Towards a Standardized Data Science Competence Framework: A Literature Review Approach

Maria Potanin, Maike Holtkemper, Tobias Golz and Christian Beecks
Department of Data Science, University of Hagen, Universitätsstrasse 11, 58097 Hagen, Germany

Keywords: Data Science, Competence, Framework, Curricula.

Abstract: Described as the "sexiest job of the 21st century", the data scientist profession has attracted a lot of attention and demand over the past decade. The rapid growth of this profession, coupled with high barriers to entry and a lack of standardization, has led to challenges in defining required competencies. This study examines various frameworks and curricula that aim to teach essential data science competencies, with a focus on the needs of the industry. A systematic literature review resulted in 32 relevant articles of which 12 documents were analysed using a qualitative content analysis to synthesise the existing knowledge and integrate it into a unified competence framework based on the EDISON Data Science Framework. The results provide a comprehensive overview of the current relevant literature and propose a grouping of competencies, their knowledge and skills based on current research findings. These findings will be presented transparently to different users from teaching, training and resource planning practice through the visualisation of different levels in a web application. This work serves as a foundation for future research efforts aimed at improving the effectiveness and relevance of data science curricula and frameworks.

1 INTRODUCTION

According to an article in the Harvard Business Review, the profession of data scientist received great attention when it was named the "Sexiest Job of the 21st Century" (Davenport and Patril, 2012, p.1). At this point, the role of a data scientist was still quite new and in high demand from both technology companies and start-ups (Davenport and Patril, 2022). Ten years later, the follow-up article "Is data scientist Still the Sexiest Job of the 21st Century?" stated that the demand for qualified workers in this field was higher than ever (Davenport and Patril, 2022, p.1). The skills required to use the latest digital technologies to extract valuable knowledge from heterogeneous data sources are a major part of the fascination of this profession (Vaast and Pinsonneault, 2021). In particular, the development of technology, combined with the expansion of the field over the past decade, has led to a proliferation of data scientists who specialise in specific areas of expertise, rather than working as generalists (Davenport and Patril, 2022).

Alongside the growing demand for new technologies, the shortage of data-oriented professionals related to Data Science is also continuing to grow. In addition to demographic change (Grenčíková et al., 2022), high barriers to entry and a lack of standardization in the field (Fayyad and Hamutcu, 2020) are possible causes. Definitions, required skills and differentiation from other related professions are not clear (National Academies of Sciences, Engineering, and Medicine, 2018; Pompa and Burke, 2017). Companies or other institutions must carefully examine which roles with the appropriate knowledge and skill profiles are required for their large data science projects and how the Data Scientist's qualifications fit into one of these roles (Davenport and Patril, 2022; Holtkemper et al., 2024). In addition, intelligent solutions to support or even completely fulfill data science tasks have been developed (Necula, 2023). This raises the question of whether technologies such as automation frameworks (Auth et al., 2019; Potanin et al., 2023; Potanin et al., 2024) will persist or could potentially make the data scientist profession obsolete (Vaast and Pinsonneault, 2021). According to Manyika et al. (Manyika et al., 2017), time-consuming tasks such as data processing or data collecting, have the greatest potential for automation. Therefore, Potanin et al. (Potanin et al., 2023) analyzed the impact of automation frameworks on today's data science competencies by investigating whether modern automation frameworks have data science competencies in their design.

Educational institutions are facing significant challenges in keeping up with the rapid changes and the complex demands of modern work environments,

Potanin, M., Holtkemper, M., Golz, T. and Beecks, C.
Towards a Standardized Data Science Competence Framework: A Literature Review Approach.
DOI: 10.5220/0013364200003932

In *Proceedings of the 17th International Conference on Computer Supported Education (CSEDU 2025) - Volume 2*, pages 569-581
ISBN: 978-989-758-746-7; ISSN: 2184-5026

as evidenced by the findings of Pompa (Pompa and Burke, 2017). A study by Wu (Wu, 2017) revealed considerable variation in job titles and required qualifications for data scientists across different employers. The interdisciplinary nature of data science, which combines elements of computer science, statistics, mathematics, and domain-specific knowledge, has resulted in different fields developing their own definitions of data science (National Academies of Sciences, Engineering, and Medicine, 2018). The lack of standardization in both education and practice has resulted in various frameworks and curricula, some of which differ significantly in content (Schmitt et al., 2023). The lack of clarity on this matter makes it difficult to devise a curriculum that is fit for purpose (Schmitt et al., 2023).

The aim of this study is to investigate different frameworks or curricula that focus on essential competencies in data science and to integrate them into a unified competence framework. Establishing a consensus on the essential skills and knowledge enables intentional design of educational research in data science. This clarity will benefit the industry by providing a clearer understanding of the competence profile required for a data scientist and other roles in a data science project. We therefore pose the following research question: *How can competencies from existing curricula and frameworks in data science be merged into an integrated competence framework?*

Using Webster and Watson's systematic literature review (Webster and Watson, 2002), 32 relevant documents out of 209 were identified and evaluated for their suitability for this study based on various quality characteristics. Twelve selected frameworks were analysed using qualitative content analysis according to Kuckartz (Kuckartz and Rädiker, 2022). As a result of the literature synthesis, we summarized the collected findings in an integrated competence framework based on the EDISON Data Science Competence Framework (CF-DS) (Demchenko and José, 2021). In contrast to a previous study (Schmitt et al., 2023), which identified similarities and inconsistencies of the EDISON CF-DS compared to five other frameworks or curricula, we focus on adding the inconsistencies and unnamed elements (knowledge and skills) to get a comprehensive view of possible facets of the domain. We extend the EDISON framework with new competence groups, map the findings from the EDISON CF-DS to the new structure and add competencies, knowledge, and skills from other literature-based data sources.

This article is organized as follows. Section 2 provides the theoretical background. Section 3 presents an overview of the research methods used to answer the research questions and the literature review process. Section 4 presents the results of the literature review and synthesis, followed by a discussion in Section 5. The paper ends with the conclusion and future work in Section 6.

2 BACKGROUND

2.1 Competence Framework Terminology

Competence frameworks have become increasingly common over the last few decades and a grown number of domains use them (Mills et al., 2020). They can be understood as frames with an organized collection of related competency statements, for which the literature of different domains seems to agree (Mills et al., 2020). However, it quickly becomes apparent that published competence frameworks differ in their use of the term "competency" or "competence". Mills et al. (Mills et al., 2020, p.4) attribute this to two dominant conceptualisations: the educational emergence of "behavioural competency frameworks" and, about a decade later, the employer-driven "functional competence frameworks". In the behavioural approach, "competency" is defined by the concept of cumulative attributes and refers to performance in general and not to a specific profession or task (Chen and Chang, 2010). In contrast, functional competence frameworks have been developed to describe performance expectations for specific professions and define standards, whereby a standard was either achieved or not (Boritz and C., 2003; Bruno et al., 2010; Rowe, 1995). Over time, various frameworks have been published that merge behavioral and functional approaches, leading to coherence in the concepts and motivation of the framework (Boritz and C., 2003; Mills et al., 2020; Shippmann et al., 2000; Lambert et al., 2014; Russ-Eft, 1995). "Competency" is often used to refer to how performance occurs (behavioural approach), and "competence" to refer to how performance occurs (functional approach), but this has not been consistent (Thistlethwaite et al., 2014; Bruno et al., 2010; Davis et al., 2008; McLagan, 1997; Stuart et al., 1995). Mills et al. (Mills et al., 2020) proposes a glossary for competency frameworks in health, which currently does not exist for business-related frameworks. It can be argued that data science projects require the current state of competencies to conduct risk assessments, which Mills et al. (Mills et al., 2020) argues is equivalent to the functional approach. However, project or HR managers, and other stakeholders in the field are also interested in the longterm development of skills. For this reason, the terminology ”competence” (in the plural form ”competencies”) of the EDISON Competence Framework is used in this paper, as it represents the basis for modification.

2.2 EDISON Data Science Framework

The EDISON Data Science Framework (EDSF) encompasses all the outcomes of the EDISON project, which ran from 2015 to 2017 (EDISON Initiative, 2023). It was dedicated to advancing the establishment of the novel profession of ’data scientist’ and aimed to define the required competencies, formulate a framework/profile of skills, develop a body of knowledge and design a model curriculum (European Commission, 2017). Within this framework, the CF-DS identifies the relevant competencies of the data science domain and links them to knowledge and skills. In addition, EDISON aimed to create a sustainable business model to ensure the continuous growth of data scientists coming from universities and trained by various professional education and training institutions across Europe (European Commission, 2017). The definition of competence is outlined in the European e-Competence Framework (European Committee for Standardization, 2014), which is also included in the CF-DS (Demchenko and José, 2021). A competence ”is a demonstrated ability to apply knowledge, skills and attitudes for achieving observable results” (European Committee for Standardization, 2014, p.5). Within the CF-DS, this concept is operationalized by associating several types of knowledge and skills with a given competence, which is used as a broad term (Demchenko et al., 2022b). Knowledge is ”information often acquired through formal education, books, or other media” (Fayyad and Hamutcu, 2020, p.10). In short, knowledge is the theoretical component required for the exercise of competence. Skills can be described as ”the ability to apply this knowledge, often gained through practice.” (Fayyad and Hamutcu, 2020, p.10). The EDISON concept consists of four dimensions:

- Competence Areas: This dimension organizes competencies into five groups: Data Analytics (DSDA), Data Engineering (DSENG), Data Management and Governance (DSDM), Research Methods and Project Management (DSRMP), and Domain-specific Knowledge (DSDK), applied to Business Analytics.
- Generic Description: This dimension further develops the competence groups based on the specific domain where data science is applied, highlighting industry-specific requirements and challenges.
- Proficiency Level: This focuses on the roles and responsibilities within data science teams, defining tasks and skills required for each role.
- Knowledge and Skills: Competencies are divided into methodological, technological, and soft skills, each supporting different aspects of data science work.

2.3 Related Work

In recent years, data science has witnessed the development of numerous competence frameworks aimed at guiding the design of educational programs (Bile Hassan et al., 2021; Danyluk and Leidig, 2021; Pompa and Burke, 2017; Ramamurthy, 2016; Rosenthal and Chung, 2020; Salloum et al., 2021; Veaux et al., 2017). The research landscape within the field is constantly evolving. Since the introduction of the EDSF as a central publication of the European Union, numerous additional frameworks have emerged, such as publications by the National Academies of Sciences, Engineering, and Medicine (National Academies of Sciences, Engineering, and Medicine, 2018), the Harvard Data Science Review (Fayyad and Hamutcu, 2020), and the United Nations (UN Global Working Group,). Of particular importance is the Association for Computing Machinery (ACM) Data Science Curriculum (Danyluk and Leidig, 2021), which delves deeply into the field of data science.

In their research paper ”Evaluation of EDISON’s data science competence framework through a comparative literature analysis” Schmitt et al. (Schmitt et al., 2023) compared the CF-DS with five other competence framework approaches. The study carries out an exemplary comparison at the curriculum level and also evaluates several introductory data science courses. The results of this study show the differences between the existing curricula and the different priorities in the education of data scientists. While the authors provide valuable insights into the differences in content between the various competence frameworks and teaching programs, they do not present a modified framework. Urs and Minhaj (Urs and Minhaj, 2023) carried out an analysis of the various data science programs offered by educational institutions, in which they catalogued the topics that were covered. They were categorized into clusters based on their titles, and their frequency across individual courses was assessed. The results indicated that certain disciplines, notably machine learning and databases, were included in almost all instructional programs, while others, such as software development or domain-specific skills, were taught comparatively infrequently. The recorded course topics were then mapped to the knowledge areas outlined in the ACM Computer Science Curricula of 2013 (The Joint Task Force on Computing Curricula, 2013). This analysis revealed that certain knowledge areas were pervasive across most teaching programs, while others were completely absent. As the studies above show, institutions vary greatly in how they design data science programs. While they are working to define the competencies students should develop, their approaches and priorities differ, resulting in a lack of a unified competence framework for the industry, such as for job advertisements (Suryan and Gupta, 2021).

3 METHODOLOGY

3.1 Literature Review

In accordance with the recommendations set by Webster and Watson (Webster and Watson, 2002), a comprehensive literature review is conducted, which involves keyword and backward searches. The methodology outlined by vom Brocke et al. (vom Brocke et al., 2009) was employed to document the process in a systematic manner. An overview of the literature review process is presented in Figure 1. The investigation is centred on the various frameworks and curricula that are currently in use within the field of data science. A comprehensive search was initially conducted across a range of scientific databases, including the ACM Digital Library, IEEE Xplore, and Science Direct. Predefined search queries were used, including *"Data Science Curricula"*, *"Data Science Competency"* and *"Data Science Framework"* (see Table 1).

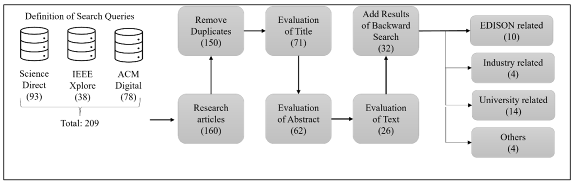

Figure 1: Literature Review Process.

During the literature review process, peripheral topics such as statistics were not investigated further, given that the major focus of the study is on the field of data science. Initially, a total of 209 documents were identified, of which 160 were classified as research articles. Following the removal of duplicate entries, 71 unique documents were identified as being potentially relevant for subsequent analysis. The evaluation of abstracts led to the selection of 62 papers for comprehensive examination. Following an evaluation of the text, the final subset of 26 articles was reviewed, augmented by results obtained from a backward search. This resulted in a final pool of 32 articles. Of these, 10 publications related to the EDISON (E) project, 4 presented industry-relevant (I) competencies, 14 focused on university programs (U), and 4 fell into other (O) categorizations (see Figure 1; Table 2).

Table 1: Findings of Keyword Terms.

Database	Keyword	Total Results	Articles	Relevant
ACM	Data Science Curriculum	61	38	20
	Data Science Framework	15	15	13
	Data Science Competency	2	2	2
Total	ACM	78	55	35
IEEE Xplore	Data Science Curriculum	10	9	6
	Data Science Framework	25	25	16
	Data Science Competency	3	3	2
Total	IEEE Xplore	38	37	24
Science Direct	Data Science Curriculum	13	10	5
	Data Science Framework	69	48	4
	Data Science Competency	11	10	3
Total	Science Direct	93	68	12
Total		**209**	**160**	**71**

3.2 Quality Appraisal

Following the literature review process, a literature analysis was conducted. As part of this process, the results from the literature review were analyzed in more detail. To ensure the integrity of the selected literature, a comprehensive critical appraisal methodology was applied, following the framework outlined by Kitchenham (Kitchenham, 2004). Each article was subjected to a rigorous quality assessment utilizing predefined checklists, in order to facilitate its categorization into one of three tiers of relevance: low,

Table 2: Findings from Keyword (K) and Backward (B) Search.

Title	Year	Cat.	Type
Model Curricula for Data Science EDISON Data Science (Wiktorski et al., 2017)	2017	E	K
Customisable Data Science Educational Environment: From competencies Management and Curriculum Design to Virtual Labs On-Demand (Demchenko et al., 2017)	2017	E	K
Data Science Model Curriculum Implementation for Various Types of Big Data Infrastructure Courses (Wiktorski et al., 2019)	2019	E	K
EDISON Data Science Framework (EDSF) Extension to Address Transversal Skills Required by Emerging Industry 4.0 Transformation (Demchenko et al., 2019b)	2019	E	K
Designing Customisable Data Science Curriculum Using Ontology for Data Science competencies and Body of Knowledge (Demchenko et al., 2019a)	2019	E	K
Big Data Platforms and Tools for Data Analytics in the Data Science Engineering Curriculum (Demchenko, 2019)	2019	E	K
EDISON Data Science Competence Framework (CF-DS, Release 4) (Demchenko et al., 2022b)	2022	E	K
EDISON Data Science Framework (EDSF): Addressing Demand for Data Science and Analytics competencies for the Data Driven Digital Economy (Demchenko and José, 2021)	2021	E	K
Data Scientist Professional Revisited: competencies Definition and Assessment, Curriculum and Education Path Design (Demchenko et al., 2021)	2021	E	K
Data science in the business environment: Insight management for an Executive MBA (Lu, 2022)	2022	E	K
A Practical and Sustainable Model for Learning and Teaching Data Science (Ramamurthy, 2016)	2016	U	K
Systematic Study of Big Data Science and Analytics Programs (Wu, 2017)	2017	U	B
Curriculum Guidelines for Undergraduate Programs in Data Science (Veaux et al., 2017)	2017	U	B
Data Science for Undergraduates Opportunities and Options (National Academies of Sciences, Engineering, and Medicine, 2018)	2018	U	B
A Functional Approach to Data Science in CS1 (Dahlby Albright et al., 2018)	2018	U	K
A Data Science Major: Building Skills and Confidence (Rosenthal and Chung, 2020)	2020	U	K
Creating a Balanced Data Science Program (Adams, 2020)	2020	U	K
A CDIO Oriented Curriculum for Division of Data Science and Big Data Technologies: The Content, Process of Derivation, and Levels of Proficiency (Zhou et al., 2020)	2020	U	K
Data Science Curriculum Design: A Case Study (Bile Hassan et al., 2021)	2021	U	K
Developing an Interdisciplinary Data Science Program (Salloum et al., 2021)	2021	U	K
A Data-centric Computing Curriculum for a Data Science Major (Fekete et al., 2021)	2021	U	K
Exploring potential roles of academic libraries in undergraduate data science education curriculum development (Shao et al., 2021)	2021	U	K
Computing Competencies for Undergraduate Data Science Curricula (Force, 2021)	2021	U	B
Rankings vs Realities Exploring Competency Differences in Graduate Data Science Programs (Li et al., 2023)	2023	U	K
Data Science and Analytics Skills Shortage: Equipping the APEC Workforce with the Competencies Demanded by Employers (Pompa and Burke, 2017)	2017	I	B
Data Science Competency in Organisations: A Systematic Review and Unified Model (Hattingh et al., 2019)	2019	I	K
Toward Foundations for Data Science and Analytics: A Knowledge Framework for Professional Standards (Fayyad and Hamutcu, 2020)	2020	I	B
Investigating Academia-Industry Gap for Data Science Jobs and Curriculum (Suryan and Gupta, 2021)	2021	I	K
Teaching Computational Modeling in the Data Science Era (Giabbanelli and Mago, 2016)	2016	O	K
Cloud Computing Curriculum: Developing Exemplar Modules for General Course Inclusion (Adams, 2020)	2020	O	K
Establishing ABET Accreditation Criteria for Data Science (Blair et al., 2021)	2021	O	K
Integrated Data Science for Secondary Schools: Design and Assessment Course Inclusion (Schanzer et al., 2022)	2022	O	K
		Total: 32	

medium, or high (Nidhra et al., 2013):

1 Does the topical domain of the research paper align with the present research objectives?

2 Has a complete framework or curriculum been developed?

3 Does it constitute a study or further education program for data scientists?

4 Are the results relevant for the present study?

Prior to analysis, responses were categorized according to their alignment with the established quality criteria. Responses that met the quality criterion were assigned a weight of 1, those that partially met the quality criterion were assigned a weight of 0.5, and those that failed to meet the quality criterion were assigned a weight of 0. Studies were then classified as high-quality if their overall score exceeded 4, low-quality if their overall score fell below 1, and medium-quality if their overall score ranged between 1 and 3. In accordance with the established quality criteria for this study, 12 research paper were classified as high-quality studies and 21 as medium-quality.

3.3 Qualitative Content Analysis

The 12 research papers (see Table 4) were systematically analyzed using a qualitative and structured content analysis according to Kuckartz (Kuckartz and Rädiker, 2022). The objective of the analysis was to identify similarities and differences between the various literature sources. As the information was named and presented differently in the publications (e.g. tables or text), the first deductive coding process was undertaken to divide relevant text passages into competence groups (if available), competencies and descriptions based on the EDISON CF-DS scheme. Descriptions include statements on particular skills or knowledge as well as additional information on the respective competencies. The EDISON CF-DS was selected as the foundation for subsequent mapping of results within a modified framework. It is the most frequently cited framework in our literature search, and further developments have been released continuously. The table structure provides an overview of competence groups and competencies, and thus categories and subcategories were partly adopted in a deductive manner for the second coding process. The categories of application development, business analytics, theoretical foundations and soft skills were added inductively, with further sub-categories and sub-sub-categories. In the final stage of the process, all codes were checked according to the final coding scheme and reassigned if necessary. Table 3 illustrates an example of the *DSAD - Application Development* category from the coding scheme. In accordance with the recommendations of Kuckartz (Kuckartz and Rädiker, 2022), all documents were coded by two of the authors to enhance the rigor of the results. The overall agreement between the coded segments was high, although not identical. The elements that had been categorised differently were discussed with the addition of the third author and assigned to the category in which two of the three authors were in agreement.

Table 3: Final coding scheme for category “Application Development”.

Competence Group (Category)	Competence (Sub-category)	Description (Sub-sub-category)
DSAD - Application Development	DSAD01 - Programming	Programming skills, Development issues
	DSAD02 - Libraries and tools	Programming libraries & tools, Development environment
	DSAD03 - Software Engineering	Software Engineering principles
	DSAD04 - Development pipelines	Version control, Development pipelines (automation)

4 FINDINGS

4.1 Descriptive Results

As a first result, Table 4 represents the identified competence frameworks derived from the literature review and analysis. The selected documents were published between 2017 and 2022, with at least one framework (or release) being published each year. This highlights the necessity for a framework that is both current and applicable across a range of contexts. The research work has a strong academic focus (66%), even though some of the frameworks include practical domains. The EDISON CF-DS, for example, was evaluated by expert groups comprising representatives from academic and industrial sectors. Other publications (44%) such as Hattingh et al. (Hattingh et al., 2019) are clearly focused on practical applications in order to develop an industry-specific skill set. The comparison of data sources is therefore inclusive of both academic and practice-oriented documents. With the exception of one research paper

Table 4: Identified Frameworks for Qualitative Content Analysis.

ID	Title	Research Method	Data
1	EDISON Data Science Competence Framework (CF-DS) (Demchenko et al., 2022a)	Content analysis	Six existing frameworks;evaluated by expert groups (academia & industry)
2	Curriculum Guidelines for Undergraduate Programs in data science (Veaux et al., 2017)	Literature analysis	NSF Workshop on Data Science Education, guidelines for undergraduate majors in Mathematics, Statistics and Computer Science
3	Data Science and Analytics Skills Shortage (Pompa and Burke, 2017)	Literature review	Academic literature, research and technical papers, government reports, working papers, industry publications and surveys
4	Systematic Study of Big Data Science and Analytics Programs (Wu, 2017)	Literature review	DSA programs in the U.S.
5	National Academies of Sciences, Engineering, and Medicine (National Academies of Sciences, Engineering, and Medicine, 2018)	Content analysis	Information-gathering activities and community conversations
6	Data Science Competency in Organisations: A Systematic Review and Unified Model (Hattingh et al., 2019)	literature review	Literature on essential data science workforce competencies
7	Toward Foundations for Data Science and Analytics: A Knowledge Framework for Professional Standards (Fayyad and Hamutcu, 2020)	Literature review	Literature on analytics and data science skills
8	Computing Competencies for Undergraduate Data Science Curricula (ACM) (Danyluk and Leidig, 2021)	2 Surveys for Academia and Industry	Results of the surveys
9	Investigating Academia-Industry Gap for Data Science Jobs and Curriculum (Suryan and Gupta, 2021)	Literature review	28 curricula from Indian and Internat. institutions
10	A Data Science Major: Building Skills and Confidence (Rosenthal and Chung, 2020)	Following Jolly et al.'s trilogy	Based on data science guidelines such as EDISON, ACM, BHEF, NIST
11	Data Science Curriculum Design: A Case Study Bile (Bile Hassan et al., 2021)	Literature review	122 Data Science degrees in the U.S.
12	A Data-centric Computing Curriculum for a Data Science Major (Fekete et al., 2021)	Standardized structure of majors (University of Sydney)	ACM guideline

(Force, 2021), the authors primarily base their findings on qualitative data and use research methods such as literature reviews.

4.2 Framework Analysis and Modification

These 12 documents were coded as described in the methodology section. As a result, a total of 8 competence groups were identified, comprising 46 distinct competencies. These encompass 269 elements of knowledge, methodological and technological skills, and soft skills. Four of the five EDISON groups were kept, whereby three originally came from the NIST (NIST Big Data Public Working Group, Definitions and Taxonomies Subgroup, 2019) definition of a data scientist:

- **Data Analytics (DSDA).** Using appropriate analytical methods and statistical techniques on available data to discover correlations and support decision-making.
- **Data Engineering (DSENG).** Applying engineering principles and modern computing technologies to design and implement new data analytics applications, and developing tools, systems, and infrastructure to support data processing throughout the data life cycle.
- **Data Management and Governance (DSDM).** Development and implementation of a data management strategy for data collection, storage, preservation, and availability for further processing steps.
- **Research Methods and Project Management (DSRMP).** Creating new knowledge by applying scientific or similar engineering methods to generate knowledge and achieve research or organizational goals.

The fifth group of the EDISON CF-DS, *Domain-specific knowledge and expertise* was modified, as the addition of industry-related articles was designed to facilitate a heightened emphasis on practice. As a result, it was renamed **Business Analytics (DSBA)**.

During the comparison, we determined that more groups should be added to categorize all identified competencies in a logical manner. For example, the EDISON CF-DS excludes personal competencies from its overview, in constrast to other documents (Bile Hassan et al., 2021; Fekete et al., 2021; Hattingh et al., 2019; Rosenthal and Chung, 2020). For competencies related to software and application development (Hattingh et al., 2019), theoretical fundamentals such as algorithms and data structures, mathematics, computational fundamentals and complexity theory (Bile Hassan et al., 2021; Danyluk and Leidig, 2021; Fekete et al., 2021; Hattingh et al., 2019; Rosenthal and Chung, 2020), as well as personal competencies like problem solving, communication, and leadership (Bile Hassan et al., 2021; Danyluk and Leidig, 2021; Hattingh et al., 2019; Pompa and Burke, 2017; Veaux et al., 2017), new competence groups are added. They are defined as follows:

- **Application Development (DSAD).** Planning and development of software products using suitable programming languages and technologies, as well as implementation of development pipelines.
- **Theoretical Fundamentals (DSTF).** Proficiency in the relevant theoretical foundations for the application of data science principles.
- **Personal competencies.** Development of personal and social skills and attributes that are crucial for success in today's job market and especially as a data scientist.

In conclusion, the final categories of competence groups are as follows: *Data Analytics (DSDA), Data Engineering (DSENG), Data Management and Governance (DSDM), Research Methods and Project Management (DSRMP), Business Analytics (DSBA), Application Development (DSAD), Theoretical Fundamentals (DSTF), and Personal competencies.*

In the subsequent phase, the EDISON framework was modified in accordance with the aforementioned eight groups. For this purpose, new competencies were created (e.g. *Business and Organization* or *Information Needs* for the Business Analytics group) or existing ones were reassigned. For enhanced presentation and comprehension, the competencies were further subdivided when necessary. The entire framework with the literature comparison is accessible **here**. The structure of the modified framework is explained using the DSAD competence group sample in Figure 2.

The Application Development competence group is comprised of the following competencies: *Programming*, *Libraries and Tools*, *Software Engineering* and *Development pipelines*. In line with the recommendations of the e-Competence Framework (European Committee for Standardization, 2014), all elements are assigned an ID, in this case acronyms. To illustrate, the competence *DSAD01 - Programming* has been divided into *Programming* and *Development issues* (Table 4). The individual elements of the literature synthesis correspond to the definition of a competence, which includes knowledge and skills. In the case of skills, a distinction is made between methodological, technological and personal skills. Acronyms

Competence group	Elements (literature synthesis)				Competence	Framework Comparison (literature analysis)												Total	
	Knowledge	Skill-Method	Skill-Technology	Skill-Personal		EDSF	Park City (De Veaux et al.)	IADSS (Fayyad & Hamutcu)	Wu	ACM (Danyluk & Leidig)	APEC (Pompa et al.)	National Academies	Suryan & Gupta	Hassan et al	Fekete et al.	Rosenthal & Chung	Hattingh et al.		%
Application Development					**DSAD01 - Programming**														
	KDSAD01	SDSAD03	DSDALANG06 DSDALANG09 DSDEV01		Programming	x	x	x	x	x	x	x	x	x	x	x	x	12	100
	KDSAD08	SDSDM15 SDSAD09	x		Development issues	x	x	x	x	x		x						6	50
					DSAD02 - Libraries and tools														
	x	SDSAD04	DSDEV02 DSDEV08		Programming libraries & tools	x	x		x	x		x		x	x	x		8	67
	x	SDSAD08	DSDEV03		Development environment	x		x		x		x				x		5	42
					DSAD03 - Software Engineering														
	KDSAD03 KDSAD05 KDSAD07 KDSTF04	SDSAD01 SDSAD06	DSDEV07		Software Engineering principles	x		x	x	x	x		x	x			x	8	67
					DSAD04 - Development pipelines (DevOps)														0
	KDSAD02	SDSAD05	DSDEV04		Version control	x	x			x		x			x			5	42
	KDSAD04	SDSAD07	DSDEV05		Development pipelines (automation)	x			x			x			x			4	33

Figure 2: Framework modification and comparison for *DSAD - Application Development*.

are also assigned to these elements (excluding personal skills). The complete element catalogs for the acronyms can be accessed **here**.

The competence group *DSAD* with its four assigned competencies shows that the competence with the subgroup *Programming* in particular was named with a 100% agreement in every framework examined and thus contains a generally relevant competence for science and practice. The Development pipelines are less well represented. *Data Science Analytics (DSDA)* with five competencies and very detailed subgroups also shows four core aspects that were found in all documents: *Machine Learning Fundamentals*, *Visualize Results*, *Basic Data Analytics* and *Statistics*. More specific subgroups such as *Outlier Identification* (Danyluk and Leidig, 2021) or *Descriptive Data Analysis* (Demchenko et al., 2022b) were only mentioned once. Seven competencies were identified from the *Data Science Engineering* category, which can also be described in detail with up to six subgroups. High overlap in the literature can be found in the areas of *Data Collection* (75%) and *Data Preparation* (92%) of the *DSENG02* competence. *Relational and Non-relational Databases* (75%) can also be found in most sources. In general, the named competence groups were very well covered in the literature with overlaps of up to 100%.

The competence groups *Data Management, Research Methods, Project Management* and *Business Analytics* were less frequently mentioned. *Data Management* comprises six competencies with up to six subgroups. *Data Management Policy* was mentioned in nine documents, resulting in an overlap of 75%. The remaining subgroups range from 8% to 67%. Particularly detailed areas such as *Identify Data Sources* (Danyluk and Leidig, 2021) or *Data Interoperability* (Demchenko et al., 2022b) are mentioned only once. *Data Quality* is addressed in 7 documents (58%). The category *Research Methods and Project Management* also achieves the highest level of agreement in the use of *Research Methods* (58%), the *Data Life Cycle* (58%) and *Project Management* (67%). The modified *Business Analytics* competence group contains the most competencies. A total of 8 were identified, with overlap ranging from 8% to 42% within the literature. There is little overlapping in the competencies *DSBA02 (Fuzzy Concepts)* and *DSBA04 (Process Optimization)*, which, in addition to Demchenko et al. (Demchenko et al., 2022b), are only covered in the form of process optimization in Hassan et al. (Bile Hassan et al., 2021). *Business Analytics* and *Business Intelligence*, *Domain Knowledge* and (agile) *Decision Making* are addressed more frequently (42%).

Theoretical Fundamentals (DSTF) consist largely of knowledge items and focus on four competencies: *Algorithms & Data structures (DSTF01)*, *Mathematics (DSTF02)*, *Computational Fundamentals (DSTF03)* and *Complexity Theory (DSTF04)*. In this case, there are overlaps especially in *DSTF01* (92%) and *DSTF02* (83%). The competence group *Personal competencies* was not subdivided according to the scheme in Table 3. The competencies *Personal Improvement*, *Communication*, *Collaboration*, *Leadership*, *Problem Solving*, and *Ethical Thinking* were included with personal skills. Most of the competencies were identified in at least 8 documents.

4.3 Web Application

The modified competence framework is designed to be made accessible to different target groups to be used widely. In the first step, the framework is published via a web application for evaluation purposes and can be accessed via the **link**. The choice of an easily accessible, easy-to-understand and yet detailed form of presentation makes it possible to address different stakeholders from education and industry. It is possible to switch between three different views.

1. Competence List: Presents the competencies in tabular form, providing a quick overview. In each table, a competence group is presented, which includes all associated competencies. A filter allows the selection of all or specific competence groups.
2. Competence Grid: Competence groups are displayed colored tiles. By selecting a competence

group, the tiles of the next level (competencies) are displayed. On the next layer, the details of the selected competence are presented, which include information, assigned knowledge and skills.

5 DISCUSSION AND LIMITATIONS

Comparing different competence frameworks reveals some limitations that arise from several aspects. On the one hand, the frameworks have different objectives, such as Suryan's focus on the skills of a data scientist in industry (Suryan and Gupta, 2021) versus Rosenthal's focus on the university education of a data scientist (Rosenthal and Chung, 2020). In addition, the length of the individual research articles varies considerably, leading to differences in the scope and depth of detail of the competence descriptions, as in the case of EDISON CF-DS (176 pages) compared to Rosenthal's Data Science Major (6 pages). Table 5 shows the comparison of the different data sources.

Table 5: Document comparison on finding frequencies.

	EDSF	Park City	IADSS	Wu	ACM	APEC	National A.	Suryan	Hassan	Fekete	Rosenthal	Hattingh
Total	118	48	40	56	87	31	72	32	38	46	40	38
in %	87	36	30	41	64	23	53	24	28	34	30	28

A total of 135 possible competence subgroups were recorded. The EDISON CF-DS (87%) and the ACM (64%) achieved the most findings. The fact that EDISON was used as the basis for coding the documents may also explain the high degree of overlap. Concise frameworks typically address broad topics like statistics, while expansive frameworks delve into specific topics with greater detail, such as decision trees. In addition, many frameworks use literature analyses or qualitative content analyses as research methods. There is a lack of comprehensive practical research, such as case studies or quantitative surveys, which examine core elements of practice.

The discourse can be particularly focused on in the context of new technologies and the flexibility of skills frameworks. The literature examined does not show any explicit use of certain key technologies such as generative AI (GenAI) or large language models. In various areas with an industry focus, such as business process management, GenAI can be used to support automated routine tasks or the discovery of process innovations (Beverungen et al., 2021). The question therefore also arises as to how this technology can be used in data science projects (Feuerriegel et al., 2023; Zschech et al., 2020) and, associated with this, which competencies are required for the various roles in such a project. For example, the gap between modeling experts and domain users could be closed (Zschech et al., 2020). However, this leads us to the limitation that current frameworks are not adaptable enough and cannot react quickly to changes and new findings. Therefore, more and more new frameworks are being developed instead of updating old frameworks. The lack of standardization on a current and adapted framework makes it difficult for practitioners and educators to keep up with the rapid development of new technologies. The framework expanded here serves to present the status quo from the literature, which is intended to include both education and practice. It shows similarities and differences between individual frameworks and which competencies are hardly represented. For example, data quality plays an important role in the development of machine learning models, but is hardly considered in the various data science programs. In order to establish it as a useful tool and to make it more adaptable to technological developments and to the needs of teaching and practice, a discourse between the different target groups (teachers, students, practitioners) is necessary. As an initial evaluation approach, two expert interviews were conducted with data scientists in practice and a group discussion was held with three university lecturers. In summary, both the expert interviews and the group discussion highlighted the relevance of a standardized data science competence framework. There was general agreement on the content of the framework and no major gaps in the listed competencies were identified. Respondents from both sides gave positive feedback on the overlap with current educational programs and the daily work of a data scientist.

6 CONCLUSION

Described a decade ago as the most attractive profession of our time, the data science profession has grown due to increasing demand and technological advances (Davenport and Patril, 2012). However, this growth has also brought challenges, particularly in defining the necessary skills. While the development of various frameworks and curricula demonstrates efforts to address these issues, it has resulted in a lack of standardization and clarity.

Our study addresses this complexity and aims to integrate different frameworks and curricula into a unified competence framework. Through a systematic literature review and qualitative content analysis, we synthesized findings from different sources and built on the EDISON CF-DS. In contrast to previous studies, we not only identified similarities and inconsistencies, but also introduced new groups of competencies to enrich the understanding of the field of data science. Our integrated competence framework provides a comprehensive view of the essential competencies, knowledge and skills required for data scientists. Initial evaluations with experts and trainers have further validated and refined our findings. Our framework highlights the importance of standardization in defining data science competencies, which benefits both academia and industry. By providing a clearer understanding of the skill profile required for data scientists, our framework facilitates the targeting of educational programs and helps organizations identify and develop talent. However, further research is needed to refine and validate our framework as the field of data science continues to evolve rapidly. In summary, our study contributes to bridging the gap between theory and practice in data science education and serves as a foundation for future research efforts aimed at improving the effectiveness and relevance of data science curricula and frameworks. It also aims to highlight automation opportunities to address skills shortages, education and industry. Based on the framework presented, future research will investigate how automation frameworks can help automate specific data science tasks. Future research should investigate which competencies employees, especially domain experts, need to develop in order to operate these automation frameworks. Based on this approach, domain experts can be specifically trained without having all the data science knowledge.

REFERENCES

Adams, J. C. (2020). Creating a balanced data science program. In Zhang, J., Sherriff, M., Heckman, S., Cutter, P., and Monge, A., editors, *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, pages 185–191, New York, NY, USA. ACM.

Auth, G., Jokisch, O., and Dürk, C. (2019). Revisiting automated project management in the digital age – a survey of ai approaches. *Online Journal of Applied Knowledge Management*, 7(1):27–39.

Beverungen, D., Buijs, J., Becker, J., and et al. (2021). Seven paradoxes of business process management in a hyper-connected world. *Business & Information Systems Engineering*, 63:145–156.

Bile Hassan, I., Ghanem, T., Jacobson, D., Jin, S., Johnson, K., Sulieman, D., and Wei, W. (2021). Data science curriculum design: A case study. In Sherriff, M., Merkle, L. D., Cutter, P., Monge, A., and Sheard, J., editors, *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, pages 529–534, New York, NY, USA. ACM.

Blair, J. R. S., Jones, L., Leidig, P., Murray, S., Raj, R. K., and Romanowski, C. J. (2021). Establishing abet accreditation criteria for data science. In Sherriff, M., Merkle, L. D., Cutter, P., Monge, A., and Sheard, J., editors, *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, pages 535–540, New York, NY, USA. ACM.

Boritz, E. and C., C. (2003). Competency-based education and assessment for the accounting profession: a critical review. *Accounting Perspectives*, 2(1):7–42.

Bruno, A., Bates, I., Brock, T., and Anderson, C. (2010). Towards a global competency framework. *Am J Pharm Educ*, 74(3).

Chen, H. M. and Chang, W. Y. (2010). The essence of the competence concept: Adopting an organization's sustained competitive advantage viewpoint. *Journal of Management & Organization*, 16(05):677–699.

Dahlby Albright, S., Klinge, T. H., and Rebelsky, S. A. (2018). A functional approach to data science in cs1. In Barnes, T., Garcia, D., Hawthorne, E. K., and Pérez-Quiñones, M. A., editors, *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, pages 1035–1040, New York, NY, USA. ACM.

Danyluk, A. and Leidig, P. (2021). Computing competencies for undergraduate data science curricula.

Davenport, T. H. and Patril, D. J. (2012). Data scientist: The sexiest job of the 21st century. *Harvard Business Report*.

Davenport, T. H. and Patril, D. J. (2022). Is data scientist still the sexiest job of the 21st century? *Harvard Business Report*.

Davis, R., Turner, E., Hicks, D., and Tipson, M. (2008). Developing an integrated career and competency framework for diabetes nursing. *Journal of Clinical Nursing*, 17(2):168–174.

Demchenko, Y. (2019). Big data platforms and tools for data analytics in the data science engineering curriculum. In *Proceedings of the 2019 3rd International Conference on Cloud and Big Data Computing*, pages 60–64, New York, NY, USA. ACM.

Demchenko, Y., Belloum, A., de Laat, C., Loomis, C., Wiktorski, T., and Spekschoor, E. (2017). Customisable data science educational environment: From competences management and curriculum design to virtual labs on-demand. In *2017 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, pages 363–368. IEEE.

Demchenko, Y., Belloum, A., et al. (2022a). *Data Science Body of Knowledge (DS-BoK)*. EDISON Community Initiative.

Demchenko, Y., Belloum, A., et al. (2022b). *Data Science*

Competence Framework (CF-DS, Release 4). EDISON Community Initiative.

Demchenko, Y., Comminiello, L., and Reali, G. (2019a). Designing customisable data science curriculum using ontology for data science competences and body of knowledge. In *Proceedings of the 2019 International Conference on Big Data and Education*, pages 124–128, New York, NY, USA. ACM.

Demchenko, Y. and José, C. G. J. (2021). Edison data science framework (edsf): addressing demand for data science and analytics competences for the data driven digital economy. In IEEE, editor, *IEEE Global Engineering Education Conference (EDUCON)*, pages 1682–1687.

Demchenko, Y., Maijer, M., and Comminiello, L. (2021). Data scientist professional revisited: Competences definition and assessment, curriculum and education path design. In *2021 4th International Conference on Big Data and Education*, pages 52–62, New York, NY, USA. ACM.

Demchenko, Y., Wiktorski, T., Cuadrado Gallego, J., and Brewer, S. (2019b). Edison data science framework (edsf) extension to address transversal skills required by emerging industry 4.0 transformation. In *2019 15th International Conference on eScience (eScience)*, pages 553–559. IEEE.

EDISON Initiative (2023). Edison project.

European Commission (2017). Education for data intensive science to open new science frontiers.

European Committee for Standardization (2014). *European e-Competence Framework*.

Fayyad, U. and Hamutcu, H. (2020). Toward foundations for data science and analytics: A knowledge framework for professional standards. *Harvard Data Science Review*, 2(2).

Fekete, A., Kay, J., and Röhm, U. (2021). A data-centric computing curriculum for a data science major. In Sherriff, M., Merkle, L. D., Cutter, P., Monge, A., and Sheard, J., editors, *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, pages 865–871, New York, NY, USA. ACM.

Feuerriegel, S., Hartmann, J., Janiesch, C., and Zschech, P. (2023). Generative ai. *Business & Information Systems Engineering*, 66:1–16.

Force, A. D. S. T. (2021). *Computing competencies for undergraduate data science curricula*. ACM, New York, NY, USA.

Giabbanelli, P. J. and Mago, V. K. (2016). Teaching computational modeling in the data science era. *Procedia Computer Science*, 80:1968–1977.

Grenčíková, A., Habánik, J., Španková, J., Húzevka, M., and Šrámka, M. (2022). Current labour market challenges in the light of future economic and demographic developments. *European Journal of Interdisciplinary Studies*, 14(1).

Hattingh, M., Marshall, L., Holmner, M., and Naidoo, R. (2019). Data science competency in organisations. In de Villiers, C. and Smuts, H., editors, *Proceedings of the South African Institute of Computer Scientists and Information Technologists 2019*, pages 1–8, New York, NY, USA. ACM.

Holtkemper, M., Potanin, M., Oberst, A., and Beecks, C. (2024). Risk assessment of data science projects: A literature review on risk identification. In *Intelligent Systems Conference*, pages 217–231. Springer.

Kitchenham, B. (2004). Procedures for performing systematic reviews.

Kuckartz, U. and Rädiker, S. (2022). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung: Grundlagentexte Methoden*. Grundlagentexte Methoden. Beltz Juventa, Weinheim and Basel, 5. auflage edition.

Lambert, B., Plank, R. E., Reid, D. A., and Fleming, D. (2014). A competency model for entry level business-to-business services salespeople. *Services Marketing Quarterly*, 35(1):84–103.

Li, D., Milonas, E., and Zhang, Q. (2023). Rankings vs realities: Exploring competency differences in graduate data science programs. In *2023 IEEE Frontiers in Education Conference (FIE)*, pages 1–4. IEEE.

Lu, J. (2022). Data science in the business environment: Insight management for an executive mba. *The International Journal of Management Education*, 20(1):100588.

Manyika, J., Chui, M., Miremadi, M., Bughin, J., George, K., Willmott, P., and Dewhurst, M. (2017). A future that works: Automation, employment, and productivity.

McLagan, P. A. (1997). Competencies: moving into the next generation. *Training & Development*, 51(5):40.

Mills, J.-A., Middleton, J. W., Schafer, A., Fitzpatrick, S., Short, S., and Cieza, A. (2020). Proposing a reconceptualisation of competency framework terminology for health: a scoping review. *Human Resources for Health*, 18(1):15.

National Academies of Sciences, Engineering, and Medicine (2018). *Data science for undergraduates: Opportunities and options*. National Academies Press.

Necula, S.-C. (2023). Artificial intelligence impact on the labour force – searching for the analytical skills of the future software engineers.

Nidhra, S., Yanamadala, M., Afzal, W., and Torkar, R. (2013). Knowledge transfer challenges and mitigation strategies in global software development—a systematic literature review and industrial validation. *International Journal of Information Management*, 33(2):333–355.

NIST Big Data Public Working Group, Definitions and Taxonomies Subgroup (2019). Nist big data interoperability framework.

Pompa, C. and Burke, T. (2017). Data science and analytics skills shortage: equipping the apec workforce with the competencies demanded by employers.

Potanin, M., Holtkemper, M., and Beecks, C. (2023). The impact of automation frameworks on today's data science competencies. In *Human Interaction & Emerging Technologies (IHIET 2023): Artificial Intelligence & Future Applications*, AHFE International. AHFE International.

Potanin, M., Holtkemper, M., and Beecks, C. (2024). Exploring the integration of data science competencies in modern automation frameworks: Insights for workforce empowerment. In *Intelligent Systems Conference*, pages 232–240. Springer.

Ramamurthy, B. (2016). A practical and sustainable model for learning and teaching data science. In Alphonce, C., Tims, J., Caspersen, M., and Edwards, S., editors, *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, pages 169–174, New York, NY, USA. ACM.

Rosenthal, S. and Chung, T. (2020). A data science major. In Zhang, J., Sherriff, M., Heckman, S., Cutter, P., and Monge, A., editors, *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, pages 178–184, New York, NY, USA. ACM.

Rowe, C. J. (1995). Clarifying the use of competence and competency models in recruitment, assessment and staff development. *Industrial and Commercial Training*, 27:12–17.

Russ-Eft, D. (1995). Defining competencies: a critique. *Human Resource Development Quarterly*, 6(4):329–335.

Salloum, M., Jeske, D., Ma, W., Papalexakis, V., Shelton, C., Tsotras, V., and Zhou, S. (2021). Developing an interdisciplinary data science program. In Sherriff, M., Merkle, L. D., Cutter, P., Monge, A., and Sheard, J., editors, *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, pages 509–515, New York, NY, USA. ACM.

Schanzer, E., Pfenning, N., Denny, F., Dooman, S., Politz, J. G., Lerner, B. S., Fisler, K., and Krishnamurthi, S. (2022). Integrated data science for secondary schools. In Merkle, L., Doyle, M., Sheard, J., Soh, L.-K., and Dorn, B., editors, *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education*, pages 22–28, New York, NY, USA. ACM.

Schmitt, K. R. B., Clark, L., Kinnaird, K. M., Wertz, R. E. H., and Sandstede, B. (2023). Evaluation of edison's data science competency framework through a comparative literature analysis. *Foundations of Data Science*, 5(2):177–198.

Shao, G., Quintana, J. P., Zakharov, W., Purzer, S., and Kim, E. (2021). Exploring potential roles of academic libraries in undergraduate data science education curriculum development. *The Journal of Academic Librarianship*, 47(2):102320.

Shippmann, J., Ash, R., Batitsta, M., L., C., L.D, E., J., K., E.P., P., and J.I., S. (2000). The practice of competency modeling. *Personnel Psychology*, 53:703–740.

Stuart, R., Thompson, J. E., and Harrison, J. (1995). Translation: From generalizable to organization-specific competence frameworks. *Journal of Management Development*, 14(1):67–80.

Suryan, K. and Gupta, R. (2021). Investigating academia-industry gap for data science jobs and curriculum. In *2021 International Conference on Data Analytics for Business and Industry (ICDABI)*, pages 453–458. IEEE.

The Joint Task Force on Computing Curricula (2013). Computer science curricula 2013 - curriculum guidelines for undergraduate degree programs in computer science.

Thistlethwaite, J. E., Forman, D., Matthews, L. R., Rogers, G. D., Steketee, C., and Yassine, T. (2014). Competencies and frameworks in interprofessional education: a comparative analysis. *Academic Medicine*, 89(6):869–875.

UN Global Working Group. Competency framework for big data acquisition and processing.

Urs, S. R. and Minhaj, M. (2023). Evolution of data science and its education in ischools: An impressionistic study using curriculum analysis. *Journal of the Association for Information Science and Technology*.

Vaast, E. and Pinsonneault, A. (2021). When digital technologies enable and threaten occupational identity: The delicate balancing act of data scientists. *MIS Quarterly*, 45(3):1087–1112.

Veaux, R. D., Agarwal, M., Averett, M., Baumer, B., Bray, A., Bressoud, T., Bryant, L., Cheng, L., Francis, A., Gould, R., Kim, A. Y., Kretchmar, M., Lu, Q., Moskol, A., Nolan, D., Pelayo, R., Raleigh, S., Sethi, R. J., Sondjaja, M., Tiruviluamala, N., Uhlig, P., Washington, T., Wesley, C., White, D., and Ye, P. (2017). Curriculum guidelines for undergraduate programs in data science. *Annual Review of Statistics and Its Application*, 4(1):15–30.

vom Brocke, J., Simons, A., Niehaves, B., Reimer, K., Plattfaut, R., and Cleven, A., editors (2009). *Reconstructing the giant: On the importance of rigour in documenting the literature*, volume 161.

Webster, J. and Watson, R. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, 26(2).

Wiktorski, T., Demchenko, Y., and Belloum, A. (2017). Model curricula for data science edison data science framework. In *2017 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, pages 369–374. IEEE.

Wiktorski, T., Demchenko, Y., and Chertov, O. (2019). Data science model curriculum implementation for various types of big data infrastructure courses. In *2019 15th International Conference on eScience (eScience)*, pages 541–547. IEEE.

Wu, H. (2017). Systematic study of big data science and analytics programs. In *2017 ASEE Annual Conference & Exposition*.

Zhou, T., Jiang, D., Wang, F., Li, X., and Zheng, L. (2020). A cdio oriented curriculum for division of data science and big data technologies: The content, process of derivation, and levels of proficiency. In *2020 8th International Conference on Digital Home (ICDH)*, pages 172–177. IEEE.

Zschech, P., Horn, R., Hoeschele, D., C., J., and K., H. (2020). Intelligent user assistance for automated data mining method selection. *Business & Information Systems Engineering*, 62:227–247.