Kurdish Studies
August 2021
Volume: 09, No: 2, pp 283-287
ISSN: 2051-4883 (Print) | ISSN 2051-4891 (Online)
www.KurdishStudies.net

DOI: 10.53555/ks.v9i2.3469

Information and Data Sciences: A Pragmatic Perspective on Their Evolution

Dr. Sandip Banerjee[1*], Hinia Jeram[2], Hage Dipmala[3]

[1*]Assistant Programmer, Department of Election, Govt. of Arunachal Pradesh
[2]Assistant System Analyst, Department of Election, Govt. of Arunachal Pradesh
[3]Assistant Programmer, Department of Election, Govt. of Arunachal Pradesh

ABSTRACT
The article explores the historical development, current state, and future directions of data science and information science. It traces the origins of data science in statistics and computer science, and information science in library science and interdisciplinary programs. The Emergence of iSchools is discussed as an important development bridging these fields. Currently, many institutions offer data science programs, with iSchools integrating data science into information science curricula. The article argues that information science brings unique strengths to data science education, including a transdisciplinary approach, human-centered focus, emphasis on context, and ethical considerations. Looking ahead, it recommends that data science education in information schools should teach fundamental skills while making the field more accessible, expose students to real-world applications, and leverage information science's established strengths. The integration of data science and information science presents opportunities to address complex challenges at the intersection of technology and society.

1. Prolegomena

At first glance, establishing definitions and distinctions between data science and information science may seem straightforward. One might assume that simply examining the meanings of 'data', 'information', and 'science' would suffice to delineate these two scientific branches and their relationships. However, if it were that uncomplicated, we wouldn't be grappling with these fundamental questions about their significance. In reality, I believe that data science and information science are more defined by their presentation, perception, and application than by formal definitions or distinctions. I will share my perspectives based on personal experiences and those of numerous colleagues and collaborators in these fields, drawing from both academic and industry settings. While my focus is primarily on the US context, some aspects incorporate and may have relevance to other global regions.

I will begin with the more straightforward task of elucidating data science and information science through their historical development and evolution (Section 2). Subsequently, I will explore how these sciences are currently taught and implemented (Section 3). Lastly, I will offer some thoughts (Section 4) on their future trajectories (projections) and what I believe their direction should be (perspectives).

2. Antecedent

Although this article aims to connect and contrast data science and information science, these disciplines originated from distinct domains with different objectives and structures. Data science has its foundations in statistics, while information science emerged from library science or various interdisciplinary programs. In this section, we will attempt to trace their origins to understand their initial purposes and later question how they are fulfilling those goals.

2.1.Data science

John Tukey's 1962 publication "The Future of Data Analysis" is widely recognized as the genesis of what we now call data science (Donoho, 2017; University of Virginia). In this seminal work, Tukey explored the burgeoning field of "data analysis," which David Donoho later equated to modern data science in his "50 years of Data Science" (Donoho,2017). The term "data science" itself was first coined by Peter Naur in 1974 in his "Concise Survey of Computer Methods," where he described it as "the science of handling data, once they have been established, while the relationship of the data to what they represent is assigned to other fields and sciences" (Cao, 2017). The discipline of data science subsequently emerged from related fields such as computer science, mathematics, and statistics as a means to process and structure rapidly expanding databases.

In 1985, statistician Chien-Fu Jeff Wu proposed data science as an alternative term for statistics during a lecture at the Chinese Academy of Sciences, reiterating this suggestion in a 1998 lecture titled "Statistics Data Science?" (Donoho, 2017). The term gained traction throughout the 1990s and early 2000s. In 2001, another statistician, William Cleveland, proposed a framework for the emerging field of data science, describing it as "a new discipline, broader than statistics" that would aid professionals in data computation (Brady, 2019; Cleveland, 2001; Yan & Davis, 2019). The International Council for Science: Committee on Data for Science and Technology launched the Data Science Journal in 2002, followed by Columbia University's The Journal of Data Science in 2003 (Yan & Davis, 2019).

From the mid-2010s onwards, numerous institutions worldwide have established dedicated data science programs and departments. Alternatively, some universities opted to incorporate data science as a specialization within existing master's programs in fields such as computer science, information science, business, or statistics, providing a swift method to introduce data science education on campus. This approach has led to multiple data science offerings at a single institution, often causing confusion among students. For instance, at the University of Washington, data science concentrations and specializations are available through various departments, including the Information School, Computer Science and Engineering, and Applied and Computational Math, alongside a standalone master's program in data science. This diversity of options frequently leaves students and faculty alike grappling to understand the distinctions between these programs.

In my own definition, data science is described as "a field of study and practice that involves collection, storage, and processing of data in order to derive important insights into a problem or a phenomenon" (Shah, 2020). However, the practical implications of this definition, particularly within the context of information science, require further exploration. To fully grasp this concept, it is necessary to first examine the nature and origins of information science.

2.2. Information science

Tracing the history of information science presents challenges due to its interdisciplinary nature, drawing from fields such as library science, computer science, and communications. The core principles and definitions of this discipline have been subject to ongoing debate since its emergence as a distinct field in the 1970-1980s, with no clear resolution (Bawden, 2008; Pawley, 2005, pp. 223–238; Rayward, 1996). The American Documentation Institute (ADI) was established on March 13, 1937, aiming to enhance scientific publishing documentation and circulation through microfilm technology, as noted by Lerner (Lerner, 1994). Schultz and Garwig (Schultz & Garwig, 1969) provide a comprehensive history of ADI. Following World War II, the proliferation of scientific and technological advancements led to extensive microfilm projects, resulting in the expansion and restructuring of ADI. In 1968, it was renamed ASIS (American Society of Information Science), later becoming ASIST in 2000 with the addition of 'T' for 'Technology'.

While the evolution of ASIST parallels that of information science, the discipline's foundation has another related narrative. Rayward (Rayward, 1996) attributes the term "information science" to the post-war computer revolution and suggests that its history must incorporate various related historical studies, including the history of science and technology, printing and publishing, and information institutions like libraries, archives, and museums. Bawden (Bawden, 2008) highlights ongoing debates about the philosophical and practical foundations of information science in the early issues of the Journal of Information Science, starting in 1979. These debates encompass controversies surrounding the definition of Information Science, its philosophical principles as proposed by Bertie Brookes (considered the founder of the "cognitive approach" to information science), and the "exact nature of the discipline." Bawden concludes that discussions continue regarding the foundational aspects of this science, its multidisciplinary nature versus distinctiveness, appropriate education in the field, and the core theories underlying its application.

Although the discourse on what constitutes information science persists, the iSchools movement serves as a prime example of the field's clear recognition.

2.3. iSchools

The emergence of iSchools and the associated movement stemmed from discussions about the interrelationship between information science and library science. As technology progressed, the interdisciplinary connections between these fields became increasingly complex (Chakrabarti & Mandal, 2017; Shu & Mongeon, 2016). Experts in the field concur that the iSchool movement originated in 1988 with the 'Gang of Three,' comprising academic professionals from diverse backgrounds in information, technology, and library studies at the University of Pittsburgh, Syracuse University, and Drexel University (Chakrabarti & Mandal, 2017; Dillon, 2012, pp. 267–273; Shu & Mongeon, 2016). King (2006) stated, "The initial purpose of this small group was to share information and facilitate interaction when facing the new intellectual and professional challenges in the field of information science." The group expanded to the Gang of Four in 1990 with Rutgers University joining, and became the Gang of Five in 2001 with the addition of the universities of Washington and Michigan (Rutgers was no longer part of the Gang by 2001). By 2003, Florida State University and the universities of Illinois, North Carolina, Indiana, and Texas had joined, forming the Gang of Ten (Shu & Mongeon, 2016).

From 2003 to 2005, the group began meeting biannually to continue discussions on the intersections and evolution of the information, technology, and library fields. During this period, it was renamed the "iCaucus," and by 2005, approximately eighty schools worldwide had contributed to the movement (Shu & Mongeon, 2016). Shu and Mongeon (Shu & Mongeon, 2016) highlighted that 14 of the original 19 members offered American Library Association (ALA) accredited graduate degree programs, demonstrating the iCaucus' roots in library and information science (LIS). Many current members still maintain this accreditation. In 2008, the iCaucus initiated the 'iConference,' described by Chakrabarti and Mandal (2017) as a "forum in which iSchools' deans share their collective interests through the websites of iSchools."

3. Contemporary

In the present day, iSchools are presenting their own interpretation of data science. This section will explore that interpretation, enabling us to better comprehend how data science and information science coexist and provide intriguing solutions to identical problems.

At present, there are well over a hundred data science and related programs available nationwide, offering both undergraduate and graduate degrees. Many of these programs incorporate capstone or portfolio graduation requirements to provide students with a comprehensive experience in the field. Additionally, more than 120 colleges and universities worldwide are affiliated with the iSchool organization.

As data science gained popularity over the years, iSchools and LIS programs naturally integrated data science offerings into their information science curricula, resulting in numerous data science concentrations, specializations, certificates, and degrees. These offerings often emerged in response to market demands and opportunities. In some cases, it was a way for iSchools to stake a claim in a segment of data science that is more human-centric and context-dependent, much like information science itself (Shah et al., 2021).

However, by 2018, it became evident that creating and running such data science programs without understanding the broader landscape of data science was no longer feasible. This realization stemmed from three main factors. Firstly, the field of data science was maturing, with clear trends emerging from industry and academia that shaped specific expectations for data science practitioners. Gone were the days when any course or expertise could be labeled as data science. Secondly, students began questioning the differences between various data science programs on campus. Those unable to gain admission to one program often turned to another, assuming similarity, only to be disappointed later. The distinction between data science programs offered by information science, computer science, or business schools became a point of inquiry. Lastly, as campus resources became constrained, upper-level administrators started questioning the necessity and sustainability of multiple data science curricula. They needed to determine which data science program to highlight when pitching to potential donors.

Without a mechanism to promote campus-wide collaboration, various programs within a university may end up in competition rather than cooperation. Given that information science programs are still grappling with questions of identity, it's understandable that they found it crucial to explore the definition of data science within their field.

In 2019, the iSchools organization established the iSchool Data Science Curriculum Committee (iDSCC) to gain a better understanding of data science's structure in undergraduate and graduate programs. Through global collaboration among information science scholars and programs, they conducted research and developed proposals.

Additional research emphasizes the importance of offering students hands-on experience beyond the classroom (Elkhatib, 2017; Wang, 2018). Data science is considered highly interdisciplinary, intersecting with fields such as library and information science, mathematics, and statistics (Brady, 2019; Donoho, 2017). Specifically, data science concentrates on developing data management and programming skills that enable students to comprehend the ethical implications of information and apply this knowledge to more human-centric applications.

Both library science and information science deal with the collection, organization, retrieval, and preservation of knowledge. They share the goal of connecting people with information and knowledge (Pawley, 2005, pp. 223–238). The synergy between data science and information science is evident in their teaching and practical applications. Wang (Wang, 2018) contends that Information Science education emphasizes "human information behavior, information ecology, knowledge management, and bibliometrics." These priorities reflect the aforementioned value of human-information connectivity and are thus incorporated into the structure of LIS programs. However, information science education and practices need to enhance their technological approach, as they lag behind data scientists in computational proficiency (Pawley, 2005, pp. 223–238; Wang, 2018).

In conclusion, data science and information science must develop curricula that stress the significance of empirical work, understanding social demands and ethical concerns arising from technological advancements, expanding computational and technological innovations, and enhancing information literacy to address real-world issues. The subsequent section will delve into the specific implications for the future of both fields.

4. Futurity

The preceding sections have demonstrated that data science and information science have established themselves as distinct and robust fields. While the Data-Information-Knowledge-Wisdom (DIKW) hierarchy traditionally places information above data, it may not be appropriate to view data science and information science through this same lens. This perspective is crucial when considering their future trajectories and potential directions.

Both disciplines emerged from the need to address specific problems that other fields were not adequately tackling at the time. As they gained success and recognition in solving these issues, questions arose about their true nature and purpose. Additionally, practical considerations in education and employment necessitate identifying their unique strengths and differences, as discussed earlier. These ongoing discussions and inquiries are likely to continue, which is beneficial as they keep the fields accountable and forward-looking.

With this context in mind, let's examine the integration of data science and information science, focusing on what makes the former's offering by the latter unique and compelling. I contend that when information science programs teach data science curricula, four distinctive characteristics emerge. Educators should ensure these qualities, students should consider them when selecting a data science program, and employers should take note of them to meet their data science recruitment needs.

- ✓ **Multidisciplinary:** Information science is inherently interdisciplinary, but it goes beyond that. Rather than simply operating across disciplines, information science promotes collaboration between fields, providing a transdisciplinary framework (Gibbs, 2017). While many programs and fields claim to do this, information science has embraced transdisciplinarity from its inception. In this field, you'll encounter numerous individuals, projects, and opportunities that genuinely embody transdisciplinarity in a meaningful way. Many significant data science challenges span multiple disciplines and can be more effectively addressed by working not just across but in conjunction with these various fields. Information science can provide robust training for this approach.

- ✓ **Human-centered approach:** Information science distinguishes itself by prioritizing human considerations. This principle is crucial in data science processes, as it transforms statistical and computational problem-solving into a more comprehensive discipline when data is examined with proper context and human needs in mind. Unlike other fields that separate "data science for social good" or "data and society," information science inherently integrates societal concerns and social welfare into its core practices.
- ✓ **Contextual significance:** Understanding data-derived information necessitates considering the context of data collection, storage, analysis, and application. Information science consistently emphasizes context, with professionals in the field trained to view it as a fundamental concept. This approach naturally lends itself to extracting meaningful insights from data science problems. Any data science education within an information science curriculum must stress this contextual importance.
- ✓ **Ethical and responsible focus:** Information science has long prioritized fairness, equity, accountability, and ethics in its investigations, predating the recent attention to computational and data biases. It would be a missed opportunity not to leverage this established strength in data science studies.

Having examined information science's potential contributions to data science, let's consider how to implement these ideas in practice. Note that 'iSchools' is used here to collectively refer to information science schools and programs.

- Fundamentally, all data science programs, regardless of their institutional setting, must teach essential data science skills. Students should master statistical analysis and understand concepts such as overfitting, spurious correlations, and Simpson's paradox (Wagner, 1982) before engaging in discussions about social perspectives on bias and fairness (often covered under "data science for social good" or "responsible AI").
- Many perceive data science and computational sciences as domains exclusive to those with strong coding abilities and a "hacker mentality." While programming skills are essential, data science encompasses much more. To prevent discouraging potential learners, particularly women and certain minority groups, from pursuing data science, it's crucial to create a more accessible entry point. This approach should explore fundamental concepts without immediately delving into complex datasets and programming challenges. For instance, Edwin Hubble's 1929 paper (Hubble, 1929), which provides empirical evidence for the Big Bang theory, is easily comprehensible. It features a simple table with 24 observations and a clear scatterplot demonstrating the relationship between a distant object's distance from Earth and its recession velocity. This regression problem can be understood without coding or database expertise and relates to concepts familiar to high school students.
- Although the initial approach to data science should be gradual, it shouldn't limit the field's potential depth. Classroom constraints and infrastructure limitations often hinder work with large-scale datasets or real-world problems, but these shouldn't be disregarded. While students may not launch new products in class, they can still gain exposure to concepts like A/B testing. This can be achieved through studying relevant research, inviting industry professionals as guest speakers, or organizing field trips to observe real-world applications.
- The majority of data science programs in iSchools are professionally oriented, catering to students seeking career advancement and job opportunities. While some may pursue doctoral studies, this is uncommon based on experience. Consequently, it's vital to pay close attention to industry trends (not fads) and incorporate them into the curriculum. Internships offer an excellent opportunity for students to gain firsthand experience in the field.

Summary

The evolution and integration of data science and information science have led to a unique and compelling approach to addressing complex data challenges. Information science programs offering data science curricula bring four distinctive characteristics: a transdisciplinary framework, a human-centered approach, an emphasis on contextual significance, and a focus on ethical and responsible practices. These qualities position information science-based data science programs to effectively tackle real-world problems that span multiple disciplines and require consideration of societal impacts.

To implement these ideas in practice, it is crucial to:

1. Teach essential data science skills while maintaining accessibility for diverse learners.
2. Create approachable entry points to data science concepts without sacrificing depth.
3. Expose students to real-world applications and large-scale datasets despite classroom constraints.
4. Align curricula with industry needs while preserving the unique strengths of information science.

By leveraging these strengths and addressing practical considerations, information science programs can continue to contribute significantly to the field of data science, producing well-rounded professionals equipped to navigate the complex landscape of data-driven decision-making in various contexts.

References

1. Bawden, D. (2008). Smoother pebbles and the shoulders of giants: The developing foundations of information science. Journal of Information Science, 34(4), 415–426.
2. Brady, H. E. (2019). The challenge of big data and data science. Annual Review of Political Science, 22, 297–323.
3. Cao, L. (2017). Data science: A comprehensive overview. ACM Computing Surveys, 50(3),
4. 1–42.

5. Chakrabarti, A., & Mandal, S. (2017). The ischools: A study. Library Philosophy & Practice.
6. Cleveland, W. S. (2001). Data science: An action plan for expanding the technical areas of the field of statistics. International Statistical Review, 69(1), 21–26.
7. Dillon, A. (2012). What it means to be an ischool. Journal of education for library and information science.
8. Donoho, D. (2017). 50 years of data science. Journal of Computational & Graphical Statistics, 26(4), 745–766.
9. Elkhatib, Y. (2017). Navigating diverse data science learning: Critical reflections towards future practice. In 2017 IEEE international conference on cloud computing technology and science (CloudCom) (pp. 357–362). IEEE.
10. Gibbs, P. (2017). Transdisciplinary higher education: A theoretical basis revealed in practice.
11. Springer.
12. Hubble, E. (1929). A relation between distance and radial velocity among extra-galactic nebulae. Proceedings of the National Academy of Sciences, 15(3), 168–173.
13. King, J. L. (2006). Identity in the iSchool movement. Bulletin of the American Society for Information Science, 13-15.
14. Lerner, R. G. (1994). From documentation to information science: The beginnings and early development of the american documentation institute–american society for information science.i.24701