This is a repository copy of *Data science curriculum in the iField*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/188812/

Version: Published Version

Article:
Zhang, Y., Wu, D., Hagen, L. et al. (10 more authors) (2023) Data science curriculum in the iField. Journal of the American Society for Information Science. American Society for Information Science, 74 (6). pp. 641-662. ISSN 0002-8231

https://doi.org/10.1002/asi.24701

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Check for updates

Received: 1 September 2021 | Revised: 23 April 2022 | Accepted: 5 July 2022

DOI: 10.1002/asi.24701

RESEARCH ARTICLE

JASIST WILEY

Data science curriculum in the iField

Yin Zhang[1] | Dan Wu[2] | Loni Hagen[3] | Il-Yeol Song[4] | Javed Mostafa[5] | Sam Oh[6] | Theresa Anderson[7] | Chirag Shah[8] | Bradley Wade Bishop[9] | Frank Hopfgartner[10] | Kai Eckert[11] | Lisa Federer[12] | Jeffrey S. Saltz[13]

[1]School of Information, Kent State University, Kent, Ohio, USA

[2]School of Information Management, Wuhan University, Wuhan, China

[3]School of Information, University of South Florida, Tampa, Florida, USA

[4]College of Computing & Informatics, Drexel University, Philadelphia, Pennsylvania, USA

[5]School of Information and Library Science, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[6]Department of Library and Information Science and Data Science, Sungkyunkwan University, Seoul, South Korea

[7]Connecting Stones Consulting, Sydney, Australia

[8]Information School, University of Washington, Seattle, Washington, USA

[9]School of Information Sciences, University of Tennessee, Knoxville, Tennessee, USA

[10]Information School, University of Sheffield, Sheffield, UK

[11]Stuttgart Media University, Stuttgart, Germany

[12]National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA

[13]School of Information Studies, Syracuse University, Syracuse, New York, USA

Correspondence
Yin Zhang, School of Information, Kent State University, Kent, OH, USA.
Email: yzhang4@kent.edu

Abstract

Many disciplines, including the broad Field of Information (iField), offer Data Science (DS) programs. There have been significant efforts exploring an individual discipline's identity and unique contributions to the broader DS education landscape. To advance DS education in the iField, the iSchool Data Science Curriculum Committee (iDSCC) was formed and charged with building and recommending a DS education framework for iSchools. This paper reports on the research process and findings of a series of studies to address important questions: What is the iField identity in the multidisciplinary DS education landscape? What is the status of DS education in iField schools? What knowledge and skills should be included in the core curriculum for iField DS education? What are the jobs available for DS graduates from the iField? What are the differences between graduate-level and undergraduate-level DS education? Answers to these questions will not only distinguish an iField approach to DS education but also define critical components of DS curriculum. The results will inform individual DS programs in the iField to

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2022 The Authors. *Journal of the Association for Information Science and Technology* published by Wiley Periodicals LLC on behalf of Association for Information Science and Technology. This article has been contributed to by U.S. Government employees and their work is in the public domain in the USA.

develop curriculum to support undergraduate and graduate DS education in their local context.

1 | INTRODUCTION

Data have far-reaching and increasing significance in our lives. Although Data Science (DS) has been around since the 1960s, it has gained increasing interest and attention from a wide range of domains and disciplines due to the data revolution in more recent years. Consequently, there has been ongoing debate and exploration from various perspectives regarding what constitutes DS (e.g., Brodie, 2019; Cao, 2018a; Hayashi, 1998; Ozsu, 2020). While DS has been recognized as multi- and interdisciplinary in nature (Cao, 2017; Raban & Gordon, 2020), individual disciplines and academic units have started to reflect, discuss, and establish a particular disciplinary identity in the DS research and education landscape (e.g., Donoho, 2017; Shah et al., 2021; Siebes, 2018; Vicario & Coleman, 2020; Virkus & Garoufallou, 2019, 2020).

As the demands of a data-informed society grow and evolve rapidly, there is an increasing need for workers skilled in dealing with data challenges in various roles and capacities (Berman et al., 2018; Carter & Sholler, 2016; Saltz & Grady, 2017). In response to this need, many disciplines have begun developing data-related programs (Wu, 2019). Despite the growing number of DS programs and courses, however, most such programs have been based on existing curricula from statistics, computing and informatics, business, and information technology programs (Cao, 2018b).

As a leading player in the DS field, information sciences (IS)/library and information science (LIS), broadly referred to as the Field of Information (iField), has started to explore, reflect, and position an iField perspective of DS (e.g., Shah et al., 2021; Virkus & Garoufallou, 2019, 2020) more comprehensively. To better develop and advance DS education in the iSchool context, the Board of Directors of the iSchools organization (https://ischools.org), a collective of information schools dedicated to advancing the information field, decided in March of 2019 to create the iSchool Data Science Curriculum Committee (iDSCC). The iDSCC's primary aim is to build and recommend a DS education framework for iSchools that reflects the core concepts, values, expertise, and strengths of the iField.

This paper reflects the efforts and outcomes of the iDSCC's work in exploring and identifying an iField approach to DS curriculum. The findings and discussion will contribute to defining and highlighting the critical aspects of the DS curriculum that distinguish an iField approach in the broader DS education landscape.

2 | RELATED WORK

To put the iDSCC's work into context and build upon previous efforts, a review of related work and background was conducted and is summarized in this section. The review helps the iDSCC understand the current state of DS education landscape and identify the major questions and challenges facing DS education in the iField.

2.1 | Evolution of iField

The origins and evolution of the information field—the iField—are linked to the growth of forms of information and the increased complexities of its access and use. References to the "information explosion" pepper historical accounts of the field (e.g., Buckland, 1999; Larsen, 2017; Saracevic, 1979, 1999). The promulgation of digital forms of information coincides with the launch of the iSchool Movement (Bruce, 2011; Larsen, 2017). However, the evolution of the iField has long been framed not only in relation to information technology, but also to professional principles and scientific insights shaping the application of those technologies (e.g., Saracevic, 1979). These remain strong threads through the discourse of the iField with a focus on relationships between information, people, and technology (https://ischools.org/About) and a deliberate and deep coupling of computational content with study of the contextual dynamics of creation and use of information (Dillon, 2012).

In many ways, the iField is a metafield with interdisciplinary culture embedded in its history (Bruce, 2011; Dillon, 2012; Larsen, 2017; Seadle & Greifeneder, 2007). The iSchools, for example, encompass the disciplines of computer science, library and information science, business informatics, knowledge management, business, sociology, psychology, philosophy, ethics, linguistics, and media, with a range of applied fields and disciplines such as astronomy, medicine, biology, health, history, religion, archaeology, musicology, literature, and art. While each individual iSchool will have its own strengths and specializations, the iSchool member schools (https://ischools.org/members/directory/) share a commitment to building on these interdisciplinary approaches to harness the power of information and technology for the benefit of individuals and communities (https://ischools.org/stories/ischool-movement/).

We see increasing reference to the iField's capacity to take transdisciplinary approaches (Galliers, 2003; Golub, 2019; Larsen, 2017). This transdisciplinary perspective is also visible in the breadth of research interests of iSchool faculty (Holmberg et al., 2013) and the themes discussed in iSchool conferences (e.g., Anderson & Parker, 2019). While diversity with "no neat boundaries" (Dillon, 2012) may be challenging, it is this very rhizomorphic characteristic that enables dynamic, open, and interconnected engagement within the iField (Burnett & Bonnici, 2013). Over the decades we see evidence of a deeply held appreciation of the integrative diversity within and across schools associated with the iField (e.g., Bruce, 2011; Buckland, 1999; Galliers, 2003; Saracevic, 1979, 1999). King (2006) celebrated this richness and the broad embrace of intellectual interests for the value in relation to enormous potential future options. Interdisciplinary collaborations and diversity of perspectives characteristic of the iField can thus offer a strategic advantage to an emerging field like DS (Marchionini, 2016).

Another persistent thread throughout the history of the iField relates to a concern for the needs of society and limitations of techno-centric approaches. From the early history of the teaching of information science as a distinct subject (Bawden, 2008; Farradane, 1976) through to contemporary articulations of sociotechnical framing (e.g., Bates et al., 2020), focus on people is a prominent and long-standing component of both inquiry and practice in the iField (e.g., Bruce, 2011; Larsen, 2017; Saracevic, 1999; Seadle & Greifeneder, 2007). Increasingly, interest in AI as a strategic and commercial force is accompanied by calls for better algorithmic governance and data/AI practices that more explicitly protect human rights (e.g., Bates et al., 2020; Tanweer, 2018). Societal concerns such as topics of human rights and social justice have a traditional place in iSchool curriculum, teaching human-centered approaches to tackle information and sociotechnical concerns (e.g., Rosenbaum & Fichman, 2019; Saltz et al., 2018). The iField's concern with the entire data lifecycle is thereby inextricably linked to the socio-cultural issues associated with data collection and use (Marchionini, 2016) and the provision of services to target communities, serving public good, democracy, and equality.

2.2 | DS education overview

There has been ongoing debate as to what constitutes DS. Based on a comprehensive review of the evolution of DS, Cao (2017) observes discussion has expanded from data-focused disciplines and domains (e.g., statistics, computing, and informatics) to non-traditional data-related fields such as the social sciences and management. There have been various perspectives regarding what defines DS (e.g., Brodie, 2019; Cao, 2017, 2018a; Hayashi, 1998; Ozsu, 2020). For example, Hayashi (1998) notes DS is an integration of statistics, data analysis, related methods, and results, involving data design, collection, and analysis. O'Neil and Schutt (2013) observe that DS integrates mathematical, statistical, computer technology, and domain knowledge, which reflects its remarkable interdisciplinary characteristics.

DS has been recognized as multi- and interdisciplinary in nature (Cao, 2017; Raban & Gordon, 2020). Many disciplines and academic units have started to reflect, discuss, and establish a particular disciplinary identity in the DS research and education landscape. For example, after reflecting on 50 years of DS in the context of mathematics and statistics, Donoho (2017) presents a vision of broad DS based on six data activities: (1) Data Gathering, Preparation, and Exploration, (2) Data Representation and Transformation, (3) Data Computation, (4) Data Visualization and Presentation, (5) Data Modeling, and (6) Science about DS. Donoho (2017) further points out that collectively, these activities are beyond what a single field can study or teach, and that Data Modeling is the major focus of statistics and mathematics. From a computer science perspective, Siebes (2018) considers DS to be a language based on computer science for datafied sciences. From a business and industry perspective, Vicario and Coleman (2020) emphasize DS's practical role of improving operations and serving as a driving force in creating knowledge, and Bailer and Fisher (2020) further add that DS, as it relates to business and industry, involves the entire process of problem elicitation and formulation from the beginning to the end of communicating and reporting the results.

To address the increasing needs of the DS workforce, various disciplines have also developed and offered data-related programs and courses. A recent review of graduate-level DS education programs shows that Mathematics and Statistics, Computer Science, Business, and Library and Information Science (LIS) are the leading disciplines offering such programs (Wu, 2019). Some institutions also offer an interdisciplinary DS and education program and the challenges of teaching DS with an interdisciplinary teaching team are also discussed (Bates et al., 2020; Wu, 2019). While many disciplines introduce DS curricula and the number of DS programs and courses is growing, it has been observed that they are mostly based on the re-labeling and combination of existing curricula from statistics, computing and informatics, business, and information technology curricula (Cao, 2018b). For example, the field of business and economics tends to combine DS education with business applications and consideration of business processes (Miah et al., 2020), and the field of statistics tends to address why and how to set up DS programs under the discipline of statistics (Aerts et al., 2021).

2.3 | DS education in iField

In recent years, there have been active discussions at conferences to explore, reflect, and position an iField perspective of DS (e.g., Albright & Mehra, 2020; Blake & Brown, 2019; Gunderman, 2019, 2020; Sundqvist et al., 2020; Taylor et al., 2019). Based on a bibliometric analysis of DS research output, Virkus and Garoufallou (2019, 2020) examine disciplinary contributions and identify the main themes discussed in the publications from the LIS perspective. The study shows the DS field is highly interdisciplinary, and that there has been a continuous increase in research output from LIS since 2015. The major themes discussed in LIS publications span six areas: (1) DS education and training; (2) knowledge and skills of the data professional; (3) the role of libraries and librarians in the DS movement; (4) tools, techniques, and applications; (5) DS from the knowledge management perspective; and (6) DS from the perspective of health sciences. In a position paper, Shah et al. (2021) argues that the iSchools approach to DS research and education encompasses three distinct yet integrated core characteristics: human-centered, socially responsible, and contextual grounding.

So far, most LIS schools have started offering DS curricula to various extents (courses, specializations, and programs) at the undergraduate and/or graduate level and engaged in exploring the uniqueness of the discipline in the DS education landscape. After reviewing and comparing both DS and information science programs, L. Wang (2018) notes that DS and information science programs are closely related with similar missions and concerns that complement each other. Hagen (2020) proposes that key aspects such as information behavior and ethics and knowledge of programming and statistics, should be included in DS education (Hagen, 2020). Song and Zhu (2017) point out that DS education in the iField should be user-, tool-, and application-based, and recommended people, technology, and data as the baseline for this. To provide more effective DS education, Song and Zhu (2017) further propose that the DS lifecycle should be considered in course design, and specific courses should be provided, including communication skills, system thinking, project management, big data technologies and model building techniques (Song & Zhu, 2016). Because of the practical nature of DS, the cultivation of DS students should be encouraged by learning beyond the classroom with practical projects to solve real-world problems (Elkhatib, 2017).

2.4 | Gaps and questions

After reviewing related literature, a survey of DS programs and curricula in various disciplines and the iField, and several iDSCC sessions engaging the community at conferences (iDSCC, 2019a, 2019b, 2020, 2021), it is evident iField educators have put significant efforts into DS education, but major questions remain to be answered:

1. What is the iField identity in the multidisciplinary DS education landscape? More specifically, what are the characteristics of iField DS education?
2. What is the status of DS education in iField schools?
3. What knowledge and skills should be included in the core curriculum to serve well for iField DS education?
4. What are the jobs available for DS graduates from the iField?
5. What are the differences between graduate-level and undergraduate-level DS education?

Answers to these questions will contribute to the ongoing discussion and efforts in identifying an iField approach to DS education and defining the critical components of the iField DS curriculum in the broader DS education landscape. In addition, results will guide and inform individual DS programs in the iField to develop curriculum to support undergraduate and graduate DS education in their local context.

3 | METHODOLOGY

The iDSCC has adopted a research approach to address the fundamental questions facing the DS education in the iField and has conducted a series of studies using multiple methods for data collection and analysis.

3.1 | Delphi study

As an initial step, the iDSCC engaged the 16 DS educators, researchers, and practitioners serving on the committee through a group exercise in a multi-round Delphi survey for the following two purposes:

- To identify critical topics and best practices in DS education in the iSchool/iField context, and
- To build group consensus on core curriculum recommendations and best practices in DS education.

iDSCC was formulated by 2018–2020 iSchools Chair Sam Oh and DS scholar and educator Javed Mostafa and Il-Yeol Song. The leadership team recruited committee members who were active researchers and professionals in DS who presented at iField conferences. Considerations were also given to be inclusive in terms of geographic locations of iSchools. In addition, an open call for membership was issued at a well-attended panel at the ASIST 2019 Annual Meeting.

The Delphi method is a useful and effective research methodology to examine a new phenomenon or evolving topic when there is no clear or knowable answer and the collective opinions of domain experts are needed and valued (Linstone & Turoff, 1975; Kochtanek & Hein, 1999; Baruchson-Arbib & Bronstein, 2002; Zhang & Salaba, 2009).

This Delphi study focused on the following major areas/questions regarding DS education in the iSchool context:

1. What are the unique characteristics of DS education in iSchools that differentiate it from other disciplines and best represent iSchools' core concepts, values, expertise, and strengths that can be used for branding?
2. What are the jobs or careers for students who successfully complete a DS program in an iSchool (or an academic unit which represents the iField)?
3. What are the theoretical and/or generalizable knowledge areas that students must learn?
4. What are the skills and competencies that students must achieve?
5. What are the best practices that are essential for students to translate knowledge and skills to real-world problems/projects?
6. What are the knowledge areas, skills, and practices that should NOT be given major emphasis or should be avoided?

The three-round Delphi study occurred over 9 weeks in early 2020. Specifically,

- In the Round 1 survey (February 2–11, 2020), the panelists were asked to independently and anonymously provide up to five suggestions to each of the six questions regarding DS education. A total of 280 individual suggestions with respective rationales were received. These original, individual suggestions were then synthesized into 63 items, with similar suggestions being merged and presented in the subsequent survey rounds for review and rating by the panel.
- In the Round 2 survey (February 23–March 5, 2020), panelists had the opportunity to make new suggestions that were not covered in Round 1. As a result, two new suggestions were raised and added for a total of 65 suggestions for the Round 3 survey.
- The Round 3 survey (March 19–April 4, 2020) was used to build group consensus regarding suggestions for the key areas of DS education. The panelists rated the importance of the suggested items after reviewing their peers' anonymous suggestions and summarized rating scores and rationales.

The results of the general Delphi study were further examined and deliberated by the two iDSCC subcommittees with their respective focus on graduate level and undergraduate level DS education.

3.2 | DS program and curriculum analysis

3.2.1 | Graduate level DS program and curriculum

To understand the status of graduate level DS education in information schools worldwide, the iDSCC graduate subcommittee conducted a review of all iSchools' programs in the organization directory and other information schools. Other information schools were drawn from the website "Datascience.Community" (http://datascience.community), which provides an index of DS programs worldwide. The criterion for indexing programs on this site is described as "Most of the programs are not named data science, but they all focus on producing data people."

As of June 2020, there were 116 members in the iSchools organization directory and 618 institutions or programs in the "Datascience.Community" website. Given the lack of agreed upon definition for "data science," our inclusion criterion were programs with a particular emphasis on developing skills related to data mining, data analysis, and data modeling. Based on this criterion, the subcommittee manually reviewed all the iSchools websites and identified 72 DS graduate programs. For the "Datascience.Community" source, the subcommittee used the keywords "informatics" and "information" to search the "department" field and removed any schools already included in the iSchools directory from the sample. The final sample included 96 DS graduate programs, with 72 programs from iSchools and 24 programs from the "Datascience.Community" website. The schools with the sample DS programs will uniformly be called "iField schools" hereafter.

A review of the DS programs revealed that a majority of the 96 graduate programs are application oriented. This is not surprising as "Data Science" aims to use data to *understand and analyze* real-world phenomena, and this discipline is applied to solving real-world problems. Therefore, DS education programs pay more attention to improving students' critical thinking skills. Based on the manual review, detailed information regarding the programs was collected, including program names, program descriptions, courses, course descriptions, and program outcomes. Content analysis was conducted to analyze the program descriptions and program outcomes to identify competencies. Content analysis is a research approach for systematically describing qualitative data to transform the text into well-ordered conceptual categories and to gain insights (Wildemuth, 2016). It is an interactive approach because the research team frequently meets during the stage of interpretive convergence (Saldana, 2015) to discuss all issues and ensure consistent evaluations. Further, the content analysis approach was also used to categorize and analyze all 2,084 courses to refine the course classification. Finally, course description texts of 883 courses were clustered to further validate the reliability of the classification results by using quantitative mining technologies.

3.2.2 | Undergraduate level DS program and curriculum

The iDSCC undergraduate subcommittee conducted a survey of iSchools DS programs' description and learning outcomes. In addition, the subcommittee conducted a detailed course-level analysis of undergraduate data programs. Two sets of independent data were collected:

First, to study the existing DS programs, all 122 iSchool Deans and Directors were contacted twice through the mailing list (all-ischool-leaders; all-ischool-leaders@ischools.org) to gather program descriptions and program learning outcomes on any undergraduate DS programs during June and July of 2020. The recruitment email explained the information gathered would be analyzed to produce common elements for a model iSchool DS curriculum that could potentially be used by other iSchools starting such programs. The iSchools may be grouped by region (North America (54), Asian Pacific (35), Europe (33)). As a result, nine DS programs from three regions responded and provided program information (5 North America, 3 Asia, 1 Europe). The total number of undergraduate DS programs in the iSchools is not known. The programs reviewed were a convenience sample of those self-reporting to have DS programs, but efforts were made to gather programs from each region. While the nine schools may not be representative of all iSchools, they provided a corpus to draft a snapshot of current DS program descriptions and learning outcomes. To analyze the information provided, all nine iSchools programs' descriptions and learning outcomes were reviewed and sorted by thematic coding.

Second, a course-level analysis was conducted about the DS curriculum at LIS schools/iSchools. The school sample included those in the United States appearing in any of the three directories as of August 18, 2020: (1) American Library Association (ALA) accredited programs database search, (2) Association for Library and Information Science Education (ALISE) institutional directory, or (3) iSchool Directory. Although we identified the boundaries of LIS schools and iSchools from the directories, identifying the DS programs from these schools was extremely challenging because relevant courses have diverse names and are offered through various departments. For the feasibility of data collection, we decided to identify any programs from this list that include "data" in the program title and collected course offerings and course descriptions in those programs. As a result, a total of 120 courses offered in 12 data programs at the undergraduate level were included in the course sample. For the analysis, each of the courses was coded using an iterative coding approach by creating new thematic categories as they arose while reading each of the course descriptions. A total of 13 thematic categories arose. The manual content analysis was validated using quantitative mining techniques using the same course description data.

3.3 | Job analysis

To understand the demand and requirements for DS-related jobs in the marketplace, advertisements for DS-related jobs from the United States, China, the United Kingdom, and Germany were gathered and analyzed. These countries have widely been recognized as leading markets that hold the greatest opportunities for data scientists. Specifically, the collective sample included the following datasets:

- The ALA JobLIST archive contained 24,058 job advertisements during August 2006 to April 2018. Entries with "data" as part of the position title were retrieved and manually reviewed, resulting in 391 entries that served as a DS job sample in the United States specifically targeted at LIS programs.
- Indeed, the largest global employment website, targets a broad range of jobs. Job advertisements from the Indeed website with "data science" as part of the position title were collected during July to August 2019, resulting in 1,312 entries that served as a general DS job sample in the United States.
- For a European sample, following the data collection approach using Indeed for the U.S. sample, 519 job advertisements for DS jobs were retrieved from Indeed's national website in Germany and 333 equivalents in the United Kingdom during July to August 2020. Job advertisements written in German were translated into English for easier comparison.

- LaGou is a well-known job search website for employment positions in China. From March to July 2019, 2,239 job advertisements containing the keyword "data" in the job title were obtained from the LaGou website. Using existing studies to extract iSchools' feature words (Wu et al., 2012), the frequency statistics of the feature words of job descriptions were used as a filtering criterion to find iSchool related jobs. As a result, 480 entries were selected and translated to English to serve as a DS job sample in China targeted iField graduates.

The advertisement data were analyzed using quantitative mining techniques to identify representative DS job titles, job market needs, and requirements and compare them in different countries. The analysis results are reported thereof in Table 4 and Figure 3 of the results section.

4 | RESULTS AND ANALYSIS

The findings of the series of studies described in the methodology are presented and analyzed below to answer the critical questions facing DS education in the iField.

4.1 | Delphi study of DS education in the iField

Focusing on the characteristics and critical topics of DS education, the graduate and undergraduate subcommittees reviewed, deliberated, and selected most relevant statements for their respective levels of education, based on the results of the Delphi study. Common characteristics of iSchool DS education for both graduate and undergraduate levels include:

- Takes a human-centered approach.
- Considers the moral, ethical, and societal aspects of data and the impact on society at large.
- Provides a holistic view of the data lifecycle and practical learning and teamwork.
- Addresses real-world problems.
- Does not require students to learn advanced programming or algorithm knowledge but pays more attention to data processing, analysis, and application.
- Teaches core knowledge and skills in data information literacy and competencies with a different level of requirements and expectations between levels noted.

The highly recommended knowledge and skills suggested from the Delphi study for DS education in the iField include the following with the supporting rationales:

- *Basic programming and data analytics/visualization/interpretation/communication* were the most highly recommended items. While basic programming and analytics are considered technical skills, visualization/interpretation/communication are considered as storytelling skills to add value to data by adding a layer of creative and informed judgment.
- *Basic math and statistics/machine learning fundamentals* were another highly recommended set of technology-related items. While students are not expected to have a strong background in mathematics and statistics, they should have basic math and statistical knowledge to work on advanced data projects, interpret the results, and produce insights. Basic knowledge of common machine learning and deep learning methods such as clustering, classification, and linear regression were also highly recommended to strategically work with data.
- *Ethics and implications of data in society* were considered a key knowledge area for students to select and analyze data and create DS products in ways that minimize bias and apply/use them ethically. Data ethics cover such issues as bias in data, fairness, and transparency. Students should be able to critically reflect on the societal impact of DS projects. It was recommended to incorporate ethics throughout the curriculum instead of covering separately.
- There was notable discussion regarding the importance of *data processing and management*, which not only includes cleaning but also digitization, preservation, and archiving data. Data, discussed in this context, is not limited to digital-born data. In addition, it also includes data management and curation of both big and small data.
- The ability to *select, implement, and evaluate proper tools*, among numerous DS approaches and techniques, was also highlighted.
- The value of *data science project management*, the ability to effectively manage the complexity of DS projects, was also highlighted.
- *Data and information literacy* was highly recommended. Data literacy is important in general and in an applied disciplinary context such as for digital humanities specifically. Data literacy is important for interpretation of DS projects because it informs what is behind data, such as biases involved with generation of data.
- *Database knowledge and skills* received divided recommendations. While the group agreed that database knowledge and skills are fundamental and important

for data modeling and data processing/management, a few did not consider them a core requirement of modern data scientists.

Differences between graduate and undergraduate education noted from the Delphi study are summarized below:

- In knowledge and skills, undergraduates focus on fundamental data and information literacy and competencies, whereas graduate students pay more attention to practical applications and need higher skills in problem-solving, project management, big data analytics, and machine learning.
- In application, graduate students are expected to understand data and business problems and address real-world problems and practical applications more than undergraduates.
- In the data lifecycle, graduate students are expected to have a deeper grasp of the whole data lifecycle than undergraduates.
- In jobs, design-focused jobs are equally important for undergraduates and graduate students, such as data designer, information system designer, user requirement engineer with graduate students expected for more leadership/managerial/advisory jobs, domain-specific jobs, and traditional/service DS jobs.

4.2 | DS program and curriculum analysis

4.2.1 | Graduate level DS program and curriculum

The program descriptions and program outcomes of the sample programs were content analyzed and coded for competencies and skills. The competencies were coded in three categories: (1) professional, (2) personal, and (3) practical, while skills were coded into four categories: (1) programming languages, (2) big data management platforms, (3) operating systems, and (4) domain theory. Table 1 provides the coding results with the top five most frequent secondary codes under each primary code displayed.

The analysis of the program descriptions revealed professional competencies related to the data lifecycle, such as data processing, analysis, and management, which are central to DS competency formation at iField schools (see Table 1). Visualization is also valued as an important method for illustrating and understanding text and other data types. As a practice-oriented education, the ability to extract value from data to assist in real-world problem solving is critical to personal competencies. This requires students to master data processing and analysis skills and uncover the value behind the data, which is the core of DS. All data processing and analysis skills serve to solve real-world problems, and data analysis that is divorced from real-world problems will lose its meaning. DS graduate programs emphasize the accumulation of students' domain knowledge and methods and the shaping of personal characteristics to emphasize the value of people in DS.

The content analysis and open coding of the sample course titles and descriptions yielded thematic categories, followed by axial coding that generated new thematic categories based on thematic category relationships. There were 12 categories that emerged as summarized in Table 2.

The results show iField schools provide application-oriented DS graduate programs with the highest percentage of courses falling into the category of practice and project management (34.98%). Data mining, analytics, and modeling came in second. The third and fourth most popular courses (10.51% and 8.83%, respectively) are library and information science and humans and society. This reflects that these courses are based on LIS theory and focus on applications in solving real-world problems in developing skills to unlock data value. At the same time, the focus on techniques, methods, and tools is accompanied by a curriculum design that reinforces the human and social dimensions, with attention to privacy, intellectual property, ethics, and other data and DS concepts. The application-oriented, human values focus is further reflected and reinforced in the DS graduate curriculum design.

Quantitative mining technologies were employed to cluster the courses to explore the data further and triangulate the coding classification results. Because of the short course names, the course name clustering results are dispersed, making it impossible to depict the co-occurrence relationship. As a result, we only performed a cluster analysis on the course description texts of 883 (42.37%) courses. As shown in Figure 1, the main nodes in the clustering diagram are (1) management, (2) model, (3) information system, (4) programming, (5) statistics, and (6) database. This clustering analysis result complements the manual coding result with more specific details.

4.2.2 | Undergraduate level DS program and curriculum

When reviewing undergraduate DS program descriptions and learning outcomes, it was found that each institution varied considerably in the amount of detail provided based on formatting norms across universities. In some

TABLE 1 Competencies and skills in Data Science graduate program

	Primary code	Secondary code (top 5)	N
Competency	Professional	Data analysis	290
		Data management	258
		Data processing	176
		Statistics	172
		Visualization	150
	Personal	Solve real-world problems	161
		Practice literacy	152
		Support decision making	128
		Use tools	106
		Academic literacy	98
	Practical	Understand needs	14
		Organizational skills	6
		Responsibility	6
		Entrepreneurship	4
		Continue learning	4
Skill	Programming languages	Python	52
		R	24
		SQL	16
		NoSQL	16
		SAS	10
	Big data management platforms	MapReduce	16
		Spark	12
		Hadoop	16
	Operating system	Linux	2
		Unix	4
	Domain theory	Library and information science theory	102
		Computer science theory	76
		Biological and medical scientific theory	74
		Business science theory	54
		Mathematical theory	16

instances, program learning outcomes were exhaustive and others very brief. The local approaches to writing program descriptions and program learning outcomes did not allow for easy analysis; however, some quantification of related topics was done by assigning codes to related topics mentioned in the outcomes. The following topics with the number of DS programs addressing the topic in parentheses provide an overview of iSchool coverage in the DS realm: visualization (8), ethics (7), statistics (5), machine learning (5), communication (4), human(e/istic) (4), programming (3), and leadership (2).

As a result, the following program learning outcomes were synthesized using the language and concepts found across the reviewed sample programs:

- Students study the tools and theories of information science, computer science, and statistics to master the following knowledge and abilities:
- Select from, use, and interpret results of descriptive statistical, machine learning, text-mining, and natural language processing methods effectively to discover relations and patterns contained in data and make reasonable predictions, and use data visualization to effectively display and interpret data to solve practical problems in the field of management.
- Master the relevant technologies and methods of big data acquisition, storage, processing and analysis, transmission, and application; use and modify standard techniques.

TABLE 2 Categories of Data Science graduate courses in iField schools

Categories of courses	Percentage of total courses
Practice and project management	34.98%
Data mining, analysis, and modeling	14.88%
Library and information science	10.51%
Human and society	8.83%
Data processing, storage, and management	8.73%
Mathematics and statistics	6.14%
Computer science and artificial intelligence	6.09%
Algorithms and programming	4.32%
Research design and methodology	2.21%
Introduction to data science	1.20%
Big data and big data applications	1.15%

- Apply appropriate methods and techniques to model, store and query structured and unstructured data sources.
- Communicate the results of analyses accurately and effectively, in writing, orally, and visually to different audiences.
- Demonstrate the ability to plan, manage, and document projects.
- Describe human-centered design of data and digital objects.
- Understand social and ethical contexts of information and technology, including the development of the data industry and relevant policies, laws, regulations, and standards in the field of DS.

Following a similar approach, the following DS program description was synthesized with feedback from the entire iDSCC membership:

Data Science is an interdisciplinary field concerned with the integration of methods, processes, systems, and tools from Computer Science, Informatics, and Statistics, to discover, validate, and apply knowledge and actionable insights from data, across a broad range of application domains. A data science program focuses on improving students' ability to solve practical problems; integrates domain knowledge with big data to understand, analyze, apply, and manage data, and conduct data-centered research and engage in data analysis and visualization. Students will gain the critical analytical skills needed to assess the feasibility, benefits, limitations, risks, and ethical implications of applying data sciences methods in different settings. Job opportunities for these graduates include Data Scientist, Data Analyst, Business Analyst, Advanced Analytics Professional, and Database Administrator.

The undergraduate subcommittee also analyzed the course descriptions by first creating a network graph using cooccurrences of terms as an edge, and unique terms as a node to identify the major topic clusters in the undergraduate DS course offerings as shown in Figure 2 and summarized below:

- The most central of all the clusters is on introductory elements of information and working with data. Introductory and big picture elements of the courses are a focus of this cluster with words like "foundation, introduces, and overview."
- Another cluster of courses focuses on managing data and working with data in a variety of ways. The terms are action-based, such as "reuse, implementation, and management," which emphasizes the cluster's focus on actively working with data in a firsthand manner. This cluster includes many of the primary elements of working with data such as organizing, managing, and representation.
- The third cluster is heavily focused on math and statistics. Words such as "probability and stat" illustrate the focus of statistics in this cluster. Words such as "calculus, function, and quantitative analysis" highlight the focus of mathematics in DS courses.
- The fourth cluster focuses on data science fundamentals and the ethics of data science. Phrases like "data mining, data science, and analytic" illustrate this cluster's focus on the fundamentals of data science and analyzing data in these courses. Words such as "privacy, context, and society" connect to concerns over the ethics of working with data.
- The fifth cluster focuses on programming. Words like "computation, language, and python" illustrate the focus of programming skills and knowledge in these DS courses.

Additionally, the subcommittee used an iterative coding approach to further analyze the course descriptions. Thirteen thematic categories emerged in the course descriptions, as summarized in Table 3, with the following common and unique categories:

- *Data Mining and Analytics* was the most common category (92% of schools), followed by *Statistics/math courses* (58%).
- *Data, Human, Computer, and Society*. iSchools provide DS courses to develop social perspectives for students that are rooted in social science theories to make decisions involving information, people, and

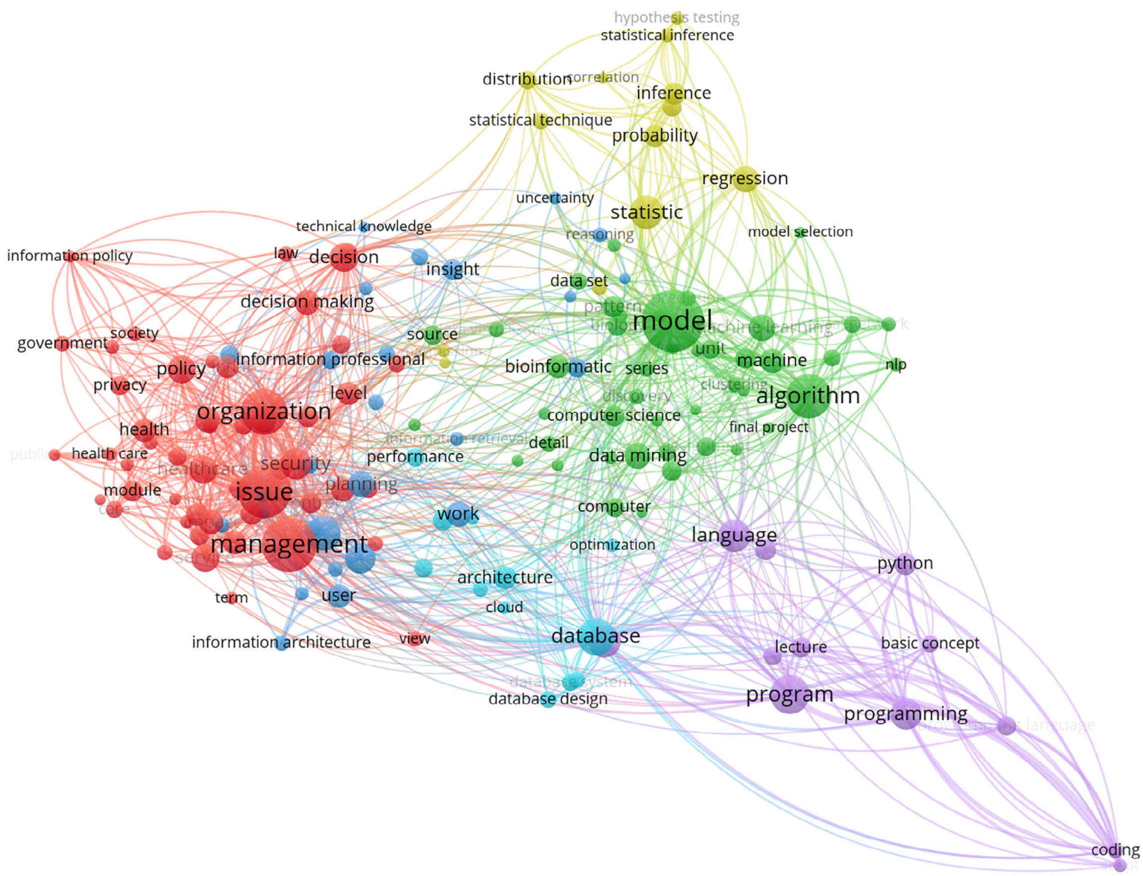

FIGURE 1 Data science courses clustering

technology. These courses provide principles of data and data science based on legal principles and social science theories such as privacy, intellectual property, ethics, socio-technical systems theories, and social informatics.

- *Data Management*. These courses teach data curation and management, following principles such as the FAIR data concept that considers data sharing and reuse throughout the data curation lifecycle.
- *Data Literacy*. These courses teach how data are created and consumed, and the effects of data practices in society. In addition, these courses cover data policies that provide guidelines for data practices.
- *Data Visualization*. These courses introduce theories of visual design as well as computer scripting to enable students to visualize actionable information extracted from large sets of data.
- *Project Management*. Some courses teach fundamentals of project management such as project selection methods and work breakdown structures. Other courses create opportunities for students to apply their learned skills to solve specific domain problems through team projects.
- The "*Others*" category includes HCI (human–computer interaction) and GIS (geographic information system) courses.

4.3 | DS jobs analysis

4.3.1 | Job titles

As explained above, our data analysis only includes job advertisements that targeted graduates of degrees traditionally taught by iSchools. Table 4 shows a summary of the aggregated job titles in four countries from various data sources.

According to the collected recruitment information, job opportunities in DS were broad and mainly included: (1) data scientist, responsible for cleaning, managing, and organizing big data; (2) data analyst, responsible for collecting, processing, and performing statistical data analysis; (3) data architect/data engineer, responsible for creating, managing, and maintaining the data management system/data warehouse/data sources; and (4) data

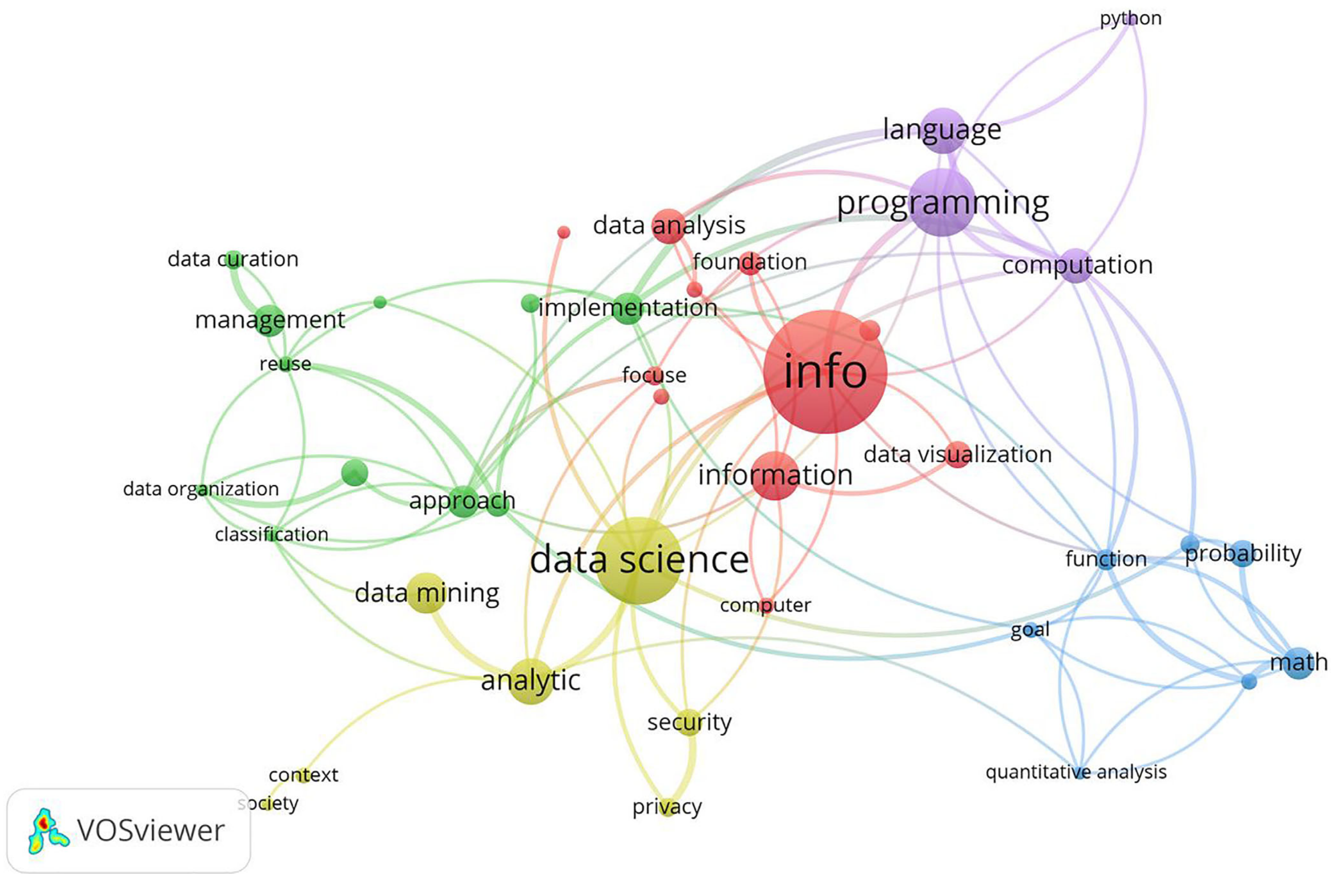

FIGURE 2 Major topic clusters in current undergraduate Data Science course offerings

TABLE 3 Categories of course offerings from the 12 universities

Categories of courses	Number of courses	Number of universities providing relevant courses	Percent ($n = 12$)
Data mining and analytics	28	11	92%
Statistics/math	10	7	58%
Programming	20	6	50%
Data management	11	6	50%
Data, human, computer, society	10	6	50%
Data visualization	7	6	50%
Introduction to data science	5	5	42%
Database	7	4	33%
Project management	4	4	33%
Others (HCI and GIS)	9	3	25%
Information systems and technology	4	2	17%
Machine learning	2	2	17%
Data literacy	3	1	8%
Total	120		

Note: The first six cells indicate the major clusters observed in the quantitative text-mining results in Figure 2. Categories in bold reflect the most highly rated items from the Delphi study.
Abbreviations: GIS, geographic information system; HCI, human–computer interaction.

TABLE 4 Comparison of job titles

The United States				The United Kingdom		Germany		China	
Position title (ALA JobLIST)	***N***	**Position title (indeed)**	***N***	**Position title**	***N***	**Position title**	***N***	**Position title**	***N***
Data Services Librarian	21	Data Science Manager	66	Data Scientist	118	Data Scientist	119	Data & Big Data Engineer	123
Data Librarian	13	Data Science Director	58	Senior Data Scientist	82	Data Scientist/Data Engineer	14	Data & Big Data Development	119
Social Sciences Data Librarian	10	Data Science Engineer (Senior)	36	Lead Data Scientist	48	Data Scientist Advanced Analytics	12	Big Data Architect	89
Bibliographic Database Designer	8	Data Science Engineer	26	Junior Data Scientist	18	Professional Data Scientist	9	Algorithm Researcher & Engineer	19
Data Annotator	6	Data Science Analyst	23	Head of Data Science	7	Big Data Analyst	9	Data & Big Data Expert	17
Data Management Specialist	6	Data science Manager (Senior)	22	Senior Data Scientist Returners Programmer	4	Data Scientist for Marketing Technology	7	Data & Big Data R & D Engineer	15
Research Data Librarian	6	Data science Intern	21	Data Scientist/Data Engineer	4	Freelance Data Scientist	7	Java Development & Engineer & Development Engineer & Architect	12
Social Sciences and Data Librarian	5	Data science Consultant	16	Medical Data Scientist	3	Senior Data Scientist	4	Big Data Director	10
Data Curation Librarian	5	Head of data Science	14	Safety Data Analyst	3	Lead Data Scientist	1	Big Data Software Engineer	7

science manager/consultant, responsible for understanding user or business needs and problem-solving, playing the decision-maker or advisor role.

However, DS job titles vary from country to country, and there is no uniform standard. In the United States, the job titles released by the ALA JobLIST are closely related to the LIS field, like "data librarian" or "database designer/manager." Most job titles released by Indeed directly contain "data science" and are not limited to specific fields, like "data scientist" or "data analyst." When comparing both groups above, DS jobs in the LIS field appear more oriented toward data/database management, while there is little demand for technical positions such as data analysis and data architecture. In addition, big data-related job titles are a major feature of Chinese DS job advertisements. DS job titles in China prefer to use "Big Data" rather than "Data Science," which is more related to big data engineering and big data development.

4.3.2 | Job requirements

Job requirements or responsibilities in the recruitment advertisements were extracted and analyzed with stop words removed. VOSviewer software was used to cluster the job requirement keywords. Figure 3 shows the keyword clusters of recruitment requirements in the United States, the United Kingdom, Germany, and China.

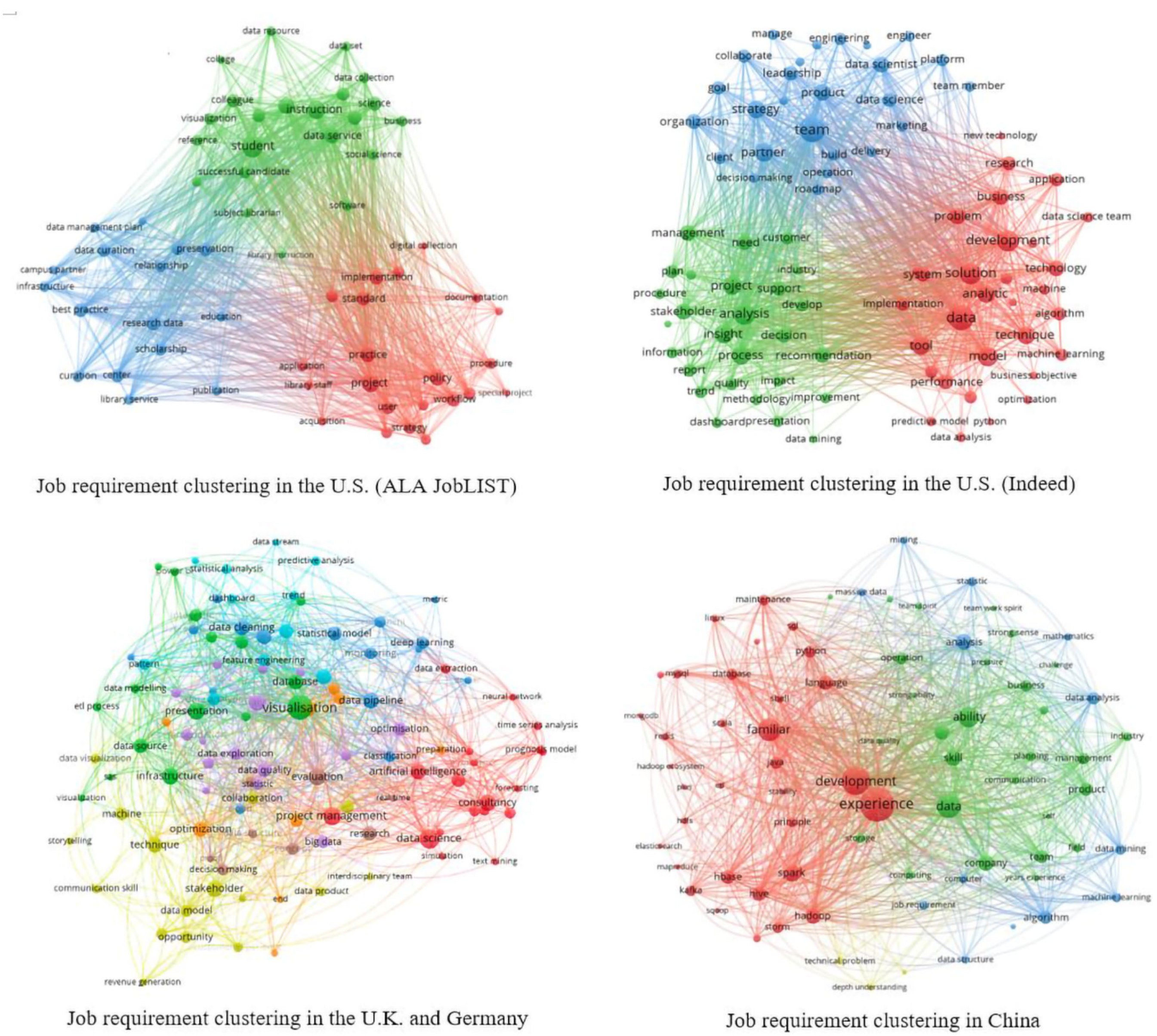

FIGURE 3 Comparison of job requirements clustering

For DS recruitment requirements, the skills required were divided into two fundamental areas:

1. Professional skills. Different DS positions put forward different requirements for professional skills, but in general, professional skills include the data science literacy required to tackle problems in the areas of data management, data analysis, data visualization, model building, programming algorithms, and text mining.
2. Soft skills. Soft skills are becoming increasingly important in the workplace, especially for DS graduates. The terms "team," "leadership," and "collaborate" appeared frequently.

There are also differences in occupational requirements across countries. The United States has more macro-level job requirements, with data decision-making and data services being important nodes for job advertisement clustering. Like the job title analysis, the requirements for "data preservation and management," "library services," and "publication" are unique in the job advertisements for ALA JobLIST, which are integrated with the LIS field. The United Kingdom and Germany have data visualization as the main competency requirement. Chinese recruitment advertisements are more detailed, involving multiple related techniques from data storage to data analysis. Work experience is more important for Chinese job seekers. In addition, Chinese companies also pay attention to programming ability (Python) and proficiency in data storage/computing platforms (e.g., Hadoop, Scala, and Storm).

5 | DISCUSSION

The iDSCC's 2 years of research inquiry and community engagements to address important questions facing DS education in the iField yielded rich findings. These findings are discussed and reflected around those questions and in the context of related work in the iField and broader DS education landscape.

5.1 | iField approach to DS education

How the iField DS programs and curricula differ from those in adjacent fields such as statistics, math, computer science, engineering, and business was frequently discussed, partially because the elements included in our recommended curriculum seem quite like those recommended by other disciplines (e.g., programming, machine learning, data mining, data management, and communication skills). For example, in an overview of the Masters' degree programs in Business Analytics, Analytics, and DS and related job market (Bowers et al., 2018), it was found that besides technical and data analytical skills, soft skills are considered important in job postings and deserve more attention in the degree programs. In another study, Wu et al. (2021) compared DS graduate programs in iField (library and information science), Business, CS (computer science), and Statistics offered in the top 50 universities per 2020 QS World University Rankings. It was found that DS programs in different disciplines showed the following similarities:

- the focus on the development of practical skills;
- requirement of a certain level of computer or statistical foundation from applicants; and
- a similar course structure with varied proportion of courses in various categories.

Wu et al. (2021) also noted differences in DS programs across fields:

- CS and Statistics programs focus more on students' data processing and analysis skills, while business programs focus on training big data talents in the field of finance or economics to solve specific business problems.
- iField programs do not merely rely on technologies and algorithms, but also place more emphasis on developing students' information literacy, with the aim of being able to understand users' demands and improve the interpretation of data and models. Building on long established social science traditions within their schools, iField programs also offer courses such as ethics, information privacy, information policy, regulation, and law for students, which reflect a "human-centered" data science education.

Based on previous study findings, Committee discussions, and comparing to DS curricula in other disciplines, DS is naturally an "applied" and "practice-oriented" field of study where an effective integration to the domain and given context is important. The Committee concluded that the uniqueness of iField DS lies in the spirit of DS programs and in its totality, not simply in the superficial elements listed in the recommended curriculum, which may seem like the recommendations made by adjacent disciplines. This is because DS is, by foundation, a practice- and application-oriented field of study, thus, each discipline is making its DS education relevant to its own disciplinary educational goals. Focusing our discussion on what we are, and what/how it should be taught might be more fruitful than forcing us to find differences. In the following, we elaborate the uniqueness and spirit of iField DS based on our findings (i.e., transdisciplinary, human-centered, forward-looking qualities, and its attention to how DS is taught).

Schools in the iField supply an ideal context for DS education, being transdisciplinary by nature and part of a meta-discipline studying the implications of data and information in relation to technology and people. The longevity of the iField's human focus in relation to data and information and its interdisciplinarity strength make DS a natural "fit" within iField curricula. In this legacy we can see the origins of the distinct components of the iField approach to DS: human-centered, socially responsible, and rooted in context (Shah et al., 2021). In addition, the context and tradition of the iSchool movement contributes to a more holistic DS education, incorporating data, technology, and human considerations. The movement's core values stress the impact of our work on society at large so that iField DS is assured of never focusing merely on improving technical performance.

Furthermore, the interactive and integrative aspects of the iField cultivate forward-looking qualities needed to meet the challenges of the connectivity era, not only in terms of increasing amounts of data but bringing structure to unstructured data (Song & Zhu, 2017). Considering the ambiguity often flagged as a challenge in DS work, one could argue such creative curiosity is an essential part of DS education.

The Delphi study revealed that respondents considered *how* DS is taught to be as important to defining an iSchool approach to DS education as *what* is taught in the curriculum. In fact, some asserted that it could be more important as a distinguishing characteristic. What is also clear from the investigations undertaken by members of the committee is that the iField offers two broad, intersecting pathways for the study of DS: one path foregrounding DS and the other embedding it in the more traditional areas of inquiry already part of iSchool curricula. Using diverse lenses to examine data challenges, data scientists learn to acknowledge the impact of context on technologies and data. By "keeping the human in the data scientist" (Anderson & Parker, 2019), DS education in the iField places attention on people and organizations as they affect and are changed by data. Instead of merely treating ethical and legal issues as "considerations" or "compliance issues," iField schools prepare students who can add value to DS practices by asking what they *ought* to do, not simply what they *can* do, and further considering who benefits, for what purpose, and at what cost.

5.2 | DS jobs

The DS job analysis showed the market demand for the competencies and skills of DS students. The ALA JobLIST showed a clear focus on data management or data services in libraries at academic institutions. In contrast, data gathered from other job websites showed a broader targeted audience with clear expectations for technical requirements and diverse employment opportunities. Overall, besides technical skills, employers also emphasized soft skills such as teamwork, leadership, and communication, which are essential for working in DS projects. Another major takeaway from the job analysis is the job title variances and varied roles. The detailed content analysis of responsibilities specified in the job descriptions helps address the previously observed problem of DS role ambiguity (Saltz & Grady, 2017).

The job data sample for the U.S. market in this study had over 200 subject domain areas as part of the degree requirement specifications for DS jobs, reflecting the market need for a broad range of education programs in DS. For example, a master's degree in LIS, information management, and information science tended to be needed for jobs in data services, data curation, and data management. In addition, some domain-specific DS jobs added to the debate of the generalist versus specialist approach to DS education. There has been criticism of the proliferation of DS generalists and a call for practical experts with a focus on professional applications to meet the diverse varied demands of the data workforce with specialists with complementary knowledge and skills to work as a team (Hardoon, 2021; Irizarry, 2020).

Looking at the future trends of new skills of the digital economy, three categories of "new foundational skills" are identified: (1) skills distinctly human (e.g., critical thinking, creativity, communication, relationship building), (2) business enabler skills (e.g., project management), and (3) technological skills (e.g., software development, programming, digital security) and more widely applicable ones (e.g., analyzing data) (Burning Glass, IBM, & Business-Higher Education Forum, 2017). In this large context, the report calls for data science and analytics workforce development and education programs to embrace developing talent for various roles and career pathways in the areas of data engineering, data governance, data lifecycle, data privacy and security, and data product development, with data literacy as a possibility for students in any field of study. iField DS education is well positioned to meet this expectation as found in this study.

One fast-growing, data-driven domain is the biomedical area that is becoming increasingly driven by big data and DS approaches. In the United States, biomedical researchers will soon have to follow the National Institutes of Health's (NIH) Data Management and Sharing Policy (NIH, 2021), which will require all NIH-funded researchers to submit data management and sharing plans as of January 2023. Similar policies in Europe, Asia, and Australia also need data sharing, creating a rapidly growing landscape of open biomedical data. Recognizing this trend, iField graduates are expected to work and collaborate with domain experts who do not have experience with managing and sharing their data or using existing open data for their research.

Additionally, an increasing focus on using real-world data, such as electronic health record data, for biomedical artificial intelligence and machine learning applications will require a workforce that is prepared to make data "AI-ready" by providing expert curation to prepare unstructured or noisy data, address concerns related to data privacy, and assure ethical use of data to avoid bias and unintended consequences, particularly those that affect disadvantaged or marginalized groups. Such work will require not only technical knowledge and skills, like data curation, machine learning, and algorithms, but also an understanding of the ethical and social implications of AI/ML applications, making iField graduates excellent candidates for work in these areas. This is particularly the case in the post-pandemic economy and with the emerging trend of using automating AI (AutoAI or AutoML) that is capable of autonomously and efficiently ingesting and processing big data and creating models based on target goals, and yet human expertise is indispensable (D. Wang et al., 2019; Lund et al., 2021). Challenges of workforce adaptation associated with the growing demand of analytics, AI, automation and the acceleration of digitization are concerns for productivity commissions globally (Mortimer-Lee & Pabst, 2022). To leverage these current and future technologies and take advantage of the long-term trends accelerated by them requires a workforce with skills that extend beyond technological capability to encompass social and cognitive ones as well. In this regard, iField DS particularly benefits from the strong social informatics traditions within iSchools. Working within academic and practice-based contexts where the impact of (emerging) technologies on people and information are the norm helps prepare iField graduates for the uncertainty and dynamics of these emerging sociotechnical ecosystems. The transdisciplinary totality of this iField approach also facilitates working with data across disciplines, contexts and sectors, helping students refine their understanding of various knowledge systems while also honing the critical social intelligence needed for collaboration and cooperation in any workplace.

5.3 | iField DS curriculum

Examining existing iField DS programs and course offerings and connecting DS jobs for iField students can help curriculum design to meet the DS job market demand. The job analysis can reflect the kind of talents that need to be cultivated in DS education, preparing students with the knowledge and skills to meet the needs of jobs. The curriculum of DS programs can be adjusted according to the job demand analysis. Our analysis does not reveal any major gaps between the current iField DS offerings and the job market demand. For example, according to job requirements, skills such as data management and data analysis are commonly needed, and these are commonly covered in iField DS programs. In addition, the commonly needed soft skills and project work for DS jobs are also common in iField coursework that emphasizes practical projects or training that requires teamwork, leadership, or interpersonal communication skills. However, the evolving needs of the DS field require DS professionals to have corresponding competencies and skills. Educating students to acquire these competencies and skills should be the goal of DS education.

During the iDSCC monthly group discussions, observations were noted about unique contributions of iField DS education:

- Consideration of the legal, policy, and ethical issues in DS curriculum design, such as legal issues in health informatics or privacy in the digital age. These curricula provide DS students with an overview of the challenges, theories, and solutions related to these issues throughout the entire DS lifecycle.
- Integration of human factors into DS curriculum design, such as human values in DS or usability assessment. These curricula cover the role of human factors principles, human–computer interactions, and the application of usability assessments in the development and use of DS technology.
- Cultivation of creativity and leadership in DS curriculum design, such as critical thinking via analytics or leadership and people management. Such curricula emphasize the cultivation of critical and innovative thinking habits of DS students, rather than solely focusing on the use of data technologies.

Based on the above study findings, committee discussions, and conference sessions during 2019–2021, the undergraduate subcommittee recommends a core set of iField DS knowledge, skills, and competencies for undergraduate students, which is publicly available (tinyurl.com/e8exzmsc). Compared to a recommendation made by the ACM Data Science Task Force (ACM, 2021), iField undergraduate DS programs offer a unique competency called "Human-Centered Data Science." Human-Centered Data Science competency is an umbrella competency that supplies backdrop to iField DS education and is recommended to be included as a common thread that interweaves throughout the program. Another unique competency is Data Literacy (an ability to understand and use data effectively to inform decisions), which is a foundational basis for other recommended competency skills and knowledge.

Overall, DS education in iField schools integrates the human-oriented concept into the classroom, pays attention to the cultivation of students' personal characteristics, and focuses on the value and impact of DS application on society. Courses of iField DS programs found in this study can also supply the "human-centered" education. iField schools can collaborate with academic units to supply more diversified courses for DS students so that the iField DS education can be more adaptable to the needs of society while supporting its uniqueness. iField educators are uniquely prepared for supplying the DS curriculum given their educational training background, teaching and research expertise, and the interdisciplinary nature of iField and the iSchool culture.

It should be noted that findings and recommendations should be put in the context of the samples used for the Delphi study, DS program and curriculum analyses, and job analysis. Due to the lack of a clear definition of DS and the iField, the samples have their limitations and potential biases. The findings and recommendations about DS curricula and competencies are intended to help inform DS education in iField schools. Individual programs may adjust and customize according to their own institutional structure, resources, vision, curricula, and culture before adoption.

5.4 | Undergraduate versus graduate DS education

In designing iSchools DS curriculum, the differences between graduate-level and undergraduate-level DS education need to be understood. Because very few schools are offering both undergraduate and graduate DS programs, this study did not follow a design with paired samples for a strict comparison. Instead, the two iDSCC subcommittees conducted their respective systematic study on graduate-level and undergraduate-level DS education. In addition, input was sought as part of the Delphi study and a regular committee meeting discussion was held on the topic. The observations and recommendations, as highlighted below, help shed light on the current practices at various DS education levels in the iField schools:

- In terms of the knowledge areas covered, data mining and analysis are the primary common knowledge modules considered. They serve as the foundation for developing DS-related competencies and skills for both undergraduate and graduate DS education in iField schools.
- The focus on theoretical learning appears to be more evident in the clustering results of undergraduate courses. The courses focus primarily on information, data science, and other fundamentals. In contrast, the graduate curriculum contains data coding and modeling, data organization and management, and other technical and practical courses to support students in becoming data scientists.
- Graduate programs are diverse, covering a broad range of knowledge areas and methods, whereas undergraduate programs focus more narrowly on DS.

The interdisciplinarity of DS is a driving factor for the differences in curricula between undergraduate and graduate education. In contrast to other fields, normally there is not a consecutive DS program that starts at the undergraduate level and builds upon these foundations for more advanced topics at the graduate level. Instead, the graduate courses mostly address students from various disciplines, with the goal of providing DS skills that then can be applied in the respective discipline. The requirements vary and often depend on the field of the faculty providing the program—for example, a computer science department offering a DS master's might expect computer science skills or even a BS in computer science from its applicants. However, the more interdisciplinary the applicants, the more diverse their foundations to be built upon. From a DS perspective, this means that even at the graduate level, basic technical skills often must be learned first. At the undergraduate level, this foundation of different disciplines and prior experiences is not a factor. With more time available and less diverse backgrounds and expectations of the students, the undergraduate programs tend to be more technical and often have a higher overlap with computer science than the graduate programs.

5.5 | Challenges and considerations

Although we have answered important questions facing DS education in the iField, there are still challenges that need to be addressed:

- Most DS programs in iSchools are designed to address a growing need for combining technical skills with social good to solve important data and information problems. However, the field of DS continued to evolve while these programs were being designed and run through curriculum development and instructor hiring. As the DS field has matured, it is time for these programs to rethink their curricula, hiring practices, and differentiating factors with respect to other DS programs on campus. In many institutions, larger efforts to develop a stronger, richer, and deeper offering of DS programs have been underway over the last few years. The iSchools need to both fit into these efforts and carve out what makes the iSchool flavor of DS special (Shah et al., 2021).

- Another big challenge for iSchool DS programs is to balance immediate need and opportunity with long-term and sustainable development. Hiring qualified instructors who meet the high bar of skills and pedagogy has been challenging, but even more difficult is retaining good instructors and program directors that can provide a longer-term vision and continuity. Often, students who join iSchool DS programs are either misinformed about or have misunderstood the uniqueness of DS in an iSchool. If they are looking to find a job as a software engineer, either we have failed them or done a poor job of constructing and conveying our curriculum objectives through strong actors such as the program director/coordinator/chair and instructors.
- Practically, as more disciplines and academic units are offering DS programs, iField schools and programs need to articulate our values to institutional administrators, colleagues, employers, and students alike about what and how we teach and what iField DS education contributes to the broader DS education spectrum and career opportunities for graduates from our programs. The findings of the curriculum and job analyses offer supporting evidence that DS programs from multiple disciplines can co-exist to meet the workforce market and those academic programs can be natural collaborators such as computer science and statistics in the sample iSchool undergraduate DS programs.
- Locally, there will always be translational challenges for iSchool DS programs. One size cannot fit all in relation to DS curricula any more than we have seen in other iSchool education programs. Concerns about localization of DS curriculum were prevalent in our findings. Not only will each school have to respond to the demands of their local market and student catchment when designing their curriculum, but also to the institutional dynamics at the point of design and re-design of their program.
- The demands of designing a forward-looking iSchool DS curriculum focused on society's contemporary and future data challenges will be ongoing. At no point can we rest easy and assume that we have arrived at a successful formula for our programs. The sociotechnical complexities of information and data alongside the social inequalities that will continue to emerge as we build ever more complex networked systems will call upon us to refresh and review our content and our commitment to social justice. However, the core values reflected in iField discourse over the decades suggest that the capacity to build up the collective resilience of our staff and our students to engage in the uncertainties of complex data and AI-informed futures is comfortably in our DNA.

6 | CONCLUSION, LIMITATIONS, AND FUTURE WORK

The interdisciplinary nature of DS offers great opportunities for individual disciplines to develop programs that uniquely contribute to the broader DS education landscape from their disciplinary perspective and meet the demand for a diverse DS workforce. The iDSCC identified key questions facing DS education in the iField and conducted a series of studies that resulted in rich findings. The findings offer a snapshot of the current state of DS education in iField schools at both graduate and undergraduate levels and highlight the career opportunities for iField DS graduates. The findings suggest an iField approach to DS education and identify a core set of competencies and skills recommended for inclusion in the core curriculum along with suggested considerations for adoption in the local context of individual programs and institutions.

DS is rapidly evolving with the swift advances in new technologies, powerful big data tools, widespread applications of AI-augmented solutions to many different domains, and strong societal demands of human-centered DS disciplines such as ethics, privacy, and AI-human collaboration. As such, DS education in the iField should also accommodate this evolution with emphases on soft and applied DS skills. DS students should be educated not only to solve data-intensive problems in iField disciplines but also to work on projects at various stages of DS lifecycles by bridging among various DS stakeholders.

As the complexity and diversity of the DS landscape is likely to continue with diverse perspectives as well as new challenges and opportunities, we need agile, flexible approaches to both the content and practice of DS. Developing pedagogies that promote critical reflection on DS practices will provide iField DS students with a socially rich perspective of data practices in context that foregrounds concern about the vulnerability of data to misinterpretation, misuse, and misappropriation, raising questions about how social values can be integrated into data infrastructure. In addition, stressing the importance of work-integrated and lifelong learning, iField DS students will possess the shapeshifting skills necessary to succeed in the dynamics of this still emerging field.

iField schools are well positioned for an interdisciplinary and transdisciplinary approach that enables us to build bridges to other disciplines and contexts to enrich understandings of DS practice. At the same time, such an approach creates challenges in terms of how to "sell" what is special about the iField approach to DS education to our institutions, our students (and their parents), and prospective employers. However, the iField has always been about navigating the challenges of multiple perspectives and debates about core identities. It is a quality that King (2006) and Burnett and Bonnici (2013) alert us to be careful to preserve and which will help us future-proof our brand of DS education.

This study has limitations and suggests future work regarding DS education in iField schools:

- Due to the lack of a clear and agreeable definition of DS and the iField, the samples in this study were based on labels of data/data science and member status at the time of the study (e.g., iSchool membership, LIS schools in ALISE, or ALA directory). The findings and recommendations regarding DS curricula and competencies should be put in the context of the respective scope. Given the evolving nature of DS and iField, continued research in this line is needed to update and further explore DS curriculum.
- Given very few schools offering both undergraduate and graduate DS programs, the program and course survey and analysis were conducted by separate subcommittees using a different sampling approach. Future research can adopt a design with paired samples for an in-depth case study of those programs that offer DS programs at both levels for a better comparison and understanding about the differences between undergraduate and graduate DS education.
- Like any job and career opportunities analysis, the samples and results are only a snapshot of a time and geographic location and reflect only the past trends up to the time of data collection. In particular, the European job advertisements were posted during the peak time of the COVID-19 pandemic which had a strong impact on the job market. Another consideration for such analysis is the targeted audience of the job data source and platforms where the job advertisements are posted. Given the widespread geographic locations of iField schools, the large, diverse job advertisement datasets were gathered from different platforms and at various times for analysis and comparisons across four countries in three continents. Future efforts may follow a similar approach and include other important markets such as India and Latin America.
- Finally, the iDSCC's charges have been aiming at disciplinary characteristics of DS curriculum. Given the wide geographic range of iField schools, an examination of differences of DS programs in different geographic areas of the world would be helpful for institutions to adopt general curriculum recommendations locally.

ACKNOWLEDGMENTS

This paper is a result of the iSchool Data Science Curriculum Committee (iDSCC) formed by 2019–2021 Chair of iSchools Organization Sam Oh and co-chaired by Il-Yeol Song and Javed Mostafa. Other iDSCC members who contributed to the work include Myongho Yi. We are incredibly grateful for the valuable feedback and input from the community of the iField throughout the 2-year process of the committee work. We also thank those individuals who participated in our surveys and provided data source for DS jobs in the ALA JobLIST archive. The research assistance from Yuan Hu, Fangli Su, Brenna Hubschman, and Taylor Fields is appreciated.

ORCID

Yin Zhang https://orcid.org/0000-0002-6611-2094
Dan Wu https://orcid.org/0000-0002-2611-7317
Loni Hagen https://orcid.org/0000-0002-6532-0852
Il-Yeol Song https://orcid.org/0000-0001-7706-959X
Sam Oh https://orcid.org/0000-0001-8488-2079
Theresa Anderson https://orcid.org/0000-0003-1792-3728
Chirag Shah https://orcid.org/0000-0002-3797-4293
Bradley Wade Bishop https://orcid.org/0000-0002-5022-2707
Frank Hopfgartner https://orcid.org/0000-0003-0380-6088
Lisa Federer https://orcid.org/0000-0001-5732-5285
Jeffrey S. Saltz https://orcid.org/0000-0002-8913-1095

REFERENCES

ACM Data Science Task Force. (2021). *Computing competencies for undergraduate data science curricula*. Association for Computing Machinery. https://www.acm.org/binaries/content/assets/education/curricula-recommendations/dstf_ccdsc2021.pdf

Aerts, M., Molenberghs, G., & Thas, O. (2021). Graduate education in statistics and data science: The why, when, where, who, and what. *Annual Review of Statistics and Its Application*, *8*, 25–39.

Albright, K., & Mehra, B. (2020). *Information for a sustainable world: Addressing society's grand challenges*. Paper presented at the Proceedings of the 83rd Annual Meeting, ASIS&T 2020.

Anderson, T. D., & Parker, N. (2019). Keeping the human in the data scientist: Shaping human-centered data science education. *Proceedings of the Association for Information Science and Technology*, *56*(1), 601–603.

Bailer, A. J., & Fisher, N. I. (2020). Discussion of "A review of data science in business and industry and a future view.". *Applied Stochastic Models in Business and Industry*, *36*(1), 20–22.

Baruchson-Arbib, S., & Bronstein, J. (2002). A view to the future of the library and information science profession: A Delphi study. *Journal of the American Society for Information Science and Technology*, *53*(5), 397–408.

Bates, J., Cameron, D., Checco, A., Clough, A., Hopfgartner, F., Mazumdar, S., Sbaffi, L., Stordy, P., & de la Vega de Leon, A. (2020). *Integrating FATE/critical data studies into data science curricula: Where are we going and how do we get there?* Paper presented at the ACM FAT* 2020.

Bawden, D. (2008). Smoother pebbles and the shoulders of giants: The developing foundations of information science. *Journal of Information Science*, *34*(4), 415–426.

Berman, F., Rutenbar, R., Hailpern, B., Christensen, H., Davidson, S., Estrin, D., Franklin, M., Martonosi, M., Raghavan, P., Stodden, V., & Szalay, A. S. (2018). Realizing the potential of data science. *Communications of the ACM*, *61*(4), 67–72.

Blake, C., & Brown, C. (2019). *Information anyone, anywhere, any time, any way*. Paper presented at the Proceedings of the 82nd Annual Meeting, ASIS&T 2019.

Bowers, M. R., Camm, J. D., & Chakraborty, G. (2018). The evolution of analytics and implications for industry and academic programs. *Interfaces*, *48*(6), 487–499.

Brodie, M. L. (2019). What is data science? In *Applied data science* (pp. 101–130). Springer.

Bruce, H. (2011). The audacious vision of information schools. *Journal of Library and Information Science*, *37*(1), 4–10.

Buckland, M. (1999). The landscape of information science: The American Society for Information Science at 62. *Journal of the American Society for Information Science*, *50*(11), 970–974.

Burnett, K., & Bonnici, L. J. (2013). Rhizomes in the iField: What does it mean to be an iSchool? *Knowledge Organization*, *40*(6), 408–413.

Burning Glass, IBM, & Business-Higher Education Forum. (2017). *The quant crunch: How the demand for data science skills is disrupting the job market*. Boston, MA. https://www.bhef.com/sites/default/files/bhef_2017_quant_crunch.pdf

Cao, L. (2017). Data science: A comprehensive overview. *ACM Computing Surveys*, *50*(3), 1–42.

Cao, L. (2018a). What is data science? In *Data science thinking* (pp. 29–58). Springer.

Cao, L. (2018b). Data science education. In *Data science thinking* (pp. 329–348). Springer.

Carter, D., & Sholler, D. (2016). Data science on the ground: Hype, criticism, and everyday work. *Journal of the Association for Information Science and Technology*, *67*(10), 2309–2319.

Dillon, A. (2012). What it means to be an iSchool. *Journal of Education for Library and Information Science*, *53*(4), 267–273.

Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, *26*(4), 745–766.

Elkhatib, Y. (2017). *Navigating diverse data science learning: Critical reflections towards future practice*. Paper presented at the IEEE International Conference on Cloud Computing Technology and Science.

Farradane, J. (1976). Towards a true information science. *The Information Scientist*, *10*(3), 91–101.

Galliers, R. D. (2003). Change as crisis or growth? Toward a trans-disciplinary view of information systems as a field of study: A response to Benbasat and Zmud's call for returning to the IT artifact. *Journal of the Association for Information Systems*, *4*(1), 13–352.

Golub, K. (2019). *From a library and information science department to a transdisciplinary universitywide iSchool: A model of Linnaeus University*. Paper presented at the 27th Symposium of the European Association for Library and Information Education and Research, BOBCATSSS'19.

Gunderman, H. C. (2019). *Exploring learning in a global information context*. Paper presented at the Proceedings of the Association for Library and Information Science Education Annual Conference, ALISE 2019.

Gunderman, H. C. (2020). *Transforming LIS education in an interconnected world*. Paper presented at the Proceedings of the Association for Library and Information Science Education annual conference, ALISE 2020.

Hagen, L. (2020). Teaching undergraduate data science for information schools. *Education for Information*, *36*(2), 109–117.

Hardoon, D. R. (2021). Is data science education a jack of all trades? *Harvard Data Science Review*, *3*(1). https://doi.org/10.1162/99608f92.d1aaf1d7

Hayashi, C. (1998). What is data science? Fundamental concepts and a heuristic example. In *Data science, classification, and related methods* (pp. 40–51). Springer.

Holmberg, K., Tsou, A., & Sugimoto, C. R. (2013). The conceptual landscape of iSchools: Examining current research interests of faculty members. *Information Research*, *18*(3). http://InformationR.net/ir/18-3/colis/paperC32.html

iDSCC. (2019a). *Model data science curriculum for iSchools: The iSchool Data Science Committee (iDSCC) update*. Conference presentation at the iConference 2019.

iDSCC. (2019b). *Data science education in the iSchool context*. Conference presentation at the 82nd Annual Meeting, ASIS&T 2019.

iDSCC. (2020). *iSchool Data Science Curriculum Committee (iDSCC) update*. Conference presentation at the iConference 2020.

iDSCC. (2021). *iSchool Data Science Curriculum Committee (iDSCC) update*. Conference presentation at the iConference 2021.

Irizarry, R. A. (2020). The role of academia in data science education. *Harvard Data Science Review*, *2*(1). https://doi.org/10.1162/99608f92.dd363929

King, J. L. (2006). Identity in the I-school movement. *Bulletin of the American Society of Information Science and Technology*, *34*(4), 13–15.

Kochtanek, T. R., & Hein, K. K. (1999). Delphi study of digital libraries. *Information Processing & Management*, *35*(3), 245–254.

Larsen, R. L. (2017). iSchools. In J. D. McDonald & M. Levine-Clark (Eds.), *Encyclopedia of library and information sciences* (4th ed., pp. 2536–2541). CRC Press.

Linstone, H. A., & Turoff, M. (Eds.). (1975). *The Delphi method* (pp. 3–12). Addison-Wesley.

Lund, S., Madgavkar, A., Manyika, J., Smit, S., Ellingrud, K., Meaney, M., & Robinson, O. (2021). *The future of work after COVID-19*. McKinsey Global Institute.

Marchionini, G. (2016). Information science roles in the emerging field of data science. *Journal of Data and Information Science*, *1*(2), 1–6.

Miah, S. J., Solomonides, I., & Gammack, J. G. (2020). A design-based research approach for developing data-focused business

curricula. *Education and Information Technologies*, *25*(1), 553–581.
Mortimer-Lee, P., & Pabst, A. (2022). *Covid-19 and productivity: Impact and implications* (No. 62). National Institute of Economic and Social Research.
National Institutes of Health (NIH). (2021). *NIH strategic plan for data science*. https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf
O'Neil, C., & Schutt, R. (2013). *Doing data science: Straight talk from the frontline*. O'Reilly Media.
Ozsu, M. T. (2020). A systematic view of data science. *IEEE Database Engineering Bulletin*, *43*(3), 3–11.
Raban, D. R., & Gordon, A. (2020). The evolution of data science and big data research: A bibliometric analysis. *Scientometrics*, *122*(3), 1563–1581.
Rosenbaum, H., & Fichman, P. (2019). Algorithmic accountability and digital justice: A critical assessment of technical and sociotechnical approaches. *Proceedings of the Association for Information Science and Technology*, *56*(1), 237–244.
Saldana, J. (2015). *The coding manual for qualitative researchers* (3rd ed.). Sage.
Saltz, J. S., Dewar, N. I., & Heckman, R. (2018). *Key concepts for a data science ethics curriculum*. Paper presented at the Proceedings of the 49th ACM Technical Symposium on Computer Science Education.
Saltz, J. S., & Grady, N. W. (2017). *The ambiguity of data science team roles and the need for a data science workforce framework*. Paper presented at the IEEE International Conference on Big Data.
Saracevic, T. (1979). An essay on the past and future (?) of information science education—I: Historical overview. *Information Processing & Management*, *15*(1), 1–15.
Saracevic, T. (1999). Information science. *Journal of the American Society for Information Science*, *50*(12), 1051–1063.
Seadle, M., & Greifeneder, E. (2007). Envisioning an iSchool curriculum. *Information Research*, *12*(4). http://InformationR.net/ir/12-4/colis/colise02
Shah, C., Anderson, T., Hagen, L., & Zhang, Y. (2021). An iSchool approach to data science: Human-centered, socially responsible, and context-driven—A position paper. *Journal of the Association for Information Science and Technology*, *72*(6), 793–796.
Siebes, A. (2018). Data science as a language: Challenges for computer science—A position paper. *International Journal of Data Science and Analytics*, *6*(3), 177–187.
Song, I. Y., & Zhu, Y. J. (2016). Big data and data science: What should we teach? *Expert Systems*, *33*(4), 364–373.
Song, I. Y., & Zhu, Y. J. (2017). Big data and data science: Opportunities and challenges of iSchools. *Journal of Data and Information Science*, *2*(3), 1–18.
Sundqvist, A., Berget, G., Nolin, J., & Skjerdingstad, K. I. (2020). *Sustainable digital communities*. Paper presented at the Proceedings of the 15th International Conference, iConference 2020.
Tanweer, A. (2018). *Data science of the social: How the practice is responding to ethical crisis and spreading across sectors* (Doctoral dissertation).
Taylor, N. G., Christian-Lamb, C., Martin, M. H., & Nardi, B. (2019). *Information in contemporary society*. Paper presented at the Proceedings of the 14th International Conference, iConference 2019.
Vicario, G., & Coleman, S. (2020). A review of data science in business and industry and a future view. *Applied Stochastic Models in Business & Industry*, *36*(1), 6–18.
Virkus, S., & Garoufallou, E. (2019). Data science from a library and information science perspective. *Data Technologies and Applications*, *53*(4), 422–441.
Virkus, S., & Garoufallou, E. (2020). Data science and its relationship to library and information science: A content analysis. *Data Technologies and Applications*, *54*(5), 643–663.
Wang, D., Weisz, J. D., Muller, M., Ram, P., Geyer, W., Dugan, C., Tausczik, Y., Samulowitz, H., & Gray, A. (2019). Human–AI collaboration in data science: Exploring data scientists' perceptions of automated AI. *Proceedings of the ACM on Human–Computer Interaction*, *3*(CSCW), 1–24.
Wang, L. (2018). Twinning data science with information science in schools of library and information science. *Journal of Documentation*, *74*(6), 1243–1257.
Wildemuth, B. M. (Ed.). (2016). *Applications of social research methods to questions in information and library science*. ABC-CLIO.
Wu, D. (2019). *Research on data science curriculum* [Conference presentation]. Association for Information Science & Technology Annual Meeting, Melbourne, Australia.
Wu, D., He, D., Jiang, J., Dong, W., & Vo, K. T. (2012). The state of iSchools: An analysis of academic research and graduate education. *Journal of Information Science*, *38*(1), 15–36.
Wu, D., Sun, Y., & Xu, H. (2021). Multidisciplinary comparative research on graduate education in data science. *Library Tribune*, *41*(11), 108–117.
Zhang, Y., & Salaba, A. (2009). What is next for functional requirements for bibliographic records? A Delphi study. *The Library Quarterly*, *79*(2), 233–255.

How to cite this article: Zhang, Y., Wu, D., Hagen, L., Song, I.-Y., Mostafa, J., Oh, S., Anderson, T., Shah, C., Bishop, B. W., Hopfgartner, F., Eckert, K., Federer, L., & Saltz, J. S. (2022). Data science curriculum in the iField. *Journal of the Association for Information Science and Technology*, 1–22. https://doi.org/10.1002/asi.24701