Journal of Statistics and Data Science Education

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ujse21

Data Skills for Everyone! (?)–An Approach to Assessing the Integration of Data Literacy and Data Science Competencies in Higher Education

André Coners, Benjamin Matthies, Carolin Vollenberg & Julian Koch

To cite this article: André Coners, Benjamin Matthies, Carolin Vollenberg & Julian Koch (13 May 2024): Data Skills for Everyone! (?)–An Approach to Assessing the Integration of Data Literacy and Data Science Competencies in Higher Education, Journal of Statistics and Data Science Education, DOI: 10.1080/26939169.2024.2334408

To link to this article: https://doi.org/10.1080/26939169.2024.2334408

JOURNAL OF STATISTICS AND DATA SCIENCE EDUCATION
2024, VOL. 00, NO. 0, 1–26
https://doi.org/10.1080/26939169.2024.2334408

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS

Check for updates

Data Skills for Everyone! (?) – An Approach to Assessing the Integration of Data Literacy and Data Science Competencies in Higher Education

André Coners[a], Benjamin Matthies[b], Carolin Vollenberg[a], and Julian Koch[a]

[a]Department of Technical Business Administration, South Westphalia University of Applied Sciences, Hagen, Germany; [b]Münster School of Business, Münster University of Applied Sciences, Münster, Germany

ABSTRACT
The proficient handling of data can undoubtedly be regarded as a key skill for the future. However, the need for data competencies is not limited to traditional professions in the information technology environment but is rather necessary across industries and work fields. Consequently, there is a call to integrate such Data Literacy and Data Science competencies into higher education teaching across the breadth of study programs. This study, descriptive in nature, sheds light for the first time on the status quo of this integration. For this purpose, a "Data Science Dictionary" has been developed, that structurally maps a corresponding curriculum of the *German Informatics Society* (GI). Using quantitative content analysis, more than 13,950 distinct modules from three German study programs (Business Administration, Business & Information Systems Engineering, Computer Science) at two different types of universities are examined. As a result of this comparative study, concise "teaching portfolios" are compared between disciplines, whereby the prevalence, proportion, and depth of the competencies integrated into the study programs become transparent. Thus, this approach can provide a basis for discourse on the future integration of data competencies into, for example, business study curricula; furthermore, it can track the resulting progress in a longitudinal study. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received July 2022
Accepted January 2024

KEYWORDS
Content analysis; Curriculum; Data literacy; Data science; Higher education

1. Introduction

Data is the "new oil" of the global economy (World Economic Forum 2019, p. 49). As a consequence, the ability to manage data is undoubtedly one of the key future skills required in both professional and private life (Kirchherr et al. 2019). The *World Economic Forum* (WEF) report "Data Science in the New Economy" (2019), for example, clearly emphasizes that Data Literacy and Data Science (hereafter referred to as DL and DS) competencies will be crucial for maintaining the competitiveness of economies in the future and that the need to develop these competencies is by no means limited to the typical information technology (IT) jobs, but is rather necessary across many industries, regions, and fields of work (Kirchherr et al. 2019).

The integration of DL and DS competencies in higher education is a central pillar of competitive economic strategy (GI 2018; Schwab-McCoy, Baker, and Gasper 2021). The *University Education Forum on Digitization* (2017), for example, describes the qualified handling of data as the central competence that should be taught by universities in the 21st century. Consequently, the profound integration of DL and DS content also makes current study programs more attractive to students (Schwab-McCoy, Baker, and Gasper 2021). However, as indicated above, the development of these competencies should not be limited only to traditional study programs in the field of IT and computer science. Instead, DL and DS education should be instituted across the breadth of higher education, that is, across domains, disciplines, and fields of study (University Education Forum on Digitization 2017; GI 2018; Kirchherr et al. 2019). In line with this, there is also frequent reference to a "democratization of [DS] education" (Kross et al. 2020). The *Future Skills Initiative*, a network of approximately 3000 leaders in business and education, addresses this need and has dedicated itself to promoting curricula that establish data skills as essential cross-cutting competencies across all fields of study (Initiative Future Skills 2019). Following this demand, the *German Informatics Society* (german: *Gesellschaft für Informatik*) has started to work on the precise formulation of the cornerstones of DL and DS education. First, in a policy paper (German Informatics Society 2018), the data competencies to be integrated into higher education were outlined. Building on this, a detailed proposal was made for a curriculum that incorporated DL and DS (German Informatics Society 2019).

The study described in this article addresses the request that the teaching of DL and DS competencies should be or rather must become, a deeply embedded topic in all fields of study. In this context, it is also essential to measure the corresponding progress, as emphasized by (Bichler, Heinzl, and van der Aalst 2017, p. 79): "It is important to reflect this development also

CONTACT Benjamin Matthies benjamin.matthies@fh-muenster.de Münster School of Business, Münster University of Applied Sciences, Corrensstr. 25, 48149 Münster, Germany.

Supplementary materials for this article are available online. Please go to www.tandfonline.com/ujse.

in our curricula." Accordingly, the first studies with this goal can be found in the recent literature (see Hardin et al. 2021; Schwab-McCoy, Baker, and Gasper 2021). In line with this, our study aims to develop a feasible assessment approach that describes (a) the distribution and (b) the intensity of integrating DL and DS competencies in higher education. The question of a corresponding analysis could be: To what extent are DL and DS competencies already embedded in the curricula of specific higher education disciplines, and how can this status quo be characterized? To elaborate a substantive study based on such a question, we examine module handbooks, an integral medium for documenting and communicating study programs and their curricula (see also Section 2). Specifically, we analyze a database of more than 13,950 individual module descriptions of bachelor courses from three study programs (Business Administration, Business & Information Systems Engineering, and Computer Science) as well as two types of universities (research universities and universities of applied sciences). The first differentiation (study programs) was chosen to distinguish a varying degree of integration according to the increasing degree of technical (IT) focus. Including two types of universities further broadened the insights and allowed for comparing results between the two groups. To analyze this extensive textual database, we applied a computer-aided, quantitative content analysis (Krippendorff 2019). The measurement tool developed for the analysis is a "Data Science Dictionary" (hereafter referred to as DSD), which maps the contents of a representative DS curriculum (EDISON Consortium 2018; GI German Informatics Society 2019) based on 12 thematic categories with significant keywords. The DSD is thereby a solution for measuring the presence of related topics in the examined module descriptions. In summary, the research question (RQ) of our study is as follows:

RQ: How can the status quo of integrating DL and DS competencies into higher education be measured and assessed, differentiated by:

(a) study programs (Business Administration, Business & Information Systems Engineering, and Computer Science) and
(b) type of university (research universities and universities of applied sciences)?

In accordance with the research design, the contribution of our study is 2-fold. The main contribution is that a robust approach is proposed for assessing the status quo of DL and DS competencies integration regarding distribution and intensity. These variables are also combined as dimensions in a concise teaching portfolio. Assessments based on this portfolio can enable targeted discussions for further development of such competencies taught in higher education, for example, business degree programs. In the future, discussing which data skills business graduates should have will be relevant and to what degree these skills are already included in the current study programs. The repeatability of the proposed measurement approach (DSD) also makes it possible to set up a long-term study on the gradual progress of integrating DL and DS competencies across fields of study. A further contribution is that exciting insights related to higher education research are provided by comparing types of universities and study programs, especially regarding the higher education landscape. Based on these contributions—and the fact that this descriptive study is the first of its kind—our conclusions include specific implications for higher education practice and research.

The article is structured as follows: Section 2 includes a more in-depth discussion of typical DL and DS curricula. In addition, the history and informative characteristics of module handbooks are explained, and the relevant background of higher education research is presented. Section 3 describes the research approach and the developed measurement instrument (DSD). Section 4 presents the results, followed by a discussion (Section 5) of their implications and limitations. The final Section 6 provides an outlook for further research.

2. Research Background

2.1. Data Science Curricula

In recent decades, there has been a shift toward competence-oriented teaching in higher education (Bergsmann et al. 2015). In this context, "competencies" can be generally understood as learning outcomes (Bergsmann et al. 2015). However, they can be described more specifically as "context-specific dispositions which are acquired and which are needed to cope successfully with domain-specific situations and tasks" (Blömeke, Zlatkin-Troitschanskaia, and Kuhn 2013, p. 3). For our study, an examination of data competencies first requires that this broad area of competence be defined and made assessable. For this purpose, it is first of all necessary to better understand fundamental concepts that can be difficult to separate: DL and DS. DL can be understood as "the ability to collect, manage, evaluate, and apply data in a critical manner" (Ridsdale et al. 2015, p. 2). Accordingly, the term represents a fundamental understanding of how decisions are informed by data and how to make use of those data to support evidence-based decision making (Ridsdale et al. 2015). DS, in contrast, is a more interdisciplinary field and can be understood as a process that is both comprehensive in scope and specific in its application. Scientifically sound methods, procedures, algorithms, and systems are implemented to extract generalizable knowledge from both structured and unstructured data (Dhar 2013).

Since our study aims to comprehensively investigate the overall competencies taught in higher education about data understanding, processing, and analysis, neither of these two terms described is intended to be referred to exclusively. However, to comprehensively and precisely capture the investigated competencies, the design of our study adapts to a specifically defined DS curriculum in terms of content. The rationale is that the specific concept (DS) is more capable of covering the more general concept (DL) than the other way around. This is why we will uniformly refer to both as "DS" in the following.

Several proposals for DS curricula, in which structure and content are often quite comparable, can already be found (Anderson et al. 2014; De Veaux et al. 2017; Kauermann and Seidl 2018; Mike 2020; Schwab-McCoy, Baker, and Gasper 2021). Cleveland (2001) early on proposed a curriculum for DS that in its essence encompasses computer science, statistics, mathematics, and application domains; its basic concept is still valid today and provides a basis for various alternative proposals, as noted by Kauermann and Seidl (2018). Nevertheless, although comparable in their essential characteristics, the individual proposals may have individual focuses. For example, the *Data Science Education Framework* (DSEF) proposed by Song and Zhu (2017) strongly focuses on project management skills. In contrast, the *Data Science Knowledge Framework* proposed by the *Initiative for Analytics and Data Science Standards* (IADSS) (Fayyad and Hamutcu 2020) focuses on Big Data. The *EDISON Data Science Framework* (EDSF) (EDISON Consortium 2018), as another example, has a more emphasized focus on knowledge of research principles and methods in comparison.

Building on previous frameworks, the *German Informatics Society* (GI) has further elaborated a specific DS curriculum in 2019. A working group with renowned partners from research and practice was involved in its development. Since this curriculum forms the content framework for the analysis of module handbooks, its main characteristics shall be discussed in more detail. The DS curriculum of the GI (2019) consists of 14 competence areas and is closely related to the EDSF (EDISON Consortium 2018). These areas define the specific learning content required for the training of Data Scientists (see Table 2 for a detailed list). A characteristic feature is the consistent integration of fundamental competencies from the subjects of computer science, mathematics, and statistics, including specific competencies in data security, data ethics, data governance, data integration, and data visualization. In-depth analytical competencies are taught using courses on data mining, machine learning (deep learning), and business intelligence. The curriculum is completed by a transfer of these competence fields in domain—and organization-specific use cases. To customize this curriculum to the varying requirements of specific student target groups, these competence areas were profiled according to the Anderson-Krathwohl taxonomy (Anderson et al. 2001) for the professionalization of three ideal-typical groups of students (referred to as “persona”). This means that the curriculum defines specific degrees of training scope and depth depending on the prior knowledge of the students. An example: a student with a bachelor’s degree in computer science (target group “Persona A”) needs to cover DS content at a different level (e.g., in terms of the profile of mandatory and optional courses) than a student from a completely different discipline with less prior knowledge relevant to DS (target group called “persona C”).

This curriculum is particularly well suited as the framework for the analysis of module handbooks because the DS competence categories are both comprehensive and represented in clearly separable categories. This precise categorization makes it particularly suitable as the basis for a measurement tool. In addition, the curriculum defines specific degrees of training scope and depth depending on the prior knowledge of the students, which are categorized according to the persona groups. This differentiation beneficially clarifies the results, especially concerning the targeted discussion of different training profiles for varied educational backgrounds and goals. The development of the measurement tool derived from this curriculum will be described in detail in the Section 3 (Research Design).

2.2. Module Handbooks in Higher Education

The Bologna process in Europe, and with it the modulization of the study programs, began in 1999 and aimed to introduce a harmonized and more compatible system for the higher education in Europe (Federal Ministry of Education and Research 2015). The goals were to increase students’ international mobility and create a transnational labor market by increasing the quality and standardization of the competencies needed for the future (Federal Ministry of Education and Research 2015; Westphal 2018).

To reach these goals, participating governments promoted a uniform system for higher education at European universities and established cross-national standards. Various activities and structural guidelines were defined, including the creation of comparable educational qualifications and a system of two-level degrees within Europe, the Bachelor and Master degrees (Kerres and Schmidt 2011), the introduction of a system of easily understandable and comparable degrees (Diploma Supplement), the introduction of a credit point system (European Credit Transfer System/ECTS), and the promotion of mobility, European collaboration, and dimensions (European Ministers of Education 1999; Buttner and Vocke 2004).

All of these actions require study achievements that are recognizable in each program and possess a certain degree of comparability (Buttner and Vocke 2004). For this reason, the modularization of study curricula was incorporated into the Bologna Declaration (Kerres and Schmidt 2011). Modularization thus pursued the holistic goal of the Bologna Declaration—creating comparability, making educational pathways more flexible, and ensuring employability (Buttner and Vocke 2004). The presentation of study programs, by describing each program with individual modules, helps to ensure an effective structure (Buttner and Vocke 2004). In addition, the modules provide students with reliable information about the study program, such as contents, requirements, and integration into the overall concept of the study program, or the relationship with other modules offered (Standing Conference of the Ministers of Education and Cultural Affairs 2000). Modules are therefore the building blocks of every study program and can be understood as teaching and learning units that are self-contained in terms of time and subject matter (Winter 2018). An exemplary module description of “Machine Learning” is attached in the Appendix (see Figure A1) to illustrate the contents and structure.

A module handbook represents the documentation of all modules offered to students within a study program. Hence, the module handbook is “the” central document that fundamentally controls teaching and learning activities in higher education today and is often considered a synonym for “Bologna” (Kerres and Schmidt 2011). Besides the study and examination regulations of specific modules, module handbooks also serve as essential documentation of the study program itself (Fregin et al. 2016). Consequently, the module handbook is also the basis for accreditation procedures in which the “studyability” of a study program is examined (Kerres and Schmidt 2011, p. 175).

The analysis of module handbooks is the main topic of this article. We have found a few comparable studies in the literature and will illustrate three exemplary module handbook analyses in the following. First, the study conducted by Fregin et al. (2016) was based on an empirical investigation into the study program’s design and on the competencies that would be acquired. Specifically, they analyzed the extent to which module handbooks meet the Bologna requirements for competence orientation in terms of leadership, ethics, and responsibility in higher education. This analysis enabled differentiated statements about the formal and institutionalized anchoring of competencies in university courses. A second study was conducted by Kerres and Schmidt (2011) and investigated specific criticisms of the Bologna reform and the modularization of study programs. For this purpose, they analyzed whether the specifications for the description of study programs defined by the education ministers' conference were met. Their analysis provided conclusions about the "study reality" (Kerres and Schmidt 2011, p. 175) and the extent to which the *Standing Conference of the Ministers of Education and Cultural Affairs* specifications for the description of the study programs have been implemented practically. The evaluation showed a diversity of implementation practices of Bologna requirements at universities and indicated a variety of the module handbooks and their deviation from the specifications (Kerres and Schmidt 2011). Third, a recent study from Föll and Thiesse (2021) focused specifically on examining Information Systems (IS) curricula (in total, 90 programs, and 3700 distinct modules) using efficient text mining techniques. As a result, an overall view of the differences and similarities of the discipline's curricula was created. This provided, among other things, a contribution to targeted curriculum design.

In contrast, our study examines the textual database of module handbooks through quantitative content analysis to assess the status quo of integrating DL and DS competencies in higher education. This means that the focus is more on the investigation of very specific teaching content than on the analysis of documentation practice or the holistic characterization of entire study programs. However, compared with the studies mentioned, even this narrowed focus contributes to the investigation of characteristic features of Germany's higher education landscape. Moreover, the development and application of a topic-specific dictionary is a methodological novelty in this context.

2.3. Higher Education Research

The system of higher education in Germany is divided into "(Research) Universities" and "Universities of Applied Sciences" (hereafter referred to as UNI and UAS). This dual system characterizes the tertiary level of higher education in the country (Fichtl and Piopiunik 2017).

UNIs provide science-based education and conduct basic research (German Rectors' Conference 2021a). They are research- and theory-based institutions and the most common and obvious kind of higher education in Germany. Most have a longer history than that of the UASs. Some UNIs have specialized in certain subject areas, such as technical universities or medical universities. The original aims of establishing UASs were to expand the range of courses, to include courses with a shorter duration of study, and to provide a stronger practical orientation. Later, application-oriented research and application-oriented training as a basis for later professional activity were added as central features of these universities (Federal Ministry of Education and Research 2020; German Rectors' Conference 2021a). According to the *Federal Ministry of Education and Research Germany*, UASs drive innovation as well as new and improved products and services (Federal Ministry of Education and Research 2020).

Higher education research is a field of study that focuses on the institutions of high education as the differences between UNIs and UASs—and is currently considered a young field of research. "Interest in topics of higher education research has increased strongly in the German-speaking area in recent years" (Society for Higher Education Research 2021). Nevertheless, higher education research remains small in comparison to the size of the higher education system (Teichler 2015). "Notably, comparative studies got momentum and internationalization became a major theme" (Teichler 2015, p. 1). For example, Lirk and Zumbach (2015) compared the perceptions of (natural) scientific working environments as reported by students at UNIs and UASs. Another comparison between UNIs and UASs was conducted by Fichtl and Piopiunik in 2017. They examined the role of graduates of German UASs in comparison with that of graduates of German UNIs in terms of research and development (R&D) activities in the German labor market and surveyed differences in labor market outcomes, such as income, cognitive skills, and regional mobility. Another notable study by Föll and Thiesse (2021) placed a specific focus on comparing curricula and investigated the extent to which teaching content in the Information Systems discipline differs in UNIs and UASs.

We contribute to this line of research by (a) analyzing differences in the delivery of a very specific course content (i.e., DS), by (b) simultaneously conducting a comparative study not only by the type of university but of three courses simultaneously, and by (c) using a database of module handbooks (more than 13,950) that has not been used to this extent before.

3. Research Design

3.1. Research Approach and Process

Our study, descriptive in nature, assesses the integration of DL and DS competencies in higher education. In doing so, our study follows a deductive research approach (Hyde 2000) that is based on an established and generalizable framework, an established DS curriculum from the GI (2019), and attempts to determine to what extent this framework is prevalent in a particular research context (three study programs, and two university types).

Content analysis, the method used, is a research technique for making valid inferences from texts to the contexts of their use (Krippendorff 2019). This definition clearly emphasizes the inferential nature of content analysis; its application typically requires well-grounded inferential frameworks (based on existing theories, previous research, or experience) and rigorous procedural (coding) rules to produce meaningful answers to research questions that are based on valid observations of unstructured texts (White and Marsh 2006). Within this framework, our study first applies qualitative content analysis and then quantitative content analysis. In the first step, qualitative content analysis (Mayring 2004) is used to infer concepts relevant to the dictionary development through the use of an inductive process. Subsequently, for the implementation of the actual descriptive study, a quantitative (computer-aided) content analysis is used that implements a "measurement" of texts by detecting, counting, and statistically analyzing predefined content, in the form

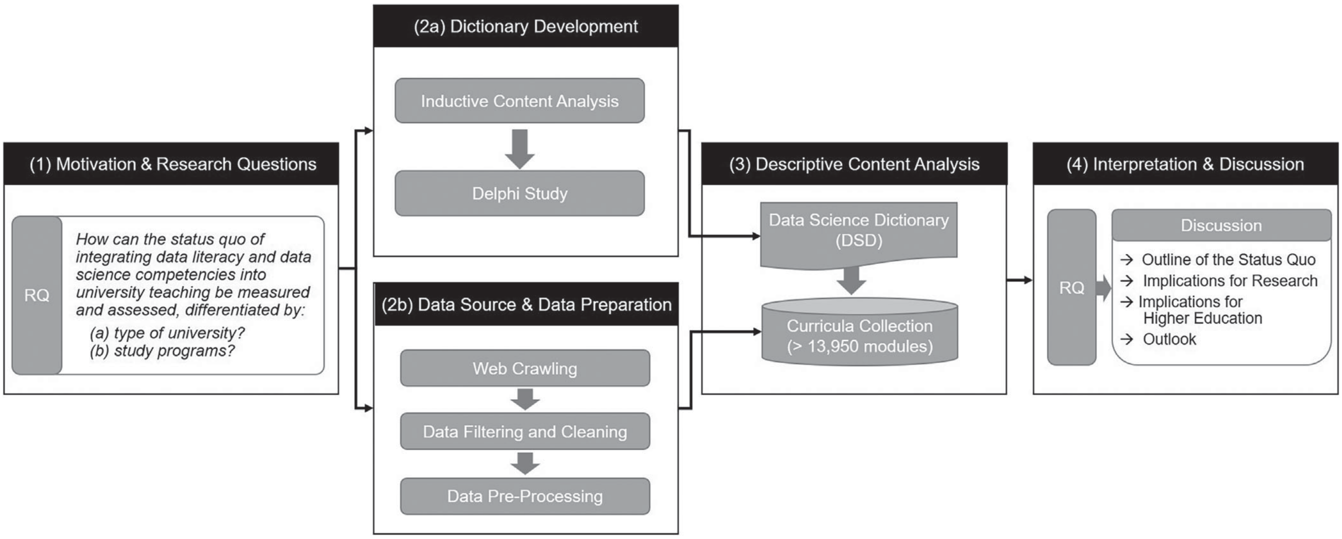


Figure 1. Research Process.

of specific keywords. In this way, quantitative statements can be generated that describe the appearance and extent of concepts contained in the texts (Krippendorff 2019).

For thorough documentation of our study, the research process is shown in Figure 1. The research process started with (1) the theoretical justification of the motivation and derivation of the research question (RQ). The corresponding discussions are presented in the first two sections. For the preparation of the descriptive content analysis, we conducted two parallel processes. On the one hand, (2a) we developed the measurement instrument in the form of a dictionary (DSD), which is supposed to representatively capture the contents of a DS curriculum (GI 2019). On the other hand, (2b) we established a database, which ultimately contained bachelor curricula (i.e., module handbooks) from 280 universities with a total of more than 13,950 module descriptions (see the following Section 3.2 for details). As part of (3) the descriptive content analysis, we applied the developed DSD to the module handbook collection and generated results to answer the RQ. Finally, (4) we interpreted the results to derive specific implications for further university practice and research.

3.2. Research Data

The process of data collection for this examination of the textual content of module handbooks began with the selection of individual study programs for inclusion, which was based on the *German Rectors' Conference* (2021b). We prioritized individual bachelor study programs in the fields of Business Administration, Business & Information Systems Engineering, and Computer Science and identified the websites of the UNIs and UASs. We then used a web crawler to search the universities' websites for the corresponding module handbooks between December 2018 and June 2019. The web crawler downloaded these module handbooks, extracted and transformed the content, and stored it in an appropriate structural text format in a database. In addition, the database also separated the learning objectives and learning content of the individual modules, as these represent the focus of our analysis. This structuring of the module handbooks subsequently enabled us to carry out various measurements and analyses, such as comparisons between UNI and UAS, study programs, or even individual module offerings.

After the data extraction and structuring, systematic data cleansing and pre-processing were completed (Feldman and Sanger 2007). First, we manually excluded modules that were only described in English (since our study took place in German universities, the language of the modules was standardized to German; the following data and results have been translated into English) or had missing data in their database, as extraction was not possible. In this way, the number of modules was reduced to 13,951 at 280 German institutions of higher education (95 UNIs and 185 UASs). Table 1 gives an overview of the included numbers of UNIs and UASs (95 and 185) and of the total number of modules. In total, we examined 94 Business Administration study programs, 86 Business & Information Systems Engineering study programs, and 100 Computer Science study programs. Second, we prepared the textual data for later coding by using two typical techniques of Natural Language Processing (Manning and Schütze 1999): first tokenization (extraction of word components in the text document collection) and then normalization (unification of nearly identical words) using a lemmatization procedure that converts different word forms to their base form or dictionary form, thus reducing variation in the data. The Data pre-processing process and also the following frequency analysis were performed with the tool WordStat 8.0.

The information value of our database is illustrated in Appendix Figure A1 by the example module "Machine Learning." Since our analysis of module handbooks seeks to measure specific teaching content in the textual module descriptions, the specific focus of the quantitative content analysis is on the chapters of "learning objectives" and "learning content." The mandatory learning objectives describe the knowledge a student gains through participation in the module. The learning content describes any subject-related, methodical, practical, and interdisciplinary content that should be taught (Standing Conference of the Ministers of Education and Cultural Affairs 2010). The result is a unique text database that allows deep insights into the distribution and composition of teaching content in German

Table 1. Research data overview.

	Business administration		Business & information systems engineering		Computer science		Total	
	UNI	UAS	UNI	UAS	UNI	UAS	UNI	UAS
Number of universities (i.e., curricula) and shares (%)	n = 94 (33.6%)		n = 86 (30.7%)		n = 100 (35.7%)		n = 280 (100%)	
	n = 24 (25.3%)	n = 70 (37.8%)	n = 27 (28.4%)	n = 59 (31.9%)	n = 44 (46.3%)	n = 56 (30.3%)	n = 95 (100%)	n = 185 (100%)
Total number of modules and shares (%)	n = 4,951 (35.5%)		n = 3,830 (27.4%)		n = 5,170 (37.1%)		n = 13,951 (100%)	
	n = 1,482 (26.5%)	n = 3,469 (41.5%)	n = 1,318 (23.5%)	n = 2,512 (30.0%)	n = 2,795 (50.0%)	n = 2,375 (28.5%)	n = 5,595 (100%)	n = 8,356 (100%)
Average number of modules per curriculum	52.67		44.53		51.7		71.9	
	61.75	49.56	48.81	42.58	63.52	42.41	84.80	64.80

Table 2. Data Science Dictionary (DSD) – categories of competence.

#	Field of Competence	#	Field of Competence
1	Basics Mathematics & Statistics	7	Data Governance
2	Advanced Mathematics &	8	Data Integration
3	Statistics	9	Data Visualization
4	Basic Computer Science	10	Data Mining
5	Advanced Computer Science	11	Machine Learning/Deep Learning
6	Cryptography & Data Security Data Ethics & Data Privacy	12	Business Intelligence

higher education as this database covers the majority of all higher education institutions in Germany.

3.3. Dictionary Development

Just as a questionnaire or interview must be structured to generate knowledge, an analysis of textual content can also be structured. The core of a sound content analysis is, therefore, a clearly defined analytical construct that organizes the categories, indicators, and underlying coding rules (Krippendorff 2019). One widespread procedure is to develop a dictionary in which distinctive thematic categories are modeled with appropriate keywords (Beattie and Thomson 2007). Both the categories used and the respective keywords should represent the thematic area to be analyzed as meaningfully and with as little overlap as possible. For this reason, safeguarding objectivity and validity in the development of this analytical construct is always in the foreground (Krippendorff 2019). Given that there is no such construct available for coding "DS," such an analytical construct was developed exclusively for our analysis of module handbooks. This dictionary—called the "Data Science Dictionary" (DSD)—is designed to quantitatively capture the existence of DS content in module descriptions. The complete construct with its categories and corresponding keywords can be found in the Appendix in Table A1. In the following, this development process (see Figure 1) is described in more detail.

As a first step, concise categories had to be created that capture the DS theme as completely and validly as possible. Such categories help to contextualize the measurement and give our study the analytical depth that it requires. To safeguard the objectivity and validity of these categories, Krippendorff (2019) recommends using established theories or expert knowledge. For our analysis, we relied on the DS curriculum proposed by the GI (2019). The categories of DS competencies described in this curriculum (see Table 2) can therefore be regarded as such a well-founded framework. Table 2 defines the particular category by number and field of competence. The categories are hereafter referred to as "#number of the category." Two categories were not included because they represent cross-cutting, that is, nonexclusive DS content (domain-specific applications, DS in organizations).

In the next step, it was necessary to define relevant keywords and to thematically allocate them to the appropriate categories. Here, too, reliance on established theories or expert knowledge is recommended (Krippendorff 2019). To identify and allocate valid keywords, we followed a two-phase process (see Figure 2). First, we developed a raw version of the dictionary with a preliminary keyword collection before conducting a Delphi study, a systematic and multistage survey of experts, to have it reviewed by an interdisciplinary panel of DS specialists. The phases are presented and explained in more detail below.

3.3.1. Phase 1: Identification & Allocation of Keywords (Inductive Content Analysis)

We developed the raw version of the DSD with a preliminary keyword collection based on two sources:

- *DS literature*: The keyword indexes of a total of 24 textbooks related to DS were automatically extracted, and duplicates of keywords were removed. Subsequently, the keywords were ranked based on the frequency of occurrence in the textbook collection. As a next step, with the help of a panel of experts, this list was further prioritized. Finally, a list with a total of 142 keywords remained (referred to as "first list").
- *DS module handbooks*: This first list was supplemented by keywords identified through inductive (explorative) content analysis (Mayring 2004) of ten module handbooks for the study program "DS" (the selected study programs were referenced in the underlying DS curriculum of the GI (2019).

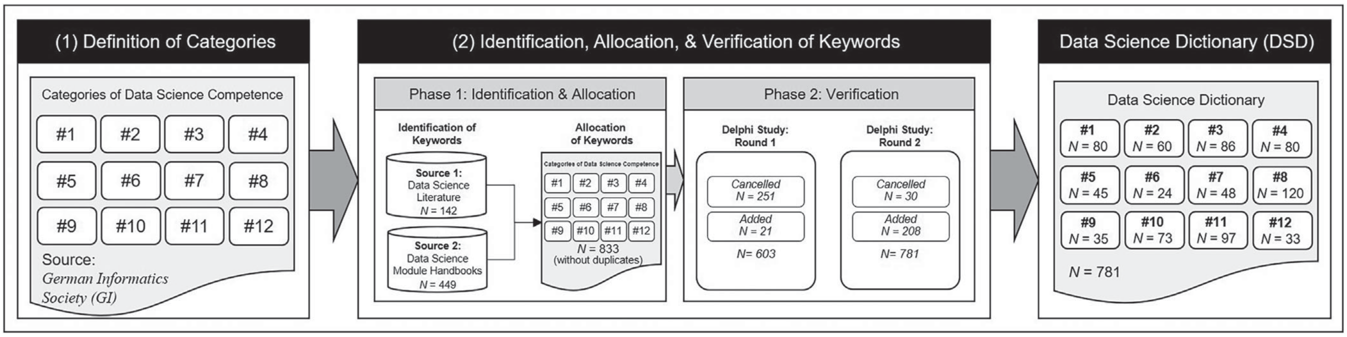


Figure 2. Data Science Dictionary (DSD) development – definition and allocation of keywords.

Three coders extracted significant words to describe "DS" from these ten module handbooks. This list was also prioritized with the help of an expert panel, resulting in 449 words (referred to as "second list").

The keywords then had to be assigned to the appropriate categories to accurately capture the content of that category. Two coders (research assistants) assigned the keywords from both lists to the categories of the DSD. The assignment of keywords was done independently in three rounds. An inter-coder agreement concerning category assignment was subsequently an acceptable 83.62% (first list) and 77.06% (second list). Differences were discussed and corrected by consensus. The mean assignment of both lists was an acceptable 79.36%, which in this case indicates a sufficient level of reliability (Landis and Koch 1977).

Subsequently, we supplemented the raw version of the DSD by adding further keywords from the underlying DS curriculum (GI 2019) that were previously missing and by making them directly assignable to the categories. Furthermore, we added keywords that were found in each category's individual subject literature to ensure an even more comprehensive mapping of each category of the DSD. Finally, the raw version of the DSD included a total of 833 unique keywords, which were assigned to 12 categories.

3.3.2. Phase 2: Verification of Keywords (Delphi Study)

The categories and keywords must be valid in the sense of the question, "Are we really measuring what we want to measure?" (Neuendorf 2002, p. 12). To test validity, one approach could be to verify the construct by an uninvolved third party for plausibility "on its face" ("face validity") (Neuendorf 2002). In phase two of the dictionary development process, we therefore conducted a two-round Delphi study "to obtain the most reliable consensus" (Dalkey and Helmer 1963, p. 458). The Delphi study method presents a systematic and multistage survey of experts. The aim is to obtain the most reliable consensus of opinion among a group of experts as a validation strategy (Dalkey and Helmer 1963). Our raw version of the dictionary was sent to a panel of seven interdisciplinary experts for review (four professors and three practitioners with experience in the field of DS). This group had been assembled to ensure that at least one expert could cover each of the following topics: mathematics, statistics, computer science, business intelligence, machine learning, and DS in general.

We conducted the Delphi study between December 2020 and March 2021 and collected data via E-mail. As the literature for the classic Delphi study suggests four rounds, we considered implementation with two to six rounds to be a practicable approach (Bradley and Stewart 2003). The number of rounds "depends on the amount of time available, whether the researcher has indicated the Delphi sequence with one broad question or with a list of questions, and consideration of levels of sample fatigue" (Hasson, Keeney, and McKenna 2000). We decided to conduct a Delphi study with two rounds, as after two rounds, we were able to reach an acceptable level of agreement between all experts (agreement rate over 90%).

In the first round of the Delphi study, the individual experts evaluated each keyword per category of the raw DSD concerning the following questions:

Does the keyword describe DS appropriately and does the word fit in the category?

Are there missing keywords in a category that need to be added?

The experts had three weeks to answer the first round. Afterward, we took one week to evaluate and summarize the first round of the Delphi study and to clarify ambiguities. Before giving feedback to the experts and sending the instructions for the second round of the Delphi study, we calculated the agreement rates between the respective experts for each category of the DSD (see Appendix Table A3). In the literature, there is no clear definition of an appropriate "agreement" among experts in a Delphi study. In some cases, agreement rates of 51% are recommended (McKenna 1994), and in other cases, an agreement rate greater than 80% is suggested (Green et al. 1999). However, the low agreement rates (some less than 50%) of round one of our Delphi study (see Appendix Table A3 column "Round 1–Agreement rate [%]") prompted a second round (Hasson, Keeney, and McKenna 2000).

To incorporate the initial feedback, we first preprocessed the dictionary to an "intermediate" DSD. Keywords that were rated as nonrepresentative for describing DS by at least two experts were eliminated (number of keywords in Appendix Table A3 column "Round 1–Canceled"). New keywords that were added (number of keywords in Appendix Table A3 column "Round 1–Added") to a respective category by at least one expert were integrated into the proposed category. The intermediate DSD as well as the eliminated and added words were separately represented to the experts for review in round two of the Delphi study.

Table 3. General proportions of modules with contents related to DS.

	Business Administration		Business & Information Systems Engineering		Computer Science	
	UNI	UAS	UNI	UAS	UNI	UAS
Total number of modules with DS content	1565		2315		4116	
	486	1079	794	1521	2218	1898
Proportion (%) of modules with DS content	31.6		60.4		79.6	
	32.8	31.1	60.2	60.5	79.4	79.9
Average number of modules with DS content per curriculum	16.65		26.92		41.16	
	20.25	15.41	29.41	25.78	50.41	33.89
Standard deviation (S.D.)	12.97		13.20		25.50	
	16.34	11.33	18.12	9.98	33.21	13.21
Min.	1		6		11	
	4	1	9	6	12	11
Median	13		24.5		37	
	15	12.5	25	24	44	36
Max.	78		84		177	
	78	71	84	55	177	78
Extent of module (in number of words)	190.2		140.6		118.3	
	199.5	186.3	130.5	145.9	116.9	120.0

In round two, the intermediate DSD derived from round one was reviewed by the experts. We also represented the eliminated and added keywords from round one of the Delphi study to cross-check them and to examine if the decision to eliminate needed to be revised or if an added word was inadequate. If at least one expert stated contradiction to elimination or addition of a keyword, the decision was revised and the keyword was kept or eliminated from the DSD. The experts were also asked to cancel from or add more words to the intermediate DSD. After round two, we examined agreement rates greater than 90% in all 12 categories of the intermediate DSD (see Appendix Table A3 column "Round 2–Agreement rate [%]").

Finally, the Delphi study resulted in a comprehensive final dictionary describing the field of DS. The final dictionary included 12 categories and 781 keywords in total. Finally, the suitability of the DSD could be tested and proven by the fact that 71% of the incorporated DS keywords could be found in the module handbooks of the diverse study programs (see a coverage analysis by DS categories in Appendix Table A2; a list of the top 25 keywords is provided in Appendix Table A5).

4. Results

4.1. Result 1: General Proportions (%) of Modules with DS Content

In the first step, we analyzed how much DS-related content (at this point without a differentiation according to the DS categories) is taught in the individual study programs (Business Administration, Business & Information Systems Engineering, and Computer Science), also differentiated by university type. Accordingly, the question we were trying to answer is: *What is the general proportion of modules with DS content in a study program?* For this purpose, we calculated the general proportions (%) of modules with DS content (see formula 1). We first identified the modules that contain any amount of DS content. All module descriptions M_i were screened per subgroup, such as study program and university type, to determine whether at least one keyword from the DSD (independently from the category) was included in the documented learning objectives or learning contents. The "i" is running from 1 to the number of the respective subgroups (= 1, …, n). Finally, M_{DS} was then divided by the total number of modules M_i per subgroup (e.g., the number of modules per study program; see also Table 1) to calculate the general proportions (%) of modules with DS content (the results are presented in Table 3):

$$Proportion\ [\%]\ of\ modules\ with\ DS\ content = \frac{\sum M_{DS}}{\sum M_i}\ x\ 100 \tag{1}$$

As a further indicator, the average number of modules M with DS content per individual curriculum C (i.e., module handbook) was calculated. The underlying question is: *How many modules with DS content are offered on average in an individual curriculum?* This indicator can be described with the following formula, where the corresponding averages per study program can be found in Table 3:

$$Average\ number\ of\ modules\ with\ DS\ content\ per\ curriculum = \frac{\sum M_{DS,\,C}}{\sum M_C} \tag{2}$$

In order to take into account the differences in documentation practices by study programs, additionally, the extent (or length) of the modules was assessed by measuring the average number of words used (not only the DS keywords). Here it became apparent that modules in the Business Administration programs include significantly more words (190.2) than those in Business & Information Systems Engineering (140.6; 35% less) and Computer Science (118.3; 61% less). Since we perform a frequency analysis of keywords, the varying amount of documentation is thus also noteworthy for the assessment of the results. The significance of these averages was also examined using further indicators (S.D., Min., Median, Max.), which indicate that there nevertheless appears to be a wider range and, in particular, significant outliers in individual modules.

In summary, the presented proportions (in %) of modules with DS content show an increase from Business Administration with 31.6% to Business & Information Systems Engineering with 60.4% to Computer Science with 79.6%. Hence, this means that the more technical the focus of the program of study, the more DS content it includes. Since there are no significant differences between the university types (UNI and UAS) within the study programs, this in turn also confirms the differences between the study programs, as the university types serve as the comparison group.

4.2. Result 2: Proportions of DS Categories (%) within the DS Modules

Our analyses also revealed which specific DS content was taught, as measured by each of the 12 categories of the DSD. Put simply, the question here is: *If a module contains DS content, what exactly is being taught?* To answer this question, we determined what the proportion (in %) of modules M with content related to specific DS categories (DSC) is within the sum of modules M with DS content (M_{DS}). Specifically, the average proportions of respective modules M_{DSC} per individual curriculum C were first calculated (separated by subgroups, such as study programs) to subsequently determine the overall average proportions through division by the total number of considered curricula C_i of the subgroup ($i = 1, \ldots, n$). C_i contains in total the individual bachelor curricula from 280 universities with more than 13,950 module descriptions. Again, the occurrence of the corresponding keywords in a module was decisive. The calculation of this proportion is represented by the following formula:

$$Proportion\ [\%]\ of\ DS\ categories = \frac{\sum \frac{M_{DSC,\,C}}{M_{DS,\,C}}}{\sum C_i}\ x\,100 \qquad (3)$$

The complete results (including statistical significance tests) are presented in Table A4 in the Appendix. Figure 3 graphically depict the individual proportions of DS categories within DS modules by study program, ordered by extent. For example, for "Basic Mathematics & Statistics" in Business Administration, 19.6% of the respective DS modules represent the content, which means contain associated keywords, of category #1. However, this does not mean that Business Administration teaches "more" math than Business & Information Systems Engineering (9.3%) or Computer Science (12.4%). It simply means that math is given more weight within the DS contents in this study program. In short, within the totality of all DS-related modules, the topic of mathematics occupies a higher share in the Business Administration program than more specific DS topics. In summary, it can be shown that DS primarily takes place in basic mathematics and statistics in Business Administration. As another example, 13.4% of the modules in Business Administration reference business intelligence topics within the appearance of DS content (#12). Compared to the other study programs, this large percentage stands out.

In Business & Information Systems Engineering and Computer Science, DS primarily takes place in classical computer science (#3), although individual characteristics are also present (e.g., clearly different values in #7). "Special topics" such as cryptography and security (#5) and data ethics and privacy (#6) are represented as a marginal area (less than 5%) in all study programs.

Examination of the significance of differences between study programs shows that almost all are highly significant ($p < 0.05^*$). Only #3, #4, #8, and #9 were not ($p > 0.05$). We confirmed the significance of the difference of means between the study programs by using the technique of one-way analysis of variance. As presented in the detailed numerical results (see Appendix Table A4), these significances can also be supported by the comparison groups (UNI and UAS) since similar results can be found between the study programs at the two types of universities. The significance of differentiation of the means between UNI and UAS was measured by using the technique of a two-sample t-test (p-values are also represented in Appendix Table A4).

4.3. Result 3: Density of DS Content within the Corresponding DS Modules

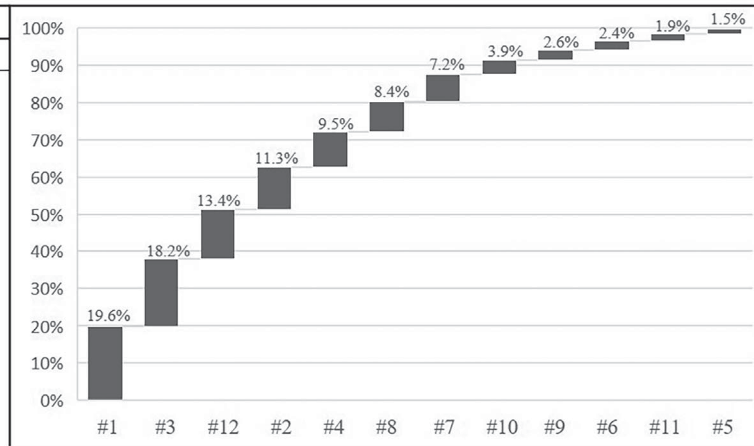
We also examined the intensity to which the 12 specific DS categories are taught in the documented modules (see Table 4). The underlying question is: *If a module contains specific DS content, to what intensity is it taught?* To answer this question, we calculated the average keyword frequency KWF per DS category DSC and corresponding DS modules M_{DSC} at the individual level of a curriculum C. To put it more simply, this measurement expresses how "dense" the proportion of category-specific keywords is in the full corpus of words used in a module description. Corresponding results can thus provide an impression of the significance of a DS-specific category in the teaching of a module. However, this measure brings a limitation in regard to the interpretability. For example, concepts like business intelligence can include important discussions of DS without requiring much DS-specific terminology. The following formula describes the calculation of this density, which was ultimately divided by the total number of curricula C_i ($i = 1, \ldots$, n) in the varying subgroups (e.g., study program) to represent an overall average density:

$$Density\ (\emptyset\ keyword\ frequency)\ per\ DS\ category = \frac{\sum \frac{KWF_{DSC,C}}{M_{DSC,\,C}}}{\sum C_i} \qquad (4)$$

In this way, we could determine that, for example, "Machine/Deep Learning" (#11) is taught in the different study programs in three significantly different densities or to different extents (p-values $<0.05^*$). In Computer Science, "Machine/Deep Learning" (#11) is taught most intensively, with 1.97 keywords per module with corresponding content, followed by Business & Information Systems Engineering with 1.38 keywords per module (medium density). Business Administration follows at a distance: only 0.58 keywords per

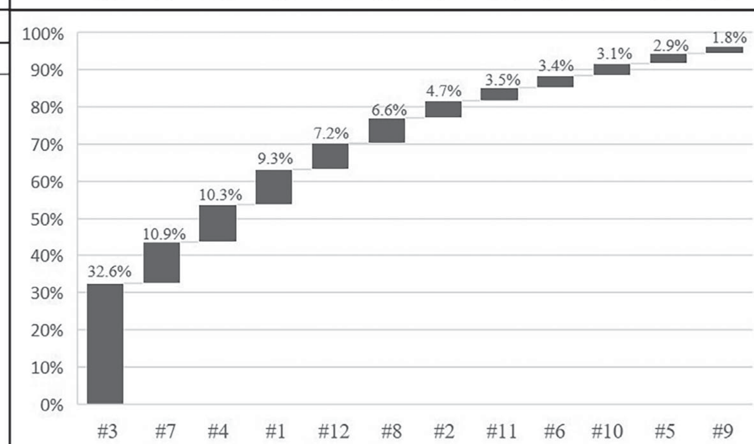
Business Administration

#	Category	Proportion (%)
1	Basic Mathematics & Statistics	19.6%
2	Advanced Mathematics & Statistics	11.3%
3	Basic Computer Science	18.2%
4	Advanced Computer Science	9.5%
5	Cryptography & Data Security	1.5%
6	Data Ethics & Data Privacy	2.4%
7	Data Governance	7.2%
8	Data Integration	8.4%
9	Data Visualization	2.6%
10	Data Mining	3.9%
11	Machine Learning/Deep Learning	1.9%
12	Business Intelligence	13.4%



Business & Information Systems Engineering

#	Category	Proportion (%)
1	Basic Mathematics & Statistics	9.3%
2	Advanced Mathematics & Statistics	4.7%
3	Basic Computer Science	32.6%
4	Advanced Computer Science	10.3%
5	Cryptography & Data Security	2.9%
6	Data Ethics & Data Privacy	3.4%
7	Data Governance	10.9%
8	Data Integration	6.6%
9	Data Visualization	1.8%
10	Data Mining	3.1%
11	Machine Learning/Deep Learning	3.5%
12	Business Intelligence	7.2%



Computer Science

#	Category	Proportion (%)
1	Basic Mathematics & Statistics	12.4%
2	Advanced Mathematics & Statistics	4.4%
3	Basic Computer Science	40.6%
4	Advanced Computer Science	10.4%
5	Cryptography & Data Security	2.9%
6	Data Ethics & Data Privacy	2.3%
7	Data Governance	7.0%
8	Data Integration	6.9%
9	Data Visualization	2.4%
10	Data Mining	2.0%
11	Machine Learning/Deep Learning	4.8%
12	Business Intelligence	1.9%

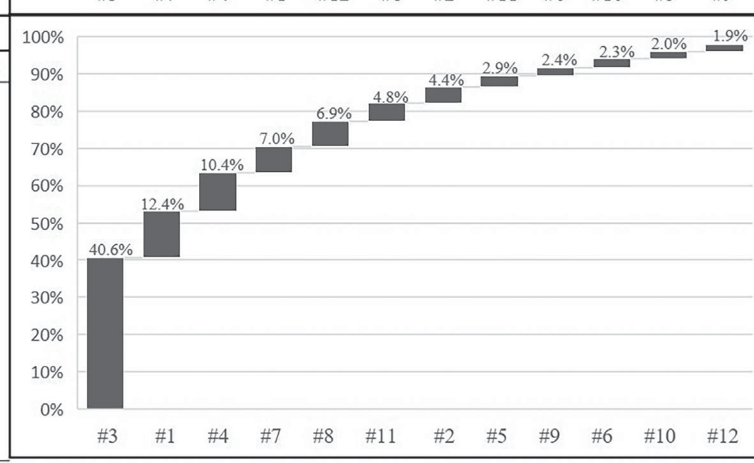


Figure 3. Proportions of DS categories (%) by study program.

the corresponding module. As another distinctive example, #12 "Business Intelligence" is also taught in three significantly different densities: first Business & Information Systems Engineering (most intensive: 3.50 keywords per module), next to Business Administration (medium intensive: 2.39), and finally Computer Science (less intensive: 1.35). These different density levels can also be considered highly significant (p-values $<0.05^*$).

These results are confirmed by the UNI and UAS comparison groups; study programs at the two types of universities show similar densities. For example, #12 "Business Intelligence" in Computer Science shows an overall density of 1.352 with an average UNI density of 1.32 and an average UAS density of 1.38 (p-value = 0.87).

4.4. Result 4: DS Teaching Portfolio

Based on the results regarding "Proportion" (Result 2) and "Density" (Result 3), a portfolio was constructed in the next step that first outlined the status quo and then served as a tool for tracking the future evolution of the integration of DS competencies in higher education. The use of portfolios for assessing integrated curricula has a long tradition (Carpenter, Ray, and Bloom 1995). Although there are different definitions

Table 4. Density of DS content per corresponding module.

	Business Administration		Business & Information Systems Engineering		Computer Science	
	UNI	UAS	UNI	UAS	UNI	UAS
#1 Basics mathematics & statistics	5.03		4.50		4.11	
			$p = 0.05$			
	4.65	5.17	4.08	4.70	3.85	4.31
	$p = 0.56$		$p = 0.24$		$p = 0.26$	
#2 Advanced mathematics & statistics	1.89		1.82		1.89	
			$p = 0.95$			
	1.95	1.86	1.57	1.95	1.94	1.75
	$p = 0.78$		$p = 0.11$		$p = 0.57$	
#3 Basic computer science	2.78		3.91		3.96	
			$p < 0.001$***			
	3.21	2.63	3.71	4.00	4.05	3.89
	$p = 0.16$		$p = 0.38$		$p = 0.65$	
#4 Advanced computer science	1.35		2.29		2.28	
			$p < 0.001$***			
	1.28	1.38	1.85	2.50	2.07	2.45
	$p = 0.58$		$p = 0.01$*		$p = 0.12$	
#5 Cryptography and data security	0.47		1.71		2.13	
			$p < 0.001$***			
	0.75	0.38	1.17	1.97	1.58	2.56
	$p = 0.13$		$p = 0.04$*		$p = 0.01$*	
#6 Data ethics and data privacy	0.76		1.31		1.19	
			$p = 0.01$*			
	0.89	0.72	1.04	1.44	0.84	1.46
	$p = 0.72$		$p = 0.18$		$p < 0.001$***	
#7 Data governance	1.50		2.37		2.19	
			$p < 0.001$***			
	1.68	1.43	2.07	2.51	2.32	2.10
	$p = 0.46$		$p = 0.07$		$p = 0.29$	
#8 Data integration	1.08		1.39		1.56	
			$p < 0.001$***			
	1.44	0.95	1.21	1.49	1.60	1.53
	$p = 0.02$*		$p = 0.11$		$p = 0.68$	
#9 Data visualization	0.59		0.91		1.30	
			$p < 0.001$***			
	0.21	0.72	0.74	0.99	1.39	1.24
	$p = 0.01$*		$p = 0.44$		$p = 0.60$	
#10 Data mining	0.90		1.66		1.63	
			$p = 0.01$*			
	1.77	0.61	1.13	1.92	2.32	1.07
	$p = 0.03$*		$p = 0.02$*		$p = 0.01$*	
#11 Machine learning/deep learning	0.58		1.38		1.97	
			$p < 0.001$***			
	0.88	0.48	1.29	1.43	2.22	1.78
	$p = 0.18$		$p = 0.62$		$p = 0.11$	
#12 Business intelligence	2.39		3.50		1.35	
			$p < 0.001$***			
	1.90	2.56	2.29	4.09	1.32	1.38
	$p = 0.09$		$p < 0.001$***		$p = 0.87$	

NOTE. *p*-values: $p > 0.05$; $p < 0.05$*; $p < 0.01$**; $p < 0.001$***.

of portfolios in the context of teaching, they are united by the basic characteristic that they are a "systematic and organized collection of evidence used by the teacher and student to monitor […] the student's knowledge" (Vavrus 1990, p. 48). With this definition comes a broad range of potential purposes; for example, portfolios can also be used to understand the composition and progress of curricula (Carpenter, Ray, and Bloom 1995). This purpose is addressed in our study by systematically organizing the presence of the 12 DS competency categories to be acquired by students in a portfolio.

Business Administration

#	Category	Proportion	Density
1	Basic Mathematics & Statistics	19.6%	5.033
2	Advanced Mathematics & Statistics	11.3%	1.887
3	Basic Computer Science	18.2%	2.776
4	Advanced Computer Science	9.5%	1.352
5	Cryptography & Data Security	1.5%	0.473
6	Data Ethics & Data Privacy	2.4%	0.763
7	Data Governance	7.2%	1.495
8	Data Integration	8.4%	1.078
9	Data Visualization	2.6%	0.586
10	Data Mining	3.9%	0.903
11	Machine Learning/Deep Learning	1.9%	0.578
12	Business Intelligence	13.4%	2.390

Business & Information Systems Engineering

#	Category	Proportion	Density
1	Basic Mathematics & Statistics	9.3%	4.495
2	Advanced Mathematics & Statistics	4.7%	1.824
3	Basic Computer Science	32.6%	3.909
4	Advanced Computer Science	10.3%	2.287
5	Cryptography & Data Security	2.9%	1.707
6	Data Ethics & Data Privacy	3.4%	1.310
7	Data Governance	10.9%	2.368
8	Data Integration	6.6%	1.394
9	Data Visualization	1.8%	0.91
10	Data Mining	3.1%	1.664
11	Machine Learning/Deep Learning	3.5%	1.384
12	Business Intelligence	7.2%	3.503

Computer Science

#	Category	Proportion	Density
1	Basic Mathematics & Statistics	12.4%	4.111
2	Advanced Mathematics & Statistics	4.4%	1.887
3	Basic Computer Science	40.6%	3.959
4	Advanced Computer Science	10.4%	2.279
5	Cryptography & Data Security	2.9%	2.128
6	Data Ethics & Data Privacy	2.3%	1.185
7	Data Governance	7.0%	2.194
8	Data Integration	6.9%	1.560
9	Data Visualization	2.4%	1.304
10	Data Mining	2.0%	1.625
11	Machine Learning/Deep Learning	4.8%	1.972
12	Business Intelligence	1.9%	1.352

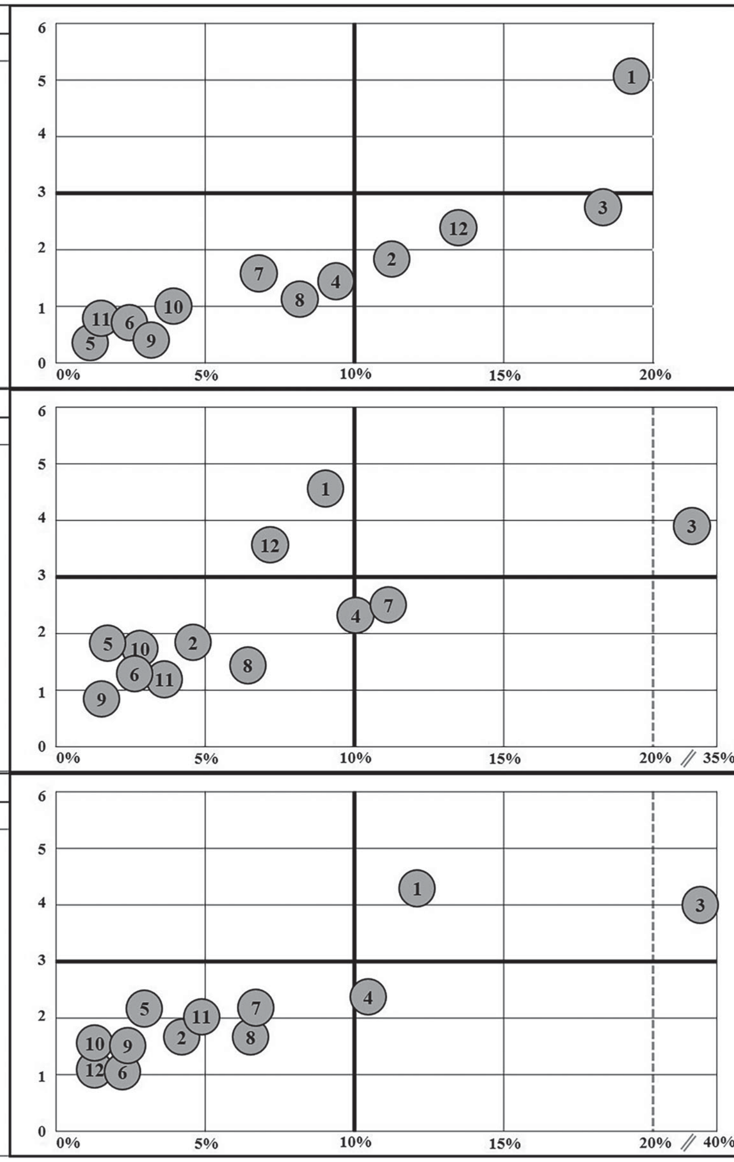


Figure 4. DS teaching portfolios.

A portfolio, as a transparent assessment tool, offers specific advantages that make it particularly suitable for our study: (a) it provides a basis for comparing and evaluating curricula, (b) it allows for better communication and discussion, and (c) it provides a basis for examining changes over time (Carpenter, Ray, and Bloom 1995; Hartley, Frontczak, and Rudelius 2015). The “DS teaching portfolio” (see Figure 4) is organized as a 2 × 2 matrix, where two independent dimensions are presented in combination:

- “Proportion”: the x-axis describes in % how often a category of the DS competencies occurs in the examined modules per study program. For comparability, the scale is limited to a maximum of 20% by default, although outliers are still shown.
- “Density”: the y-axis describes how intensively a category is addressed in a module based on the respective keyword frequency (i.e., average keyword density per module).

In combination, the matrix thus provides insight not only into the mere existence of a DS category but also into the intensity of its treatment. The upper right quadrant, for example, indicates that a DS category is not only taught very regularly but is also addressed in greater depth. The extreme counterexample is the quadrant at the bottom left, which depicts those competence categories that rarely appear in module collections and are only shallowly addressed. Therefore, this combined representation is helpful for a well-founded assessment because it allows for the identification of DS categories that appear regularly in a larger proportion of the modules but are not dealt with in comparatively greater depth (see, e.g., #3 in the Business Administration study program; Figure 4).

The portfolio illustrates that the integration of DS in Business Administration (Figure 4) takes place largely through mathematical-statistical topics (#1 and #2) and fundamental computer science (#3). The "Mathematics & Statistics" (#1) is addressed in approx. 20% of the modules and, in addition, with greater intensity (approx. five keywords per module on average). “Business Intelligence” (#12) stands out in this study program as a competence that is addressed both regularly and quite intensively (in 13.4% of the modules, with an average of 2.4 keywords). Competencies of data handling (#7 or #8) are represented less frequently. More specific competencies, such as #5 “Data Ethics & Data Privacy” or #11 “Machine Learning/Deep Learning,” on the other hand, are clearly niche topics.

In Business & Information Systems Engineering (Figure 4), the integration of DS competencies largely takes place via the fundamentals of mathematics/statistics (#1) and computer science (#3, #4). However, #3 “Basic Computer Science” stands out clearly in proportion and density. In addition, “Business Intelligence” (#12) and “Data Governance” (#7) are strongly established. These topics obviously indicate a more pronounced focus on business topics than is seen in Computer Science. Here, too, special topics in DS, such as #9, play only a minor role.

A comparatively stronger technical focus is reflected in the Computer Science portfolio (Figure 4). Basic and advanced computer science (#3 and #4) are by far the most frequently and intensively represented. As a counterexample to the previous portfolios, however, #12 “Business Intelligence” is hardly represented and treated only superficially.

In total, the portfolios clearly outline the central topics through which DS competencies are integrated into study programs. The results illustrate that this currently happens primarily via classical mathematics and statistics as well as basic and advanced computer science. However, it is much more interesting to see which contents play a minor to almost negligible role in comparison, which could ultimately trigger an interesting discourse on future developments. More specialized or advanced DS competencies, such as data ethics or data visualization, tend to be smaller niches across all study programs that are offered comparatively rarely and in less depth.

5. Discussion

5.1. Outline of the Status Quo

The main contribution of this article is to propose a robust approach for assessing the integration of DL and DS competencies in terms of distribution and intensity. Our corresponding study, descriptive in nature, sheds light for the first time on the status quo of integrating DL and DS in different contexts of higher education. The examination, differentiated by study programs (see *RQ 1*) and types of universities (see *RQ 2*), provides the necessary means of comparison and, thus, analytical depth for the following discussion of this status quo. A summary of four key findings will serve as a discussion:

(1) *The scope and depth of DS competencies increase with the degree of technical orientation*: The positioning of Business & Information Systems Engineering as a cross-disciplinary field between Business Administration and technical Computer Science (see also van der Aalst et al. 2018) is reflected clearly in the results. On average, 60.4% of the Business & Information Systems Engineering modules contain topics related to DL and DS (Business Administration: 32.6%; Computer Science: 79.6%). Additionally, more in-depth coverage of the topics according to the increasing technical degree of the study program can also be observed, such as the density of the documentation of #7 “Data Governance” or #11 “Machine/Deep Learning.” In general, this insight is not surprising. However, these results are a first step toward determining the current status quo and thereby providing a basis for public discourse. A guiding question could be, for example, which topics are relevant for classical, typically less technically oriented business programs and consequently should be integrated more extensively. Additionally, a comparison with future states via a longitudinal study would make progress clear. In this regard, however, it is deliberately not part of our study to judge the extent and depth of the specific DS competencies taught, for example, whether these competencies are sufficient for broad integration in nontechnical study programs.

(2) *DS skills are taught primarily through foundational courses in mathematics & statistics and computer science*: An examination of the exact competencies taught across all study programs reveals that the handling of data is primarily covered by general basics in mathematics and statistics as well as computer science. This is in line with the minimum requirements of common competency frameworks, such as the EDSF (EDISON Consortium 2018) or the DS Knowledge Framework of the IADSS (Fayyad and Hamutcu 2020). But in fact, DS is much more than that (Schwab-McCoy, Baker, and Gasper 2021). The more specialized topics of a DS curriculum (GI 2019), such as data ethics and data privacy, are a niche subject in isolated modules (see the following argument).

(3) *Specific DS competencies are taught only marginally and in isolation in individual modules*: The results provide further potential for discussion concerning the opinion that the purely technical handling of data is not considered sufficient; among other things, a greater awareness of the critical handling of data is required (GI 2018; Song and Zhu 2017). This involves, for example, an awareness of ethical as well as legal boundaries for the storage, processing and analysis of data. Topics such as data ethics and data privacy are therefore considered important educational content (GI 2019). After screening these more specialized modules, our study shows that the teaching of these competencies, such as data ethics, takes place almost exclusively in isolated, mostly elective modules and is less integrated into modules on other topics, such as those like marketing or market research. A provocative question could now be whether such issues, which are fundamental for dealing with data, are given enough space, especially in the classical business curricula. A corresponding discourse on the integration of such topics in business education, as has been and is being conducted for the topic of sustainability (see, e.g., Rusinko 2010), should also be carried out for the topics of DS and DL.

(4) *The integrated DS competencies do not show overall significant differences by type of university:* The two differentiated types of universities (UNI and UAS) can be considered comparison groups within our analysis of module handbooks. Based on the results, it is apparent that the level of integration of DS competencies is largely homogeneous by type of university. First, although the average number of offered modules varies by study program and type of university (e.g., UNI: 84.8 and UAS: 64.80; see Table 1), the general proportions (%) of modules with relevant DS content are again very comparable (see Table 3). Second, with specific exceptions in individual DS categories, there are no significant differences in the proportion or density of DS competencies that are taught evidently between types of universities. In sum, according to these comparison groups, these results emphasize the significance of the comparison between the study programs. Uncovering this homogeneity is also an interesting finding for higher education research. However, an as yet unanswered question for follow-up research might be how the teaching of DS competencies varies by type of university. Whether, for example, the practical application orientation of the UAS is more pronounced in this context remains unanswered.

Our study does not claim to make any specific recommendations for the design of DS curricula (e.g., with regard to the thematic weighting). Rather, this is to be left to the discourse based on these descriptive results. In this context, as shown, study programs vary widely and have different requirements for the integration of DS skills. Here, the personas proposed by the GI (2019) can support a targeted discourse by indicating characteristics of teaching content (e.g., what is mandatory or voluntary to be learned by whom). In summary, this status quo can provide directions for further discourse and comparative tracking in the sense of a longitudinal study. This is certainly also interesting to evaluate the adaptation of different curricula (such as EDSF, DSEF) and their specific content.

5.2. Implications for Research

The results of our study provide specific implications for further research. Four research perspectives are outlined below.

First, the idea and the design of our study are intended to enable longitudinal tracking as well as comparative analyses. The current results represent only the status quo of the integration of DS, which will certainly be expanded through ongoing and future efforts. Regular tracking of progress—in thematic direction and its scope—is possible, since on the one hand, module handbooks are continuously updated, and on the other hand, the measurement based on the DSD is repeatable in its basic characteristics. In addition, the possibility of further comparative analyses is interesting. An international comparison of higher education teaching could make national progress transparent and, among other things, the effect of national funding programs assessable.

Second, and in the same vein, the DSD can and should nevertheless also be further developed. In fact, “the ever-evolving nature of [DS]” (Schwab-McCoy, Baker, and Gasper 2021, p. 40) is a challenging aspect of competency integration. The keywords characterizing the DSD should be revised and extended since new subject areas (and also technologies) will certainly emerge and need to be mapped.

Third, module handbooks are a valuable and publicly available source of information about higher education teaching. It is therefore surprising that they have rarely been used in research. The research stream of curriculum mining (Kawintiranon et al. 2016; Aldowah, Al-Samarraie, and Fauzy 2019) is thus comparatively young and offers exciting research prospects. The juxtaposition of current topics in higher education curricula with the currently required competencies in practical job advertisements would allow a useful contrast of teaching and practice (Xun, Gottipati, and Shankararaman 2015). Moreover, the nature of module documentation has not yet been explored, for example, in terms of the metrics that characterize them, such as scope or heterogeneity of documentation.

Fourth, research into the higher education landscape could further elaborate on the provided insights. Based on the results, a theory-building explanation can be developed on how a new core topic—as DS undoubtedly is—is incorporated and established in existing teaching portfolios. Further, characterizing differences between types of universities could be explored more deeply. In this context, it can be particularly mentioned, that the applied keyword-based measurement so far does not consider contexts of the measured concepts. It would be interesting to explore to what degree of competence the concepts are expressed (such as “understand” vs. “apply”; according to Bloom’s taxonomy, (1956)). In particular, this investigation would contribute to the further juxtaposition of UNIs and UASs.

5.3. Implications for Higher Education

The results of our study also have implications for the practice of higher education, especially when it comes to optimizing study programs and their curricula. First and most fundamentally, the description of the status quo provides stimuli for integrating DS into study programs, such as business study curricula. If data is the “new oil” of the business world, business graduates cannot leave its handling solely to technically oriented experts, who in turn are likely to have less understanding of business. So it is certainly relevant to discuss what data skills, for example business graduates, should have and to what degree these skills are already included in the current study programs. In doing so, the thematic structure of our study, with its 12 thematic categories, also emphasizes an awareness of the complex scope of this endeavor. Second, and building on this, the results of our study provide a platform for topic-specific benchmarking against averages across the higher education landscape. A comparative evaluation of a university’s study program can provide the potential for the targeted further development of curricula if, for example, thematic gaps become apparent. The benchmark opportunity also allows students to compare the different study programs and get an overview of the main contents of an individual program.

5.4. Limitations

The limitations of our study can be discussed concerning two points: database and methodology. Regarding the first, the database that was used (more than 13,500 module descriptions from three study programs) has a limiting effect on the results. Although module handbooks generally follow a largely standardized structure, there is nevertheless a wide variation in the way the content is documented. Not every module is described in the same way in terms of scope, detail, or completeness. For example, two modules may describe similar content but use different words and different amounts of words. Similarly, it is known that there exist curricular contents that are not officially documented, but are taught informally as "Informal Curriculum" (Caza and Brower 2015). This heterogeneity would be particularly influential in exploratory (statistical) analyses (such as topic modeling). However, the deductive approach of our analysis, identification of central keywords, should accommodate this variation to a greater extent.

The methodology of deductive, dictionary-based content analysis has proven to be an efficient means of processing the large data source as well as a large number of keywords. The technique's advantage, however, also contains its limitations (Beattie and Thomson 2007). The applied methodology of deductive content analysis is always associated with potential limitations. In particular, the use of a dictionary as a measurement tool is associated with restrictions in terms of validity (Krippendorff 2019). Although the thematic categories can be considered meaningful, the selection of the keywords themselves has considerable influence on the results of this analysis. Other keywords or other compositions could certainly generate different results. Likewise, a dictionary can never be complete concerning the composition of the keywords. Nevertheless, to ensure the greatest possible validity, our study was based on an elaborate development process. The thematic categories are based on a well-founded DS curriculum (GI 2019) and the allocated keywords were validated using an iterative Delphi study with experts. However, a triangulation of this research approach, for example, through a survey among study program directors, would help to confirm the validity.

6. Outlook

The handling of data will undoubtedly be the central competence of the future (World Economic Forum 2019). In this context, the integration of DL and DS competencies into the breadth of higher education teaching—not only in technology-related study programs—will be a central pillar of maintaining or expanding competitive advantages (University Education Forum on Digitization 2017). To provide this challenge with a platform for discourse, the results of our study offer, for the first time, a comprehensive insight into the status quo of these integration efforts in the German higher education landscape.

A further discourse based on the results of our study might not only evaluate this status quo concerning the future viability of higher education teaching but also make specific suggestions for its targeted further development. In this regard, the repeatable design of this study has the potential to uncover corresponding progress via a longitudinal investigation that tracks not only the thematic direction but also the extent and intensity of the integration of DS and DL competencies into higher education.

A. Appendix

Title: Machine Learning	
Module Coordinator: N.N.	**Recommended Semester:** 6th semester
Credits: 9 ECTS	**Duration:** 1 semester
Workload: 150 h	**Course Frequency:** Each semester
Admission Requirements: None	**Language:** English
Required Knowledge: Knowledge of statistics, programming basics, and databases	

Learning Outcomes:

On successful completion, students will be able to

- understand the basic principles of machine learning.
- understand concepts and techniques of machine learning and evaluate their potentials and limitations compared to alternative approaches.
- implement common machine learning algorithms like principal component analysis, linear regression, logistic regression, common classification techniques, k-means, and neural networks.
- implement machine learning solutions with Python.
- solve practical DS problems using machine learning techniques.
- …

Module Contents:

- Introduction to Machine Learning
- Types of Machine Learning
  - Supervised Learning
  - Unsupervised Learning
  - Semi-Supervised Learning
  - Reinforcement Learning
- Notation and Definitions
- Machine Learning Algorithms
  - Linear Regression
  - Logistic Regression
  - Decision Tree Learning
  - K-Means Clustering
  - …
- …

Module Exam Type: Exam (90 minutes)

Literature:

Murphy, K. P. (2012). Machine learning: A probabilistic perspective. MIT Press.

Müller, A. C., & Guido, S. (2016). Introduction to machine learning with Python: a guide for data scientists. O'Reilly Media, Inc.

Figure A1. Exemplary module description.

Table A1. Data Science Dictionary (DSD).

#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
Basics Mathe-matics & Statistics	**Advanced Mathe-matics & Statistics**	**Basic Computer Science**	**Advanced Computer Science**	**Crypto. & Data Security**	**Data Ethics & Data Privacy**	**Data Gover-nance**	**Data Integration**	**Data Visuali-zation**	**Data Mining**	**Machine Learning/ Deep Learning**	**Business Intelligence**
Algebra	Approximation	ACID	32-bit architecture	AES	Access control	Capability Maturity Model Integration	Anomaly detection	Annotation	Advanced analytics	A/B testing	Analytical information system
Analysis	Asymptotic theory	Algorithm	64-bit architecture	Anonymization	BDSG	CMMI	API	Chart types	A-priori algorithm	Accuracy	BICC
Analysis of variance	Bayesian statistics	Arrays	Agent system	Blockchain	BSI	CMS	Application layer	Color coding	AR model	AI	Business activity monitoring
ANOVA	Bayes's statistics	ASCII	Application cache	Chiffre	Copyright	COBIT	Application programming interface	Color theory	ARIMA	Analytics framework	Business analytics
Binomial distribution	Calculus	Binary search	Arithmetic logic unit	Computer crime	Data compliance	Content management system	Communication autonomy	Coloring	ARMA	Artificial intelligence	Business intelligence
Binomial formulas	Cauchy sequence	Binary tree	Automatic synthesis	Cryptography	Data ethics	Corporate data	Crawler	Computer graphics	ARMAX	Artificial neural networks	Dashboard
Boolean algebra	Chi-square	Bitmap index	Big Data	Cyber security	Data privacy	CRC	Crawling	Data representation	Association analysis	Bagging	Dashboarding
Central limit theorem	Concordance measures	Bogo-sort	Boot-strapping	Data security	Data protection	Cyclic redundancy check	Dat heterogeneity	Data visualization	Association rule	Bayesian networks	Data analysis
Combinatorics	Contingency analysis	Bubble-sort	C#	Digital signature	Data vision	DAMA-DMBOK	Data acquisition	Diagram types	Autoregression	Boosting	Data cube
Confidence interval	Contingency table	Bucket-sort	C++	EFS	DSGVO	Data culture	Data aggregation	Editorial thinking	Autoregressive models	Branch-and-Bound	Data mart
Convergence	Differential calculus	CAP theorem	Cache coherence	Elliptic curves	Federal Data Protection Act	Data driven company	Data cleaning	Edward Tufte	Cluster	Branch-and-Cut	Data warehouse
Coordinate system	Differential equations	Coding	Cache consistency	Encrypting file system	Federal Office for Security and Information Technology	Data governance	Data cleansing	Graph drawing	Cluster analysis	C4.5	Decision engineering
Correlation coefficient	Discriminant analysis	Comb-sort	CEP	Enigma	General data protection regulation	Data governance framework	Data collection	Graphic communication	Clustering	C5.0	Decision support system
Covariance	Error probability	Comma-separated values	Cloud	Extension body	Information ethics	Data Governance Institute	Data combination	Graphic representation	Concept extraction	Caffe	Digital boardroom
Curve discussion	Exploratory statistics	Communication technology	Cloud computing	Firewall	Internet law	Data governance role	Data compression	Infographic	Cranfield paradigm	CART	Enterprise Resource Planning
Cyclic component	Exponential distribution	Computer architecture	Cloud dataflow	Forensic data analysis	ISO 27000	Data maintenance	Data consolidation	Information design	CRISP-DM	Case based reasoning	ERP
Density function	Exponential function	Computer networks	Cloud management	Hash function	ISO/IEC 38500	Data management	Data control	Information graphics	Cross Industry Standard Process for Data Mining	CHAID	Executive information system
Descriptive statistics	Exponential smoothing	Computer science	Computational linguistics	HMAC algorithm	IT audit	Data model	Data coupling	Information visualization	Cross-validation	Classification	HOLAP
Dispersion measure	Extrapolation	Constraints	Container management	IDEA	IT basic protection catalog	Data modeling	Data curation	Interaction design	Data analytics	Classification analysis	Management information system

Table A1. Continued.

#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
Distribution function	Factor analysis	Counting-sort	Data stream system	Information security	IT compliance	Data Ops	Data economy	Interactive visualization	Data classification	Computational learning theory	Management support systems
Equation system	F-test	CPU	Digital twin	Integrity measurement	IT law	Data organization	Data engineering	Knowledge visualization	Data mining	Confusion matrix	MOLAP
Expected value	Goodness of fit test	CSV	Dijkstra algorithm	Internet key exchange	Smart data	Data policy	Data enrichment	Perception theory	Data pattern recognition	Decision tree	OLAP
Frequency distribution	Graph algorithm	Data type	Distributed systems	Internet security	Social network	Data quality	Data evaluation	Preattentive attributes	DBSCAN	Deep learning	OLTP
Gauss algorithm	High dimensional statistics	Database	Embedded PC	Intrusion detection system	Traffic data	Data quality scorecard	Data exchange	Sammon illustration	Descriptive analytics	Dimensionality reduction	Online analytical processing
Geometry	Inductive statistics	Database systems	Embedded system	Intrusion prevention system		Data sovereignty	Data exploration	Statistical graphics	Diagnostic analytics	DNN	Online transaction processing
Graph theory	Inferential statistics	Debugger	Fragmentation	Irreducible polynomials		Data strategy	Data export	Theory of colors	Dimensionality reduction	EDISON Data Science Framework (EDSF)	Pivot
Hypothesis testing	Interference statistics	Debugging	Fuzzy systems	IT forensics		Data structure	Data filtering	Visual abstraction	ETS	EM algorithm	Power BI
Index numbers	Kolmogorov-Smirnov	DevOps	Geo-database	IT security		Database schema	Data flow	Visual analytics	Exploratory data analysis	Entropy	Reporting
Integral	Least squares	Disjunctive normal form	Graph database	Kerberos protocol		DMS	Data harmonization	Visual design	Exponential Smoothing	F1-score	ROLAP
Integral calculus	Least squares estimator	Entities	Hadoop	PKI standards		Document management system	Data hub	Visual discovery	Feature extraction	Gated recurrent unit	Self-Service BI
Interpolation	Likelihood	Entity relationship model	Hyper-threading	Polynomial residue classes		ECM	Data independence	Visual mapping	Feature reduction	Gaussian processes	Storyboarding
Interval estimation	Linear optimization	ER modeling	IaaS	Pseudonymization		Enterprise content management	Data integration	Visual perception	Forecasting	Generalized additive model	Storytelling
Least squares method	Linear regression models	Gnome-sort	IDE	Public key infrastructure		GARP	Data integrity	Visual thinking	Fuzzy logic	Genetic algorithm	Tableau
Limit theorem	Logarithm function	GPU	Infrastructure-as-a-Service	RC4		Information management	Data lake	Visualization	Graph mining	Greedy algorithm	
Linear equation systems	MANOVA	Graphics processor	In-Memory	S/MIME		ISO/IEC 38500	Data level	Visualization mapping	Hashimoto matrix	Hidden markov model	
Linear regression	Maximum likelihood	Hard disk drive	Integrated development environment	Security by design		ISO/TC 215	Data literacy		Information extraction	Instance based method	
Logarithm	Method of moments	Hardware	Internet of Things	Security protocols		IT governance	Data manipulation		Information retrieval	Kernel	
Matrices	Monte carlo	HDD	IoT	Single-sign on		MaRisk	Data mapping		KDD	KNN	
Matrix calculation	Multivariate distribution	Heap-sort	ITIL	Smart cards		Master data	Data migration		k-means	LASSO	
Matrix decomposition	Multivariate normal distribution	Hexa-decimal system	Java	SSL handshake		Master data management	Data pipeline		k-nearest-neighbor	Learning transfer	
Natural numbers	Multivariate statistics	HTML	Map-reduce	SSL/TLS		MDDBMS	Data pre-integration		Knowledge discovery	Logistic regression	

Table A1. Continued.

#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
Newton's method	Neyman-pearson-lemma	Hypertext Markup Language	Mass data	Stream ciphers		MDM	Data preparation		Knowledge discovery in databases	Machine learning	
Neyman-Pearson	Nonlinear regression	ICT	Memory address	Triple DES		Meta data	Data pre-processing		Knowledge extraction	Marcov chain monte carlo	
Normal distribution	Nonparametric statistics	Index structures	Memory hierarchy	Trusted platform		Meta data management	Data profiling		Knowledge retrieval	Markov decision process	
Numerics	Regression models	Information technology	Microcontroller	Zero-knowledge protocol		Meta model	Data provision		Latent dirichlet allocation	MCMC	
Parameter estimation	Resampling	Inheritance	Micro-processor			Multi-dimensional database management system	Data quality		Latent semantic analysis	MDP	
Pearson	Residual analysis	Insert-sort	NoSQL			Non-invasive data governance framework	Data redundancy		Louvain method	ML flow	
Poisson distribution	Robust statistics	Interpolation search	ODBC			SAS Data Governance Framework	Data request		MA model	Model evaluation	
Polynomials	Sensitivity analysis	IT concept	Ontology				Data scraping		Moving average	Model quality	
Power function	Simplex	IT infrastructure	Open database connectivity				Data segmentation		Network effect	Model validation	
Probability calculation	Simulation methods	IT systems	OWL				Data selection		Outlier analysis	Naive bayes	
Probability distribution	Simulation procedures	Java	OX Path				Dataset		Outlier detection	Natural language processing	
Probability theory	Six Sigma	Javascript	PaaS				Data source		Page rank	Neural networks	
Propositional Logic	Spatial statistics	JSON	Parallelization				Data stock		Parameterization	NLP	
p-value	Statistical interference	Knowledge management	PHP				Data storage		Pattern recognition	OBDD	
Random variable	Substitution principle	Linear search	Platform-as-a-Service				Data synchronization		Principal component analysis	OpenNN	
Random vectors	Sufficiency	Loops	Power-PC				Data transfer		Process mining	Overfitting	
Rational numbers	UMVUE estimator	Memory (RAM)	Reference architectures				Data transformation		SARIMA	Parameter optimization	
Real numbers	Varimax	Network architecture	RISC-V				Data transmission		SARIMAX	Parameter tuning	
Sample	Vector analysis	Non-volatile memory	SaaS				Data transport		Semantic web	PMML	
Sampling		Octadecimal system	Search algorithms				Data understanding		Sentiment analysis	Precision	
Scatter plot		O-notation	Search engines				Data workflow		Sequence mining	Predicate logic	
Seasonal component		Operating system	Shared memory				Data wrangling		Shopping cart analysis	Predictive analytics	
Set theory		Operators	Shift-and-algorithm				Datalog		Smoothing models	Predictive maintenance	
Similarity measure		Polymorphism	Simultaneous multithreading				Decoding		Social media mining	Predictive model	
Spearman		Processor	Skip-search algorithm				Design autonomy		Spatial clustering	Predictive model markup language	

Table A1. Continued.

#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
Standard deviation		Programming	Software-as-a-Service				Digital literacy		Text mining	Predictive models	
Standard normal distribution		Python	SPARC				EAI		Time series analysis	Prescriptive analytics	
Statistical inference		RDBMS	Spatial data				EDI		Volatility forecasting	Probabilistic models	
Statistics		Recursion	SQL				EII		Web content mining	Probabilistic reasoning	
Stochastics		Select-sort	Streaming				Electronic data interchange		Web mining	Probability density function	
Surface area		Sequence analysis	Swap file				ELT		Web structure mining	Probably Approximately Correct Learning	
Taylor series		Signal-slot-concept	Thread-local storage				ETL		Web usage mining	Pruning	
Test level		Snowflake scheme	Virtual machines				EventLog data			Random forrest	
Trigonomic function		software development	Virtual memory management				Extract, transform, load			Recall	
t-test		Software engineering	Virtual shared memory				Extraction			Recurrent models	
Uniform distribution		Solid-state drive (SSD)	Virtualization				Feature engineering			Regression methods	
Univariate		Solid-state storage	VxWorks				Feature selection			Reinforcement learning	
Urn model		Splay tree	x86 architecture				Geodata			Robotics	
Vectors		SSD	YARN				Hash method			ROC curve	
		Star schema					Hashing			Scikit-learn	
		Strings					Image compression			Self-organizing map	
		UML					Image processing			Self-supervised learning	
		Unified Modeling Language					Information aggregation			Sorting algorithm	
		VBA					Information exploration			Spatial clustering	
		Visual Basic for Applications					Information hub			Supervised learning	
							Information integration			Support vector machine	
							Information logistics			SVD	
							Information logistics			SVM	
							Information mapping			Swarm intelligence	
							Information needs analysis			TensorFlow	
							Information preparation			Test data	

Table A1. Continued.

#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
							Information pre-processing			Training data	
							Information procurement			Treemap	
							Information resource			Underfitting	
							Information source			Unsupervised learning	
							Information supply			Validation data	
							Integration architecture				
							Labeling				
							Logging				
							Metrics				
							Partition				
							Partitioning				
							Pipelining				
							Presentation layer				
							Query processing				
							Randomization				
							Record matching				
							Record merging				
							Schema alignment				
							Segmentation				
							Semantic heterogeneity				
							Semantic integration				
							Sensor data				
							Structural heterogeneity				
							Syntactic heterogeneity				
							Technical heterogeneity				
							Tuple				
							Virtual mediated schema				
							Web scraping				

Table A2. Data science dictionary (DSD)—coverage of keywords.

#	Category	Keywords (#)	Coverage (#)	Coverage (%)
1	Basics Mathematics & Statistics	80	75	94%
2	Advanced Mathematics & Statistics	60	45	75%
3	Basic Computer Science	86	65	76%
4	Advanced Computer Science	80	56	70%
5	Cryptography & Data Security	45	32	71%
6	Data Ethics & Data Privacy	24	20	83%
7	Data Governance	48	30	63%
8	Data Integration	120	74	62%
9	Data Visualization	35	17	49%
10	Data Mining	73	39	53%
11	Machine Learning/Deep Learning	97	77	79%
12	Business Intelligence	33	28	85%
	Total	**781**	**558**	**71%**

Table A3. Overview of the Delphi Study.

	Raw DSD	Round 1			Intermediate DSD	Round 2			Final DSD
	[#]	Agreement rate [%]	Canceled [#]	Added [#]	[#]	Agreement rate [%]	Canceled [#]	Added [#]	[#]
Category #1	93	64.52	18	0	75	96.00	1	6	80
Category #2	65	86.15	4	0	61	93.44	1	0	60
Category #3	115	26.96	72	3	46	97.83	12	52	86
Category #4	66	50.00	24	3	45	95.56	3	38	80
Category #5	59	49.15	23	3	39	97.44	4	10	45
Category #6	34	47.06	17	1	18	100.00	2	8	24
Category #7	49	79.59	5	0	44	100.00	4	8	48
Category #8	125	94.40	53	3	75	95.87	3	48	120
Category #9	41	46.34	6	0	35	100.00	0	0	35
Category #10	57	100.00	0	0	57	100.00	0	16	73
Category #11	96	98.96	29	8	75	97.92	0	22	97
Category #12	33	75.76	0	0	33	100.00	0	0	33
Total	833	68.24	251	21	603	97.84	30	208	781

Table A4. Proportion (%) of DS categories within DS modules.

	Business Administration		Business & Information Systems Engineering		Computer Science	
	UAS	UNI	UAS	UNI	UAS	UNI
1 Basics Mathematics & Statistics	19.6		9.3		12.4	
			p <0.001***			
	26.1	17.4	9.5	8.9	11.0	14.1
	$p = 0.01$*		$p = 0.53$		$p = 0.01$*	
#2 Advanced Mathematics & Statistics	11.3		4.7		4.4	
			p <0.001***			
	10.5	13.5	4.5	5.1	3.8	5.2
	$p = 0.16$		$p = 0.42$		$p = 0.02$*	
#3 Basic Computer Science	18.2		32.8		40.6	
			p <0.001***			
	18.6	16.8	32.8	32.9	41.4	39.7
	$p = 0.44$		$p = 0.98$		$p = 0.37$	
#4 Advanced Computer Science	9.5		10.3		10.4	
			$p = 0.54$			
	9.6	9.1	10.9	8.9	11.3	9.3
	$p = 0.07$		$p = 0.03$*		$p = 0.03$*	
#5 Cryptography & Data Security	1.5		2.9		2.9	
			p <0.001***			
	1.5	1.7	3.3	2.1	3.7	1.9
	$p = 0.73$		$p = 0.08$		p <0.001***	
#6 Data Ethics & Data Privacy	2.4		3.4		2.3	
			$p = 0.03$*			
	2.7	1.7	3.9	2.5	3.2	1.2
	$p = 0.24$		$p = 0.06$		p <0.001***	
#7 Data Governance	7.2		10.9		7.0	
			p <0.001***			
	7.5	6.1	10.6	11.5	6.8	7.2
	$p = 0.24$		$p = 0.41$		$p = 0.51$	
#8 Data Integration	8.4		6.6		6.9	
			$p = 0.09$			
	8.8	7.3	7.2	5.5	7.1	6.8
	$p = 0.35$		$p = 0.08$		$p = 0.61$	
#9 Data Visualization	2.6		1.8		2.4	
			$p = 0.17$			
	3.2	0.7	2.0	1.3	2.5	2.2
	p <0.001***		$p = 0.14$		$p = 0.48$	
#10 Data Mining	3.9		3.1		2.0	
			p <0.001***			
	3.3	5.6	2.8	3.7	1.6	2.5
	$p = 0.04$*		$p = 0.28$		$p = 0.02$*	
#11 Machine Learning/ Deep Learning	1.9		3.5		4.8	
			p <0.001***			
	1.7	2.7	2.7	5.2	4.2	5.6
	$p = 0.31$		$p = 0.03$*		$p = 0.02$*	
#12 Business Intelligence	13.4		7.2		1.9	
			p <0.001***			
	15.1	8.7	8.1	5.5	1.6	2.1
	p <0.001***		$p = 0.01$*		$p = 0.23$	
Total [%]	**100**		**100**		**100**	

NOTE. p-values: p >0.05; p <0.05*; p <0.01**; p <0.001***.

Table A5. Top 25 keywords in each study program.

Business Adminstration				Business & Information Systems Engineering				Computer Science			
#	Term	Frequency	Occurence (%) N = 1,565	#	Term	Frequency	Occurence (%) N = 2,315	#	Term	Frequency	Occurence (%) N = 4,116
1	ERP	344	6,6%	1	DATABASE	1,124	11,8%	1	COMPUTER SCIENCE	3,398	33,4%
2	DATABASE	323	7,7%	2	ALGORITHM	1,068	16,9%	2	ALGORITHM	2,509	23,1%
3	STATISTICS	287	9,5%	3	DATA PROCESSING	769	19,0%	3	DATABASE	1,519	8,1%
4	C#	269	11,0%	4	PROGRAMMING	640	14,7%	4	PROGRAMMING	1,459	18,2%
5	DATA ANALYSIS	183	9,3%	5	ERP	486	7,0%	5	DATA STRUCTURE	722	8,0%
6	ANALYSIS	145	5,8%	6	DATA STRUCTURE	381	7,9%	6	OPERATING SYSTEM	716	6,5%
7	MATHEMATICS	135	6,1%	7	OPERATING SYSTEM	369	5,3%	7	MATHEMATICS	461	6,3%
8	REPORTING	118	5,2%	8	SQL	301	5,0%	8	HARDWARE	453	6,6%
9	PROCESSOR	114	4,8%	9	SOFTWARE DEVELOPMENT	260	7,4%	9	SOFTWARE DEVELOPMENT	432	6,2%
10	PROGRAMMING	109	3,8%	10	DATA MODEL	251	5,7%	10	ALGEBRA	387	5,4%
11	INFORMATION TECHNOLOGY	102	5,6%	11	UML	243	6,3%	11	UML	341	4,2%
12	INFORMATION MANAGEMENT	101	3,4%	12	INFORMATION MANAGEMENT	193	3,9%	12	SQL	330	3,4%
13	ALGEBRA	99	4,0%	13	ALGEBRA	175	4,9%	13	C#	325	5,1%
14	PROBABILITY CALCULATION	95	3,8%	14	STATISTICS	173	3,9%	14	VEKTOR	305	4,4%
15	RANDOM VARIABLE	95	3,5%	15	IT SYSTEM	165	4,6%	15	ANALYSIS	285	4,3%
16	ALGORITHM	95	3,6%	16	DATA MINING	160	2,7%	16	DISTRIBUTED SYSTEMS	285	3,1%
17	DIFFERENTIAL CALCULUS	88	3,6%	17	DISTRIBUTED SYSTEMS	159	3,0%	17	INTEGRAL	276	3,2%
18	INTEGRAL	87	3,3%	18	DATA WAREHOUSE	153	3,1%	18	PROCESSOR	272	3,9%
19	LINEAR OPTIMIZATION	83	3,9%	19	SOFTWARE ENGINEERING	153	3,4%	19	DATA TYPE	236	3,9%
20	HARDWARE	81	2,8%	20	C#	151	3,3%	20	DATA MODEL	225	2,8%
21	DATA MODEL	81	3,3%	21	VECTOR	148	3,3%	21	DATA VISUALIZATION	215	2,6%
22	KNOWLEDGE MANAGEMENT	80	2,7%	22	BUSINESS INTELLIGENCE	147	2,7%	22	IMAGE PROCESSING	212	2,5%
23	LINEAR EQUATION SYSTEM	79	3,3%	23	DATA TYPE	136	4,3%	23	DIFFERENTIAL EQUATION	198	2,4%
24	DESCRIPTIVE STATISTICS	78	3,3%	24	INTEGRAL	129	3,1%	24	STATISTICS	189	2,4%
25	DATA VISUALIZATION	69	2,9%	25	PROCESSOR	127	3,9%	25	SOFTWARE ENGINEERING	189	2,6%

Supplementary Materials

The supplementary material includes the developed *Data Science Dictionary* (DSD).

Data Availability Statement

The data and code that support the findings of this study are openly available on osf.io at *https://osf.io/wsy72/?view_only=5cba426ffd574544aeb449663acb08f6*.

Disclosure Statement

The authors report there are no competing interests to declare.

References

Aldowah, H., Al-Samarraie, H., and Fauzy, W. M. (2019), "Educational Data Mining and Learning Analytics for 21st Century Higher Education: A Review and Synthesis," *Telematics and Informatics*, 37, 13–49. DOI: 10.1016/j.tele.2019.01.007

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., and Wittrock, M. C., eds. (2001), *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, New York: Longman.

Anderson, P., Bowring, J., McCauley, R., Pothering, G., and Starr, C. (2014), "An Undergraduate Degree in Data Science," in *Proceedings of the 45th ACM Technical Symposium on Computer Science Education*, eds. J. D. Dougherty, K. Nagel, A. Decker and K. Eiselt, pp. 145–150, ACM Press. DOI:10.1145/2538862.2538936

Beattie, V., and Thomson, S. J. (2007), "Lifting the Lid on the Use of Content Analysis to Investigate Intellectual Capital Disclosures," *Accounting Forum*, 31, 129–163. DOI:10.1016/j.accfor.2007.02.001

Bergsmann, E., Schultes, M.-T., Winter, P., Schober, B., and Spiel, C. (2015), "Evaluation of Competence-Based Teaching in Higher Education: From Theory to Practice," *Evaluation and Program Planning*, 52, 1–9. DOI:10.1016/j.evalprogplan.2015.03.001

Bichler, M., Heinzl, A., and van der Aalst, W. M. (2017), "Business Analytics and Data Science: Once Again?," *Business & Information Systems Engineering*, 59, 77–79. DOI:10.1007/s12599-016-0461-1

Blömeke, S., Zlatkin-Troitschanskaia, O., and Kuhn, C. and Fege, J., eds. (2013), *Modeling and Measuring Competencies in Higher Education: Tasks and Challenges*, Rotterdam: Sense Publishers.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., and Krathwohl, D. R. (eds.) (1956), *Taxonomy of Educational Objectives: Cognitive Domain*, New York: Longman.

Bradley, L., and Stewart, K. (2003), "A Delphi Study of Internet Banking," *Marketing Intelligence & Planning*, 21, 272–281. DOI:10.1108/02634500310490229

Buttner, P., and Vocke, C. (2004), "Modularisierung Von Studiengängen: Grundsatzüberlegungen [Modularization of Study Programs: Basic Considerations]," *Beiträge Zur Hochschulforschung*, 26, 6–27.

Carpenter, C. D., Ray, M., and Bloom, L. A. (1995), "Portfolio Assessment: Opportunities and Challenges," *Intervention in School and Clinic*, 31, 34–41. DOI:10.1177/105345129503100106

Caza, A., and Brower, H. H. (2015), "Mentioning the Unmentioned: An Interactive Interview about the Informal Management Curriculum," *Academy of Management Learning & Education*, 14, 96–110. DOI:10.5465/amle.2013.0164

Cleveland, W. S. (2001), "Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics," *International Statistical Review / Revue Internationale de Statistique*, 69, 21–26. DOI: 10.1111/j.1751-5823.2001.tb00477.x

Dalkey, N., and Helmer, O. (1963), "An Experimental Application of the Delphi Method to the Use of Experts," *Management Science*, 9, 458–467. DOI:10.1287/mnsc.9.3.458

De Veaux, R., Agarwal, M., Averett, M., Baumer, B. S., Bray, A., Bressoud, T. C., Bryant, L., Cheng, L. Z., 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. X., Washington, T. M., Wesley, C. L., White, D., and Ye, P. (2017), "Curriculum Guidelines for Undergraduate Programs in Data Science," *Annual Review of Statistics and Its Application*, 4, 15–30. DOI:10.1146/annurev-statistics-060116-053930

Dhar, V. (2013), "Data Science and Prediction," *Communications of the ACM*, 56, 64–73. DOI:10.1145/2500499

EDISON Consortium. (2018), "EDISON Data Science Framework (EDSF)," accessed 30 August 2021. Available at *https://edison-project.eu/edison/edison-data-science-framework-edsf/*.

European Ministers of Education. (1999), "Der Europäische Hochschulraum: Gemeinsame Erklärung der Europäischen Bildungsminister [The European Higher Education Area: Joint Declaration of the European Ministers of Education]," Accessed 4 January 2021. Available at *https://www.hrk.de/fileadmin/redaktion/hrk/02-Dokumente/02-03-Studium/02-03-01-Studium-Studienreform/Bologna_Dokumente/Bologna_1999.pdf*.

Fayyad, U., and Hamutcu, H. (2020), "Toward Foundations for Data Science and Analytics: A Knowledge Framework for Professional Standards," *Harvard Data Science Review*, 2, 23. DOI:10.1162/99608f92.1a99e67a

Federal Ministry of Education and Research. (2015), "Bericht der Bundesregierung über die Umsetzung des Bologna-Prozesses: 2012 - 2015 in Deutschland [Report of the Federal Government on the Implementation of the Bologna Process: 2012 - 2015 in Germany]" accessed 4 January 2021.

Federal Ministry of Education and Research. (2020), "Forschung-Das Wissenschaftssystem-Forschung an Fachhochschulen [Research-The Scientific System-Research at Universities of Applied Sciences]," accessed 26 April 2021. Available at *https://www.bmbf.de/bmbf/de/forschung/das-wissenschaftssystem/forschung-an-fachhochschulen/forschung-an-fachhochschulen_node.html*.

Feldman, R., and Sanger, J. (2007), *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, Cambridge: Cambridge University Press.

Fichtl, A., and Piopiunik, M. (2017), "Absolventen von Fachhochschulen und Universitäten im Vergleich: FuE-Tätigkeiten, Arbeitsmarktergebnisse, Kompetenzen und Mobilität [Comparing Graduates from Universities of Applied Sciences and Universities: R&D Activities, Labor Market Outcomes, Competencies, and Mobility]," Studien zum deutschen Innovationssystem, 14.

Föll, P., and Thiesse, F. (2021), "Exploring Information Systems Curricula," *Business & Information Systems Engineering*, 63, 711–732. DOI:10.1007/s12599-021-00702-2

Fregin, M.-C., Richter, P., Schreiber, B., Wüstenhagen, S., Dietrich, J., Frankenberger, R., Schmidt, U., Walgenbach, P. (2016), "Hier Bitte Einen Satz zu Kompetenzen Einfügen…": Kompetenzorientierung, Gesellschaftliche Verantwortungsübernahme Und Homogenisierung in Universitären Curricula Am Beispiel Der Führungsverantwortung ["Insert a Sentence on Competencies Here, Please…": Competency Orientation, Social Responsibility and Homogenization in University Curricula Using the Example of Leadership]," *Das Hochschulwesen*, 64, 117–123.

German Informatics Society. (2018), "Data Literacy und Data Science Education: Digitale Kompetenzen in der Hochschulausbildung [Data Literacy and Data Science Education: Digital Competencies in Higher Education]," Policy Paper of the Presidential Task Force "Data Science" of the German Informatics Society. Accessed 16 June 2021. Available at *https://gi.de/fileadmin/GI/Hauptseite/Aktuelles/Aktionen/Data_Literacy/GI_DataScience_2018-04-20_FINAL.pdf*.

German Informatics Society. (2019), "Data Science: Lehr- und Ausbildungsinhalte [Data Science: Teaching and Training Content," Working Paper of the German Informatics Society. Accessed 16 June 2021. Available at *https://gi.de/fileadmin/GI/Allgemein/PDF/GI_Arbeitspapier_Data-Science_2019-12_01.pdf*.

German Rectors' Conference. (2021a), "Hochschultypen [Types of Universities]," Accessed 26 April 2021. Available at *www.hrk.de/themen/hochschulsystem/hochschulrecht/hochschultypen/*.

German Rectors' Conference. (2021b), "Studiengangssuche [Study Program Search]," Accessed 16 August 2021. Available at *https://www.hochschulkompass.de/studium/studiengangsuche.html*.

Green, B., Jones, M., Hughes, D., and Williams, A. (1999), "Applying the Delphi Technique in a Study of GPs' Information Requirements," *Health & Social Care in the Community*, 7, 198–205. DOI: 10.1046/j.1365-2524.1999.00176.x

Hardin, J., Horton, N. J., Nolan, D., and Lang, D. T. (2021), "Computing in the Statistics Curricula: A 10-Year Retrospective," *Journal of Statistics and Data Science Education*, 29, S4–S6. DOI: 10.1080/10691898.2020.1862609

Hartley, S. W., Frontczak, N. T., and Rudelius, W. (2015), "Portfolios as a Method of Assessment in an Integrated Curriculum," in *Proceedings of the 2000 Academy of Marketing Science (AMS) Annual Conference*, eds. H. E. Spotts and H. L. Meadow, pp. 324–331, Cham: Springer.

Hasson, F., Keeney, S., and McKenna, H. (2000), "Research Guidelines for the Delphi Survey Technique," *Journal of Advanced Nursing*, 32, 1008–1015. DOI:10.1046/j.1365-2648.2000.t01-1-01567.x

Hyde, K. F. (2000), "Recognising Deductive Processes in Qualitative Research," *Qualitative Market Research: An International Journal*, 3, 82–90. DOI:10.1108/13522750010322089

Initiative Future Skills. (2019), "The Initiative," accessed 8 June 2021. Available at *https://www.stifterverband.org/future-skills*.

Kauermann, G., and Seidl, T. (2018), "Data Science: A Proposal for a Curriculum," *International Journal of Data Science and Analytics*, 6, 195–199. DOI:10.1007/s41060-018-0113-2

Kawintiranon, K., Vateekul, P., Suchato, A., and Punyabukkana, P. (2016), "Understanding Knowledge Areas in Curriculum through Text Mining from Course Materials," in *Proceedings of 2016 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE). Piscataway, NJ: IEEE*, pp. 161–168. DOI:10.1109/TALE.2016.7851788

Kerres, M., and Schmidt, A. (2011), "Zur Anatomie Von Bologna-Studiengängen. Eine Empirische Analyse Von Modulhandbüchern [On the Anatomy of Bologna Degree Programs. An Empirical Analysis of Module Handbooks]," *Die Hochschule*, 2, 173–191.

Kirchherr, J., Klier, J., Lehmann-Brauns, C., and Winde, M. (2019), "Future skills: Which Skills are Lacking in Germany," accessed 2 April 2021. Available at *https://www.stifterverband.org/medien/which-skills-are-lacking-in-germany*.

Krippendorff, K. (2019), *Content Analysis: An Introduction to Its Methodology*, Thousand Oaks, CA: Sage Publications.

Kross, S., Peng, R. D., Caffo, B. S., Gooding, I., and Leek, J. T. (2020), "The Democratization of Data Science Education," *The American Statistician*, 74, 1–7. DOI:10.1080/00031305.2019.1668849

Landis, J. R., and Koch, G. G. (1977), "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, 33, 159–174. DOI:10.2307/2529310

Lirk, G., and Zumbach, J. (2015), *Das Bild der Wissenschaft: Ein Vergleich zwischen Fachhochschulen und Universitäten [The Image of Science: a Comparison Between Universities of Applied Sciences and Universities]*.

Manning, C., and Schütze, H. (1999), *Foundations of Statistical Natural Language Processing*, Cambridge, MA: MIT Press.

Mayring, P. (2004), "Qualitative Content Analysis," *A Companion to Qualitative Research*, 1, 159–176.

McKenna, H. P. (1994), "The Delphi Technique: A Worthwhile Research Approach for Nursing?," *Journal of Advanced Nursing*, 19, 1221–1225. DOI:10.1111/j.1365-2648.1994.tb01207.x

Mike, K. (2020), "Data Science Education: Curriculum and Pedagogy," in *Proceedings of the 2020 ACM Conference on International Computing Education Research*, in A. Robins, A. Moskal, A. J. Ko and R. McCauley, pp. 324–325. DOI:10.1145/3372782.3407110

Neuendorf, K. A. (2002), *The Content Analysis Guidebook*, Thousand Oaks: Sage Publications.

Ridsdale, C., Rothwell, J., Smit, M., Hossam, A.-H., Bliemel, M., Irvine, D., Kelley, D. R., Matwin, S., and Wuetherick, B. (2015), "Strategies and Best Practices for Data Literacy Education (Knowledge Synthesis Report)," Dalhousie University. DOI:10.13140/RG.2.1.1922.5044

Rusinko, C. A. (2010), "Integrating Sustainability in Management and Business Education: A Matrix Approach," *Academy of Management Learning & Education*, 9, 507–519. DOI:10.5465/amle.9.3.zqr507

Schwab-McCoy, A., Baker, C. M., and Gasper, R. E. (2021), "Data Science in 2020: Computing, Curricula, and Challenges for the Next 10

Years," *Journal of Statistics and Data Science Education*, 29, S40–S50. DOI:10.1080/10691898.2020.1851159

Society for Higher Education Research. (2021), "Ziele: Warum eine Gesellschaft für Hochschulforschung? [Why a Society for Higher Education Research?]", accessed 6 July 2021. Available at *https://www.gfhf.net/home/ziele/*.

Song, I. Y., and Zhu, Y. (2017), "Big Data and Data Science: Opportunities and Challenges of iSchools," *Journal of Data and Information Science*, 2, 1–18. DOI:10.1515/jdis-2017-0011

Standing Conference of the Ministers of Education and Cultural Affairs. (2000), Rahmenvorgaben für die Einführung von Leistungspunktsystemen und die Modularisierung von Studiengängen: Beschluss der Kultusministerkonferenz [Framework for the Introduction of Credit Point Systems and the Modularization of Study Programs: Resolution of the Standing Conference of the Ministers of Education and Cultural Affairs]. Accessed 4 January 2021. Available at *https://www.kmk.org/fileadmin/Dateien/pdf/PresseUndAktuelles/2000/module.pdf*

Standing Conference of the Ministers of Education and Cultural Affairs. (2010), "Ländergemeinsame Strukturvorgaben für die Akkreditierung von Bachelor- und Masterstudiengängen. Beschluss der Kultusministerkonferenz vom 10.10.2003 i.d.F. vom 04.02.2010 [Common Structural Guidelines for the Accreditation of Bachelor's and Master's Degree Programs. Resolution of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder of 10.10.2003 as amended on 04.02.2010]," Accessed 4 January 2021.

Teichler, U. (2015), "Higher Education Research," in *International Encyclopedia of the Social & Behavioral Sciences*, eds. J. D. Wright, pp. 862–869, Amsterdam; Waltham, MA: Elsevier Science; Oxford.

University Education Forum on Digitization. (2017), "The Digital Turn: Pathways for Higher Education in the Digital Age," Executive Report. 4 January 2021. Available at *https://hochschulforumdigitalisierung.de/sites/default/files/downloads/HFD_Final_Report_English.pdf*, updated on 2017.

van der Aalst, W. M. P., Becker, J., Bichler, M., Buhl, H. U., Dibbern, J., Frank, U., Hasenkamp, U., Heinzl, A., Hinz, O., Hui, K.-L., Jarke, M., Karagiannis, D., Kliewer, N., König, W., Mendling, J., Mertens, P., Rossi, M., Voss, S., Weinhardt, C., Winter, R., and Zdravkovic, J. (2018), "Views on the past, Present, and Future of Business and Information Systems Engineering," *Business & Information Systems Engineering*, 60, 443–477. DOI:10.1007/s12599-018-0561-1

Vavrus, L. (1990), "Put Portfolios to the Test," *Instructor*, 100, 48–53.

Westphal, E. (2018), Die Bologna-Reform: "Policy Making" in Europa und Österreich [The Bologna Reform: "Policy Making" in Europe and Austria], Vienna: New Academic Press. Available at: *https://www.newacademicpress.at/wp-content/uploads/asolmerce/download-1220-westphal_bologna-reform_leseprobe.pdf*

White, M. D., and Marsh, E. E. (2006), "Content Analysis: A Flexible Methodology," *Library Trends*, 55, 22–45. DOI:10.1353/lib.2006.0053

Winter, M. (2018), "Bologna–Die Ungeliebte Reform Und Ihre Folgen [Bologna-The Unloved Reform and Its Consequences]," in *Hochschulen im Spannungsfeld Der Bologna-Reform: Erfolge Und Ungewollte Nebenfolgen Aus Interdisziplinärer Perspektive [Higher Education Institutions in the Field of Tension of the Bologna Reform: Successes and Unintended Side Effects from an Interdisciplinary Perspective Course Search]*, eds. N. Hericks, pp. 279–293, Wiesbaden: Springer. DOI:10.1007/978-3-658-21290-2

World Economic Forum. (2019), "Data Science in the New Economy: A New Race for Talent in the Fourth Industrial Revolution," Insight Report. Available at *https://www3.weforum.org/docs/WEF_Data_Science_In_the_New_Economy.pdf*

Xun, L. S., Gottipati, S., and Shankararaman, V. (2015), "Text-Mining Approach for Verifying Alignment of Information Systems Curriculum with Industry Skills," in *Proceedings of The International Conference on Information Technology Based Higher Education and Training, ITHET*, 2015, pp. 1–6. DOI: 10.1109/ITHET.2015.7217959