, K., & Labastida, I. (2018). Open Science and its role in universities (Advice Paper No. 24). Leuven: LERU.

European Commission, & Directorate-General for Research and Innovation. (2016). Realising the European open science cloud: first report and recommendations of the Commission high level expert group on the European open science cloud. Retrieved from http://bookshop.europa.eu/uri?target=EUB:NOTICE:KI0116872:EN:HTML

European Commission Expert Group on FAIR Data. (2018). Turning FAIR into reality (p. 78) [Final Report and Action Plan]. Retrieved from Directorate General for Research and Innovation website: doi: 10.2777/1524

Evans, M., and Boucher, A.R. (2015). Optimizing the Power of Choice: Supporting Student Autonomy to Foster Motivation and Engagement in Learning. *Mind, Brain and Education,* vol.9(2):pp.87-91.

Hsieh, H.-F., & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. Qualitative Health Research, 15(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Jensen, K. K., Whiteley, L., & Sandøe, P. (Eds.). (2018). RCR - a Danish textbook for courses in Responsible Conduct of Research (3rd ed.). University of Copenhagen, Department of Food and Resource Economics.

Kurepati, S. (2019, April 30). Becoming a data steward. Retrieved 16 July 2019, from Impact of Social Sciences website: https://blogs.lse.ac.uk/impactofsocialsciences/2019/04/30/becoming-a-data-steward/

Teng, J., & McGrail, K. (2018). Cultural and institutional barriers among data stewards regarding data access for research. International Journal of Population Data Science, 3(4). https://doi.org/10.23889/ijpds.v3i4.958

Teperek, M., Cruz, M. J., Verbakel, E., Böhmer, J. K., & Dunning, A. (2018, January 22). Data Stewardship – addressing disciplinary data management needs. https://doi.org/10.31219/osf.io/5w9pj

Troelsen, R., and Tofteskov, J. (2013). Underviserrollen i: Rienecker. L., Jørgensen, P.S., Dolin, J & Ingerslev, G.H. (Eds), *Universitetspædagogik*, s.201-214. København: Samfundslitteratur

Versweyveld, L. (2016). We need 500.000 respected data stewards to operate the European Open Science Cloud. Retrieved 16 July 2019, from http://e-irg.eu/news-blog/-/blogs/we-need-500-000-respected-data-stewards-to-operate-the-european-open-science-cloud

9. Appendices

Appendix 1: Search strings
Appendix 2: Description and abbreviation of educations included in the review

Appendix 1: search strings

education "data stewardship" OR "data steward" 251 results
12 relevant

course "data stewardship" OR "data steward" (google) 161 results
12 relevant

kursus data stewardship" OR "data steward" 10 results
3 relevant

kurser data stewardship" OR "data steward" 12 results

2 relevant

uddannelse "data stewardship" 40 resultater
1 relevant

uddannelse "data steward" 59 resutlater
2 relevant

training "data steward" OR "data stewardship" (google) 177
3 relevant

Appendix 2

Description and abbreviation of educations included in the review

Education	Institution	Link	Type	Abbr.
Data analysis and Stewardship	Helis Academy	https://helisacademy.com/en/data-analysis-stewardship	academic	*HA*
DAT 511 Data Stewardship: Preparation, Exploration and Handling of Big Data	Canisius College	https://catalog.canisius.edu/graduate/courses/dat/	academic	*CC*
DTL course directory	Dutch TechCenter for Life Science	https://www.dtls.nl/?s=data+stewardship&ct=course	academic	*DTL*
FAIR data stewardship: supporting FAIR interoperability	NeIC	https://www.deic.dk/en/fair-data-stewardship-supporting-fair-data-interoperability	academic	*NeIC*
Information culture and data stewardship	University of Pittsburgh	http://www.icds.pitt.edu/	academic	*UP*
Doctoral summer school on Data Stewardship	LERU	https://www.leru.org/news/leru-doctoral-summer-school-on-data-stewardship	academic	*LERU*
Training for data stewards	TU Delft	https://openworking.wordpress.com/2017/09/18/training-for-data-stewards/	academic	*TUD*
Data Science, Msc	University of Southern Denmark	https://www.sdu.dk:443/da/uddannelse/kandidat/datascience/karriere	academic	*USD*
Certified data steward program	eLearning Curve	https://ecm.elearningcurve.com/Certified_Data_Steward_Program_CDS_s/136.htm	corporate	*CDS*
Course 2M618g: IBM stewardship center for Information Server v11.5 SPVC	International Business Machines	https://www-03.ibm.com/services/learning/ites.wss/zz-	corporate	*IBM*
Data stewardship short course team	Earth Science Information Partners	https://www.esipfed.org/data-stewardship-short-course-team-course-outline	corporate	*ESIP*

Data stewardship training and certification	Information Asset	http://information-asset.com/index.php/featured-content/data-stewardship-training-and-certification/	corporate	*IA*
DSG 201 developing a strategy for data governance and stewardship	Data Management University, EW solutions	https://www.ewsolutions.com/foundations-data-stewardship/	corporate	*DMU*
DSG 301 data stewardship training	Data Management University, EW solutions	https://www.ewsolutions.com/wp-content/uploads/2016/09/20170113-EWSolutions-Course-Catalog.pdf	corporate	*DMU*
DSG 401 enterprise data stewardship and governance: full life-cycle roadmap	Data Management University, EW solutions	https://www.ewsolutions.com/wp-content/uploads/2016/09/20170113-EWSolutions-Course-Catalog.pdf	corporate	*DMU*
Certified ANA Marketing Professional Program	Association of National Advertisers	https://www.ana.net/schoolofmarketing	corporate	*ANA*
Training courses BIG DATA	aQUrate	https://www.aqurate.net/training-registration	corporate	*QU*
One-day non-technical course in data stewardship	Datavault (Business Thinking Ltd)	https://www.data-vault.co.uk/qualities-data-steward/	corporate	*DV*
Data Stewardship	Collibra University	https://university.collibra.com/knowledge/collibra-body-of-knowledge/data-stewardship/	corporate	*CU*
FAIR data stewardship and management training	Phortos Consultants	http://www.phortosconsultants.com/Fair-management	corporate	*PC*
EMC Data Protection and Management Certification	DELL	http://infotime.site/2017/06/19/dell-emc-proven-professional-certification-guide-oversigt-og-karriereveje/	corporate	*DELL*
Information Governance Professional (IGP) certification program	Association of Records Managers and Administrators	https://www.arma.org/page/Certifications	corporate	*ARMA*
ISACA	Information Systems Audit and Control	https://www.isaca.org/CERTIFICATION/Pages/default.aspx	corporate	*ISACA*

	Association			
HP accredited technical professional ATP	Hewlett Packard	https://certification-learning.hpe.com/tr/certification_skill_levels.html	corporate	*HP*