What Does Information Science Offer for Data Science Research?: A Review of Data and Information Ethics Literature

2 authors:

Brady D. Lund
University of North Texas
223 PUBLICATIONS **4,143** CITATIONS

SEE PROFILE

Ting Wang
Emporia State University
66 PUBLICATIONS **2,919** CITATIONS

SEE PROFILE

What Does Information Science Offer for Data Science Research?: A Review of Data and Information Ethics Literature

Brady Lund[†], Ting Wang

Emporia State University, 1 Kellogg Circle Emporia Kansas 66801-2523, United States

Abstract

This paper reviews literature pertaining to the development of data science as a discipline, current issues with data bias and ethics, and the role that the discipline of information science may play in addressing these concerns. Information science research and researchers have much to offer for data science, owing to their background as transdisciplinary scholars who apply human-centered and social-behavioral perspectives to issues within natural science disciplines. Information science researchers have already contributed to a humanistic approach to data ethics within the literature and an emphasis on data science within information schools all but ensures that this literature will continue to grow in coming decades. This review article serves as a reference for the history, current progress, and potential future directions of data ethics research within the corpus of information science literature.

Data as a subject and data science as a field of inquiry have received immense attention in the past handful of years, as big data becomes more ubiquitous and the applications of machine learning or artificial intelligence (AI) algorithms to process this data become more refined. Information science researchers working in Information Schools (iSchools) around the world have welcomed this development, integrating data-related curricula in their programs and encouraging data science research among faculty (Ortiz-Repiso, Greenberg, & Calzada-Prado, 2018; Song & Zhu, 2017; Wang & Lin, 2019). Meanwhile, the tremendous development of data storage capabilities due to the automation of measurement and data collection procedures, and the creation of highly sophisticated tools for analyzing and processing data, have given rise to considerable philosophical and legal concerns about the correct, legal, and appropriate approach to use data (Hand, 2018).

[†] Corresponding author: Brady Lund (Email: blund2@g.emporia.edu; ORCID: 0000-0002-4819-8162).

Review

On the one hand, information science is directly concerned with the storage and management phases of the data life cycle (Marchionini, 2016). On the other hand, one area of particular importance within the library and information science (LIS) is ethical practices: professional codes of ethics, anti-discrimination, and freedom from bias (Rubin, 2017). The human side of data science, the impact of how data is impacted by and can impact people, may then be seen as one of the areas where information science researchers and professionals can contribute expertise to the key discussions within the field of data science. Humanistic data science has already been seen in a few of the highly-notable works published by LIS researchers in recent years (Floridi, 2021; Hoffman, 2019; Noble, 2018), which may contribute to the future development of data science. However, there is a lack of understanding, at present, of how information science researchers have uniquely contributed to a discussion of data ethics. This paper will work to address these gaps by reviewing the development of the literature pertaining to the conception of data ethics and data bias and the movement towards a more humanistic data science within the information science literature.

The relevant literature for this review was retrieved within the *Library and Information Science Source* (Ebsco), *Computer Source* (Ebsco), and Google Scholar databases from July 2021 to April 2022. Each of the retrieved papers was reviewed by the authors to determine its relevance to the review and importance for supporting the complete narrative of the development of data ethics within the information science literature. This paper is designed as a narrative review (Paré et al., 2015), meaning that the articles selection aims to illustrate the key findings/concepts in relation to the central questions rather than necessarily provide a scoping review of *all* literature available.

Selected articles were chosen with the aim of addressing the following questions:

- *How do data bias and the evolution of data ethics appear in information science literature?*
- *What perspectives on data science bias and ethics within information science appear in the literature?*

This study covered the overlap of data science and information science, data bias and data ethics, and information bias and information ethics according to the findings of the collected literature to explore the possibility of the relatively young data science promoting its future development from the perspective of humanity science by learning from information science. For the purpose of this study, "information science" is used to mean the academic transdiscipline which studies all aspects of information (Borko, 1968; Taylor, 1963), whereas "library and information science" is used to refer both to the discipline of information science and professional training

and activities of librarianship (library science) that are informed by this discipline (Bates, 1999; Butler, 1951).

1   Overlap of Data Science with Information Science

The term "data" is used within the field of data science usually with the concept of electronic data or information, which has evolved over several centuries through the work of individuals like George Boole (1847), Alexander Graham Bell (1881), Ralph Hartley (1928), Alan Turing (1936), and Claude Shannon (1948), in the fields of mathematics, physics, electronics, communications, and computer science. In their work, data is the basic unit that makes electronic communications and devices like the computer possible, but there is also a basis for viewing data as a commodity that can be owned, traded, and used for prediction, and modeling, which is the basis of the modern boom in data science and engineering (Blum, Hopcroft, & Kannan, 2020; Dhar, 2013; Liang et al., 2018).

Data science emerged from a variety of disciplines over the course of half a century. Perhaps the most direct and prominent influence was from statistics, where papers dating well back into the 1960s discussed the process of collecting, storing, cleaning, and analyzing data (Ball & Hall, 1967; Horst, 1965; Tukey, 1962). Elements of what would become the field of data science are also evident within the history of bibliometrics. Eugene Garfield discusses, in his writings, bibliometrics as the use of citation *data* to "write the history of science" (Garfield, Sher, & Torpie, 1964). The natural sciences also played a significant role in the development of data science, by helping researchers understand the importance of data contexts throughout the lifecycle of data (Coombs, 1964; Spector, 1956). With the expansion of data collection and availability of data sets in the Internet age, data science has emerged as a distinct field of study (Cleveland, 2007).

Provost and Fawcett (2013) define data science as "a set of fundamental principles that support and guide the principled extraction of information and knowledge from data" with noting that, "at the moment, trying to define the boundaries of data science precisely is not of the foremost importance," as the definitions are constantly evolving as the field grows and refines itself (Provost & Fawcett, 2013). One element of the definition of data science that is clear is that it focuses broadly on changes in states of matter within the DIKW (data, information, knowledge, and wisdom) hierarchy (shown in Figure 1 below): transforming data to create information and disseminating findings to create new knowledge (Rowley, 2007). Authors have noted the considerable overlap between the modern discipline of data science and other disciplines, including informatics, statistics, computer science, and information science (Brennan, Chiang, & Ohno-Machado, 2018; Furner, 2015; Marchionini, 2016; Zhang & Benjamin, 2007).
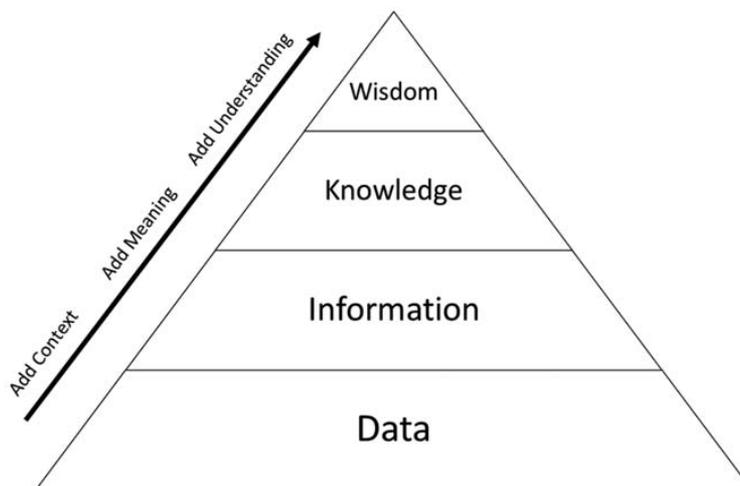
Review

Figure 1.   Depiction of DIKW Pyramid.

Among library and information science researchers, there was some early interest in data analytics in the period stretching from the 1940s to the 1960s (Batts, 1966; Janeway, 1944; Nicholson & Bartlett, 1962). However, more substantial interest appears to have emerged in the 1970s, with libraries at research universities serving as early repositories for research data, along with a growing interest within the discipline in bibliometrics (Conger, 1976; Pope, 1975; White, 1977). With the growth of information science elements within the discipline in the 1980s, terms like "data and information," "data processing," and "data analytics" become more ubiquitous in the literature (Cuadra, 1982; Griffiths & King, 1985; Wallace, 1985). In succeeding years, the work of bibliometricians carried the growth of the data science elements within information science up until the current era of enhanced interest in the topic among all factions of the discipline during the middle-to-late 2010s (Bar-Ilan, 2008; Egghe & Rousseau, 1990; Thelwall, 2004).

Data science has grown in interest at a rapid rate in recent years, including within the discipline of information science. Just in the past five years, the number of articles relating to data topics published in information science journals has grown by nearly eight-fold (Ma & Lund, 2020; 2021; Virkus & Garoufallou, 2019). Ongoing efforts are pushing for a great intwining of data science with information science, as evident in the work of authors like Wang (2018), Poole (2021), Washington Durr (2020), and Chohdary et al. (2021). While several studies have documented the increasing interest in this topic, few have attempted to parse out what, if anything, makes different disciplinary approaches towards data science unique.

2 Data Bias and the Evolution of Data Ethics

Bias in the analysis of data has long been a concern of analysts, though, until recent years, the focus was mainly on issues with the collection of data, the bias of the analysts themselves, and the imperfection of statistical models to reveal desired findings (Haussler, 1988; Kiviet, 1995; Mark, Eyssell, & Campbell, 1999). For instance, Dolly and Tillman (1974) provide a review of early instances of data bias owing to the improper design, collection, or analysis of data in educational research. Klein (1953), along with other researchers in the area of econometrics, also made early steps toward criticizing the quality of data and the impact of poor data on prediction and modeling. Many of these instances of bias were due to a lack of proper consideration of the attributes of a population, leading to a skewed sample.

Only in more recent years, with the emergence of increasingly large datasets and automated methods of analysis, have a range of pressing new concerns arise. A growing interest in artificial intelligence and machine learning, using data about the public to inform decision-making and model trends, has provoked this heightened interest (Koene, 2017; Ntoutsi et al., 2020). Algorithmic bias, flaws in the design of machine learning algorithms that are used to inform decision-making or the quality of the data that is fed to these algorithms, gained traction as a topic of substantial research interest in the mid-2010s, with the work of researchers like Bozdag (2013), Hajian et al. (2016), and Noble (2018). Similarly, the implementations of large-scale datasets to drive artificial intelligence and smart technologies can be deeply flawed when the person interacting with the AI technology fits outside the "norm" or training set utilized (Whittaker et al., 2019).

There are many types of data bias that have been discussed in the literature in recent years. Here, five of these types are explored: sampling bias, availability bias, methodological bias, contextual bias, and interpretation bias. Each of these types of bias may be known by slightly different names depending on the source (e.g. sampling bias may be called "selection bias" by some authors). In order to understand when and where each type of bias occurs, it is useful to reference a model of the data life cycle. The life cycle presented in Figure 2 is a simple model adapted from Stobierski (2021) at the Harvard Business School.

The first phase of the data life cycle is "generation," or the process of either creating or determining the qualities and quantities of data that will be collected (Ray, 2014). Data is then collected and processed (refined and organized for storage), and stored within a database, using some management processes to ensure *contextual integrity* (Nissenbaum, 2004). Next, the data is retrieved for analysis (producing some insight from the data—aka information), visualization (depictions of significant findings), and interpretation (the significance/meaning of the findings). The life

Review

cycle of data is continuous, as the data may be reused, archived, or partially or fully deleted, as is depicted as a distinct step in some other models of the data life cycle (Rüegg et al., 2014).
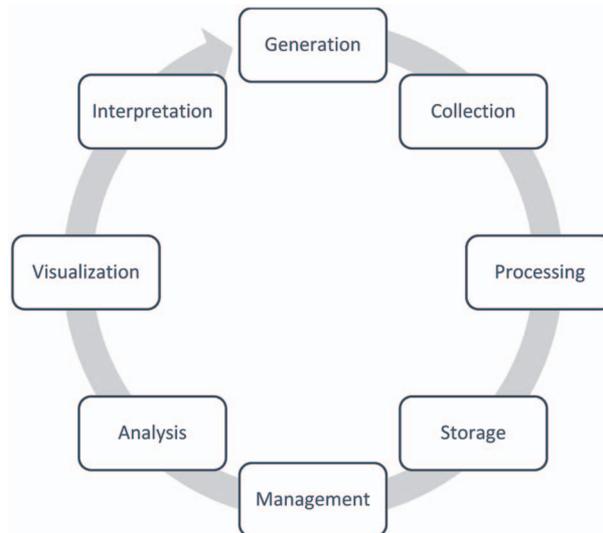

Figure 2.   Data Life Cycle Model.

Sampling Bias occurs during the planning and/or collection phase of the data lifecycle (Hahn et al., 2005). Sampling bias occurs when the parameters used to select a sample are improperly designed or executed (Smith & Noble, 2014). This bias can occur due to ethical, logical, and practical faults. Ethical faults occur when the sample collection is skewed due to some limitation owed to ethical barriers, such as a risk to certain key members of a group that prevents participation (e.g. certain victims of domestic violence) (Dattalo, 2010). Logical faults occur on the part of the data collectors when they do not select a truly representative sample (Martinez-Mesa et al., 2016). This has historically been a major, and highly visible, issue with political polling (Lynn & Jowell, 1996). An example of a practical fault is found in the work of Severson and Ary (1983), who reported sampling bias due to consent procedures with minors. Certain parents were simply unwilling to allow their child to participate in a study, and this prevented a representative sample from being collected.

Availability Bias may be classified as a type of sampling bias, which is particularly discriminatory against heterogeneous populations. It occurs when samples are produced and interpretations/decisions are made on the basis of those subjects/data that are most readily available and abundant (Dube-Rioux & Russo, 1988). The impact of availability bias is variable upon the purpose for which the data was

collected. In medical facilities, availability may lead to the misdiagnosis of a rare medical condition as one that is more common (Mamede et al., 2010). In economics, it may lead to an investor selecting a stock that is widely known or popular over one that is likely to produce the best return (Javed, Bagh, & Razzaq, 2017). In library and information science, it may result in the accumulation of library materials from a few popular, local, or cheap sources over those that offer the most diverse and intellectually challenging materials (Harmeyer, 1995).

Statistical Bias is the product of the statistical analyses selected and their execution. There is, of course, a bias that is inherent to statistical tests (Cole, 1981; Millsap & Everson, 1993). There is also bias owing to the sample characteristics of the data relative to tests selected—skewed data, outliers—that can lead to incomplete or improper analyses and interpretation (Chen et al., 2001; Kinksman, 1957; Westgard & Hunt, 1973).

Context Bias refers to bias that occurs when the context from which data was collected or analyzed is lost during the storage, analysis, or interpretation stages of the life cycle. In some cases, contextual bias can actually be an un-bias of sorts, such as in the forensic sciences where knowing information about how something originated could negatively influence interpretations of events (Osborne et al., 2014). However, in most cases, removing context from a situation is very problematic. It leads to interpretations of a situation—such as in the case of medical diagnosis or investment strategy—that may be accurate in some cases but vastly inaccurate it others (Price & Nicholson, 2019).

Interpretation Bias occurs when the interpretation or ways in which a story is told about data findings is itself biased (Wilholt, 2009). Interpretation of data can be influenced by one's past experiences and cultural biases, such that a person from one background may interpret the findings of analysis very differently from those from another background (Lund, 2022). All research publications rely on interpretations of some kind of data (Gummesson, 2003; Šimundić, 2013). Though some have noted strategies for mitigating this type of bias, such as using a blinded interpretation of findings by individuals not directly involved with the data collection and management (Järvinen et al., 2014), these are still not perfect defenses against the effect of bias.

As recognition of data bias, algorithmic bias, and AI malpractice grew over the past several years, so too did the promotion of new ethical standards and education for data scientists. Mittelstadt and Floridi (2015) and Barocas and Boyd (2017) survey current ethical beliefs and practices of data scientists and identify gaps that may exacerbate concerns with the next generation of machine learning technologies. Several authors have proposed some model or overarching philosophy for the

responsible use of big data analytics (Fairfield & Shtein, 2014; Richards & King, 2014; Zwitter, 2014).

One of the distinguishing elements of information science research from systems or computer science research is its emphasis on the *ethical* use of information/data, rather than primarily emphasizing only how best to streamline its use. Information science has a long history of debating the ethical underpinnings of what has evolved into the fields of data science/mining and artificial intelligence (Adam, 1991; Carlin, 2003; Kostrewski & Oppenheim, 1979; Shiri, 2016). This approach is one major benefit of information science's interdisciplinary nature and social science bent. The discipline's foundations within the profession of librarianship, and its codes of ethics, strongly influence the current philosophy toward data topics (Roeschley & Khader, 2020; Shachaf, 2005).

3　Perspectives on Data Science Bias and Ethics within Information Science

This section provides an overview of recent work within the literature of information science that suggests an interest in developing ethics for data science. Interest in this area of study is still growing (Ajibade & Mutula, 2020; Ma & Lund, 2021), but already a sizeable collection of research and essays exists on this topic, dating back over the past decade.

Shilton (2012) discusses the ethics of participatory personal data (PPD) from an information science perspective. PPD is "a new class of data about people and their environments, generated by an emerging network of embedded devices and specialized software" (Shilton, 2012). It includes data captured by users on personal devices, like geotagging on photos, GPS tracking on a smartwatch or other exercise device, and "smart" sensors and devices. This data clearly presents many significant ethical challenges within the purview of information science researchers and professionals:

- Surveillance concerns (who has access to this data, which could be used to reveal confidential insights about the individual),
- Storage of data (ensuring that it is not compromised), the use of this data for research purposes,
- Access to these devices and usage of personal data by the users themselves.

Andres (2016) analyzes the ethics of competitive intelligence (CI) through the lens of information science. These fields, traditionally, take starkly different tactics in approaching how data/information should be used. While CI is perhaps not as ruthless as "competitive," the name seems to imply the beliefs about how and why

data may be used differ from the standards that library and information professionals apply in utilizing and sharing information in their work. One major distinction stem from the data collection methods, for instance. Library and information professionals are taught to consider awareness and consent vital in data collection (Cooke, 2018; Rubel, 2014). These elements are important in CI as well, but definitions of "awareness" and "consent" may differ. While many entities that collect and utilize users' data today employ those lengthy consent forms that research shows that most users will never fully read or understand (Cassileth et al., 1980; Pederson et al., 2011), library and information professionals might opt for a more succinct and clear description of how the data will be collected/used (in addition to the lengthy legal form). People can often be a burden for the data analyst—but not for information professionals, who place the human being at the center of their work.

Marchionini (2016) argues that information science is one of four foundational pillars of data science (along with statistics, computer science, and knowledge domains) and can be seen as lending theory (which extends to ethical practices) to the emerging field. Information science is particularly important in terms of data curation, which is influenced by principles of knowledge organization, modeling, and storytelling that have been the expertise of library and information professionals for over a century. Information scientists, according to Marchioni, are unique in their focus on information throughout its lifecycle, including "ethical and legal conditions associated with data collection," "appraisal, data quality, and cleaning," "metadata assignment," "storage and preservation of data," and "evaluation of conclusions based on data exploration and analysis and making the data and workflows findable and reusable" (Marchionini, 2016).

Several articles published in recent years have begun to push information science towards a more defined role within data science. Hoffmann et al.'s (2018) panel discussion at the ASIST annual meeting ignited elevated exchange on the topic, which had already been brewing somewhat within the discourse of the field of IS. Each of the presenters of this panel hasengaged in scholarship that accelerates debate on topics like fairness and discrimination with AI (Hoffmann, 2019; Roberts, 2016; Wood, 2017).

One significant contributor to these discussions about data ethics from an information science perspective is Luciano Floridi. Floridi's work (e.g. Floridi, 2014; Mittelstadt et al., 2016) has received great recognition and acclaim among information science researchers, including Fyffe (2015), Wang (2018), and Bawden and Robinson (2020). Among the many important ideas lent by Floridi to information science is an ethical philosophy of data and artificial intelligence. One concern with AI is not that it poses a legitimate threat to human livelihood but that its benefits for human quality of life will be disproportionately felt by certain populations and

exacerbate disparities. These disparities can already be seen in the national investment in AI if one looks at the investment of the United States and China relative to most nations in the Global South, which have limited funds to invest in these endeavors.

Within library and information science, it is understood that information should be available to all. Ranganathan's (1931) second law of library science states, "Every Reader His/Her Book" and "Books Are For Use"—access is built into the foundation of the discipline of information science. However, access is traditional and not a guiding principle in data analytics. Data is highly valuable and is generally held in such a way, along with the technology and intellectual capital necessary to store and analyze the data (Banterle, 2019; Ballantyne, 2020). This exacerbates the class divide and further eliminates the potential for economic mobility (Rolfe, 2017; Schradie, 2020). An information science-inspired effort to expand access to the tools and technologies necessary to compete in a data-driven economy is intriguing, to say the least. But how could such an endeavor be accomplished? Already the information access mission of libraries largely fails when the picture is expanded to include countries in the Global South, where library resources and technology are far lesser than in the Global North (Lund et al., 2021; Pather & Gomez, 2010).

Roeschley and Khader (2020) provide the most extensive review of data ethics to date in their iConference paper. In their study, the researchers reviewed data science articles published in library and information science journals, identifying four themes relating to data ethics issues addressed in the literature: privacy, research ethics, ethical ecosystems, and control. The researchers note that the first theme, privacy, dates back to the earliest values of the field, set forth in the founding documents of organizations like the American Library Association and the Association for Information Science and Technology (as noted previously in this review). Research ethics, similarly, has a long history within information science, with academic librarians' long-held role in the lifecycle of scholarly research. The ethical handling and role in the lifecycle of scholarly research. The ethical handling and reporting of research data is anticipated or presently considered by many information science researchers to be a central role of the modern academic library (Cox & Pinfield, 2014; Laskowski, 2021; Pinfield, Cox, & Smith, 2014). According to Roeschley and Khader (2020), ethical ecosystems are "environments in which one would expect to see a common standard for ethical behavior." The researchers note that these environments may include business, academic, and government environments. Certainly, this role aligns with the traditional roles of libraries as complex information ecosystems, guided by a shared set of ethical standards (Walter, 2008). Lastly, control relates to maintaining the data security and understanding the legal limitations pertaining to the data (Roeschley & Khader, 2020). The term "control" conjures to mind the subject of knowledge organization, bibliographic

control, and the management of metadata. Indeed, the two roles (cataloger and data manager) can be seen as quite similar. The work of Roeschley and Khader (2020) provides a foundation for future study, which may explore more deeply the unique contributions of information science research and researchers to anti-bias and ethics literature in data science.

4 Discussion

Growth in data science research and data professionals over the coming decades is a near-certainty. The value of data to both the profit-generating and non-profit sectors of the economy are undeniable and will assuredly drive investments within higher education and among grant-funding organizations (Berman et al., 2018; Cao, 2017; Waller & Fawcett, 2013). It is not surprising that information schools would be eager to grab a piece of this field, as data science offers tremendous potential for the kind of attention and growth for which these schools are often overlooked. As this paper details, information science has much to offer to data science and, indeed, has already had an impact on the development of this growing field.

4.1 How Has Information Science Contributed to the Development of Data Science?

Historically, information science has been directly involved in the data life cycle's storage and management phases. Management of data repositories has been a growing role within academic library services over the past several decades (Antell et al., 2014; Brase & Farquhar, 2011; Read, 2008; Shen, 2015; Si et al., 2013). These roles remain important in the profession of librarianship as well as the information science research as particularly evident in the work of Carol Tenopir, among others (Tenopir et al., 2014; Tenopir et al., 2015; Tenopir et al., 2017; Tenopir et al., 2020). In more recent years, with a greater emphasis on the topics of health informatics and data-driven decision-making for health outcomes, data has taken on a greater role within the discipline (Bath et al., 2005; Chen et al., 2012; Corbett, Deardorff, & Kovar-Gough, 2014; Kerr, Norris, & Stockdale, 2008; Tremblay, 2016). Growth in the potential role of data, new avenues for AI technology, increased research funding, and the impacts of major crises like the COVID-19 pandemic led to a broader role of this topic area within the discipline (Semeler et al., 2017; Shankar et al., 2020; Siguenza-Guzman et al., 2015; Song & Zhu, 2017; van der Aalst et al., 2017). Research data management, health informatics, and data science implications for information organizations are thus three major areas where information science researchers have contributed to the literature on data science.

Certainly, information science also maintains an intimate connection with data science within their adjacent places in the DIKW pyramid. The transdisciplinary

work of Marcia Bates illustrates that the concepts of "data" and "information" cannot be truly divorced from one another. As Bates (2005) notes, data is information that can be sensed but not yet consciously understood or has not yet been rationally organized (one might react to a pinprick, but it is entirely different to understand the significance or reason for that pain). Data may be handled in much the same way as information. We see that the contemporary models of the information life cycle and information management align well with those of data life cycle and data management; indeed, researchers in these areas frequently cite one another's work (Christopherson et al., 2020; Cox & Tan, 2018). Additionally, there is an emerging interest among information literacy experts (e.g. academic librarians) to provide instruction in the related area of data literacy (Koltay, 2017; Prado, 2013; Špiranec, Kos, & George, 2019).

Bates (2015) conceives a spectrum of the information disciplines, as they span the traditional domains of arts and humanities, social and behavioral sciences, and natural science and mathematics. The author situates data science-related sub-disciplines—data mining, scientometrics, data storage, informatics—firmly within the natural sciences and mathematics end of the spectrum, on the opposite end from library science and museum studies, which lie more firmly within the arts and humanities. However, the concept of the information disciplines as a spectrum illustrates how data science may be viewed among information science researchers. There are no defined borders within a spectrum; there can be bleed-over. This differentiates the education and perspectives of the information scientist from the pure natural scientists or mathematicians. Aspects of information science like information policy, semiotics, and information behavior can inform our perspectives on data science, leading to a more humanistic perspective on the field (Aragon et al., 2016; Kuiler & McNeely, 2020).

4.2   What Contribution Has Information Science Provided to Our Understanding of Data Ethics?

One place to start when answering this question is with Safiya Noble's (2018) *Algorithms of Oppression*, a book that has been cited over 3000 times by authors from across virtually all disciplines, thus making the scope of its impact very clear. The book highlights how the design of algorithms—all phases of the data and information life cycles—is impacted by the implicit (and explicit) biases held by analysts and developers. Of course, Noble's work did not emerge independent of any prior work within academia. It was informed by "critical information studies and critical race and gender studies" (Noble, 2018). This includes work by information science faculty like Bar-Ilan (2007a; 2007b), Furner (2007), Saracevic (1999), Smith (1981), Spink et al. (2001), and Zimmer (2008). Fairness in information retrieval is

the central focus of Noble's book, a topic discussed by many other information science researchers beyond those previously cited, including Shah (Gao & Shah, 2020a, 2020b) and Jansen (Eastman & Jansen, 2003; Jansen & Schuster, 2011).

Contextualization of data is a key theme in the most impactful data science research published by information science journals. Context is an important component of how information science researchers understand the world (Madden, 2000), whether it is the life cycle of information/data or understanding the threat posed by global inequities. Wang (2018) appears to come to a similar conclusion—albeit not in the same words—when they suggest that the greatest opportunities for information science to contribute to data science are through a conception of data science research that is not purely positivistic and by which the quality of data, not merely the quantity, are questioned. Many of the human-centric theories that form the basis of much information science research today—such as Belkin's (1980) "Anomolous States of Knowledge," which describes how limitations of the human ability influence interaction—have direct applicability to how we might attain a more humanistic data science.

5 Conclusion

Data science is a rapidly growing field that represents an industry that is expected to reach a value of over $100 billion in the United States by the end of the 2020s (Acute Market Reports, 2020). That valuation alone would be enough to encourage researchers and educators in any discipline to expand their reach into data science. However, information science researchers and library and information science education programs are particularly well-situated to work within this field of data science. With a background in the life cycle of information, management strategies, and provision of information, the educational background of information science researchers translates well to handling data throughout its life cycle (Semeler & Pinto, 2020). Information science's transdisciplinary nature prepares researchers well for applying humanistic perspectives to combat data bias and impose an ethical regimen on how data is collected, stored, and used.

This review illustrates that information science researchers have already invested efforts into staking a claim within data science research, not merely by expanding faculties to hire more data science researchers from the natural science disciplines, but by educating and performing research in their own right that addresses key issues within this expanding field. Through continued investment and support, information schools have a considerable role in how data science evolves and what universities continue to thrive during a period of elevated financial scrutiny of educational institutions.

Author contributions

Brady Lund (blund2@g.emporia.edu): Writing—Original Draft; Ting Wang (twang2@g.emporia.edu): Writing—Review & Editing.

References

Acute Market Reports. (2020). *Global big data analytics market size, market share, application analysis, regional outlook, growth trends, key players, competitive strategies and forecasts, 2019 to 2027.* Retrieved from https://www.researchandmarkets.com/reports/4992328/

Adam, R. (1991). Laws for the lawless: Ethics in (information) science. *Journal of Information Science*, 17, 357–372.

Ajibade, P., & Mutula, S.M. (2020). Big data research outputs in the library and information science. *African Journal of Libraries, Archives, and Information Science*, 30(1), 49–60.

Andres, M.C. (2016). Ethical differences: A literature review of the ethics of competitive intelligence for the LIS professional. *Journal of Library and Information Sciences*, 4(1), 1–15.

Antell, K., Bales Foote, J., Turner, J., & Shults, B. (2014). Dealing with data: Science librarians' participation in data management at Association of Research Libraries institutions. *College and Research Libraries*, 75(4), 557–574.

Aragon, C., Hutto, C., Echenique, A., Fiore-Gartland, B., Huang, Y., Kim, J., …, & Bayer, J. (2016). Developing a research agenda for human-centered data science. *ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*, 19, 529–535.

Ball, G.H., & Hall, D.J. (1967). A clustering technique for summarizing multivariate data. *Behavioral Science*, 12(2), 153–155.

Ballantyne, A. (2020). How should we think about clinical data ownership? *Journal of Medical Ethics*, 46, 289–294.

Banterle, F. (2019). Data ownership in the data economy: A European dilemma. In Synodinou, T.E., Jougleux, P., Markou, C., & Prastitou, T. (eds.), *EU Internet law in the digital era*. New York, NY: Springer. https://doi.org/10.1007/978-3-030-25579-4_9

Bar-Ilan, J. (2007a). Google bombing from a time perspective. *Journal of Computer-Mediated Communication*, 12(3), 910–938. https://doi.org/10.1111/j.1083-6101.2007.00356.x

Bar-Ilan, J. (2007b). Manipulating search engine algorithms: The case of Google. *Journal of Information, Communication and Ethics in Society*, 5(2/3), 155–166.

Bar-Ilan, J. (2008). Informetrics at the beginning of the 21st century: A review. *Journal of Informetrics*, 2(1), 1–52.

Barocas, S., & Boyd, B. (2017). Computing ethics: Engaging the ethics of data science in practice. *Communications of the ACM*, 60(11), 23–25.

Bates, M.J. (1999). The invisible substrate of information science. *Journal of the American Society for Information Science*, 50(12), 1043–1050. https://doi.org/10.1002/(SICI)1097-4571(1999)50:12%3C1043::AID-ASI1%3E3.0.CO;2-X

Bates, M.J. (2005). Information and knowledge: An evolutionary framework for information science. *Information Research*, 10(4), paper 239.

Bates, M.J. (2015). The information professions: Knowledge, memory, heritage. *Information Research*, 20(1), paper 655. http://InformationR.net/ir/20-1/paper655.html

Bath, P.A., Craigs, C., Maheswaran, R., Raymond, J., & Willett, P. (2005). Use of graph thoery to identify patterns of deprivation and high morbidity and mortality in public health data sets. *Journal of the American Medical Informatics Association*, 12(6), 630–641.

Batts, N.C. (1966). Data analysis of science monograph order/cataloging forms. *Special Libraries*, 57, 583–586.

Bawden, D., & Robinson, L. (2020). "The dearest of our possessions": Applying Floridi's information privacy concept in models of information behavior and information literacy. *Journal of the Association for Information Science and Technology*, 71(9), 1030–1043.

Belkin, N.J. (1980). Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information and Library Science*, 5, 133–143.

Bell, A.G. (1881). The production of sound by radiant energy. *Science*, 48, 242–253.

Berman, F., Rutenbar, R., Hailpern, B., Christensen, H. Davidson, S., Estrin, D., …, Szalay, A.S. (2018). Realizing the potential of data science. *Communications of the ACM*, 61(4), 67–72.

Blum, A., Hopcroft, J., & Kannan, R. (2020). *Foundations of data sccience*. Cambridge, UK: University Printing House.

Boole, G. (1847). *The mathematical analysis of logic*. Cambridge, MA: MacMillan, Barclay, and MacMillan.

Borko, H. (1968). Information science: What is it? *American Documentation*, 19(1), 3–5.

Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology*, 15(3), 209–227.

Brase, J., & Farquhar, A. (2011). Access to research data. *D-Lib Magazine*, 17(1/2). https://doi.org/10.1045/january2011-brase

Brennan, P.F., Chiang, M.F., & Ohno-Machado, L. (2018). Biomedical informatics and data science: Evolving fields with significant overlap. *Journal of the American Medical Informatics Association*, 25(1), 2–3.

Butler, P. (1951). Librarianship as a profession. *The Library Quarterly*, 21(4), 235–247. https://doi.org/10.1086/617815

Cao, L. (2017). Data science: Challenges and directions. *Communications of the ACM*, 60(8), 59–68.

Carlin, A.P. (2003). Disciplinary debates and bases of interdisciplinary studies: The place of research ethics in library and information science. *Library and Information Science Research*, 25, 3–18.

Cassileth, B.R., Zupkis, R.V., Sutton-Smith, K., & March, V. (1980). Informed consent: Why are its goals imperfectly realized? *New England Journal of Medicine*, 302, 896–900.

Chen, C., Haddad, D., Selsky, J., Hoffman, J.E., Kravitz, R.L., Estrin, D.E., & Sim, I. (2012). Making sense of mobile health data: An open architecture to improve individual- and population-level health. *Journal of Medical Internet Research*, 14(4), pe112-e112.

Chen, F., Bollen, K.A., Paxton, P., Curran, P.J., & Kirby, J.B. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods and Research*, 29(4), 468–508.

Chohdary, N.I., Asghar, M.B., & Al Shaheer, M.A. (2021). Predicting LIS scholarly research directions in the era of data science. *Library Philosophy and Practice*, article 6328. https://digitalcommons.unl.edu/libphilprac/6328

Christopherson, L., Scott, E., Mandal, A., & Baldin, I. (2020). Toward a data lifecycle model for NSF large facilities. *Proceedings of Practice and Experience in Advanced Research Computing Conference, 2020*. Retrieved from https://doi.org/10.1145/3311790.3396636

Cleveland, W.S. (2007). Data science: An action plan for expanding the technical areas of the field of statistics. *International Statistical Review*, 69(1), 21–26.

Cole., N.S. (1981). Bias in testing. *American Psychologist*, 36(10), 1067–1077.

Coleman, C.N. (2020). Managing bias when library collections become data. *International Journal of Librarianship*, 5(1), 8–19.

Conger, L.D. (1976). Data reference work with machine readable data files in the social sciences. *Journal of Academic Librarianship*, 2, 60–65.

Coombs, C.H. (1964). *A theory of data*. New York, NY: John Wiley and Sons.

Cooke, L. (2018). Privacy, libraries and the era of big data. *IFLA Journal*, 44(3), 167–169.

Corbett, M., Deardorff, A., & Kovar-Gough, I. (2014). Emerging data management roles for health librarians in electronic medical records. *Journal of the Canadian Health Libraries Association*, 35, 55–59.

Cox, A., & Pinfield, S. (2014). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*, 46(4), 299–316.

Cox,, A., & Tam, W. (2018). A critical analysis of lifecycle models of the research process and research data management. *Aslib Journal of Information Management*, 70(2), 142–157.

Cuadra, C.A. (1982). *A library and information science research agenda for the 1980s: Final Report*. Santa Monica, CA: Cuadra Associates.

Dattalo, P. (2010). Ethical dilemmas in sampling. *Journal of Social Work Values and Ethics*, 7(1), 12–23.

Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, 56(12), 64–73.

Dolly, J.P., & Tillman, M.H. (1974). An overview of research artifacts supposedly causing data bias. *Proceedings of the Annual Meeting of the American Educational Research Association, 1974*. Retrieved from https://files.eric.ed.gov/fulltext/ED101009.pdf

Dube-Rioux, L., & Russo, J.E. (1988). An availability bias in professional judgment. *Journal of Behavioral Decision Making*, 1(4), 223–237.

Eastman, C.M., & Jansen, B.J. (2003). Coverage, relevance, and ranking: The impact of query operators on web search engine results. *ACM Transactions on Information Systems*, 21(4), 383–411.

Egghe, L., & Rousseau, R. (1990). *Introduction to informetrics: Quantitative methods in library, documentation, and information science*. Amsterdam, NL: Elsevier.

Fairfield, J., & Shtein, H. (2014). Big data, big problems: Emerging issues in the ethics of data science in journalism. *Journal of Mass Media Ethics*, 29(1), 38–51.

Floridi, L. (2014). Open data, data protection, and group privacy. *Philosophy and Technology*, 27, 1–3.

Floridi, L. (2021). *Ethics, governance, and policies in artificial intelligence*. London, UK: Springer.

Furner, J. (2007). Dewey deracialized: A critical race-theoretic perspective. *Knowledge Organization*, 34(3), 144–168.

Furner, J. (2015). Information science is neither. *Library Trends*, 63(3), 362–377.

Fyffe, R. (2015). The value of information: Normativity, epistemology, and library and information science in Luciano Floridi. *Portal: Libraries and the Academy*, 15(2), 267–286.

Gao, R., & Shah, C. (2020a). Toward creating a fairer ranking in search enginer results. *Information Processing and Management*, 57(1), article 102138.

Gao, R., & Shah, C. (2020b). Counteracting bias and increasing fairness in search and recommender systems. *ACM Conference on Recommender Systems*, 14, 745–747.

Garfield, E., Sher, I.H., & Torpie, R.J. (1964). *The use of citation data in writing the history of science*. Philadelphia, PA: Institute for Scientific Information.

Griffiths, J., & King, D.W. (1985). *New directions in library and information science education*. Rockville, MD: King Research Inc.

Gummesson, E. (2003). All research is interpretive! *Journal of Business and Industrial Marketing*, 18(6/7), 482–492.

Hahn, S., Puffer, S., Torgerson, D.J., & Watson, J. (2005). Methodological bias in cluster randomized trials. *BMC Medical Research Methodology*, 5, article 10.

Hajian, S., Bonchi, F., & Castillo, C. (2016). Algorithmic bias: From discrimination discovery to fairness-aware data mining. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 22, 2125–2126.

Hand, D.J. (2018). Aspects of data ethics in a changing world: Where are we now?. *Big data*, 6(3), 176–190.

Harmeyer, D. (1995). Potential collection development bias: Some evidence on a controversial topic in California. *College and Research Libraries*, 56, 101–111.

Hartley, R.V. (1928). Transmission of information. *Bell System Technical Journal, July 1928*, 535.

Haussler, D. (1988). Quantifying inductive bias: AI learning algorithms and Valiant's learning framework. *Artificial Intelligence*, 36(2), 177–221.

Hoffmann, A.L. (2019). Where fairness fails: On data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication, and Society*, 22(7), 900–915.

Hoffmann, A.L., Wolf, C.T., Roberts, S.T., & Wood, S. (2018). Beyond fairness, accountability, and transparency in the ethics of algorithms: Contributions and perspectives from LIS. *Proceedings of the ASIST Annual Meeting*, 81, 695–696.

Horst, P. (1965). *Factor analysis of data matrices*. New York, NY: Holt, Rinehart, and Winston.

Jackson, B. (2018). The changing research data landscape and the experiences of ethics review board chairs. *Journal of Academic Librarianship*, 44(5), 603–612.

Janeway, R.C. (1944). *Technique of measuring student use of the university library through circulation records* (doctoral dissertation). Urbana, IL: University of Illinois.

Jansen, B.J., & Schuster, S. (2011). Bidding on the buying funnel for sponsored search and keyword advertising. *Journal of Electronic Commerce Research*, 12(1), 1–18.

Järvinen, T.L.N., Sihvonen, R., Bhandari, M., Sprague, S., Malmivaara, A., Paavola, M., Schünemann, H.J., & Guyatt, G.H. (2014). Blinded interpretation of study results can feasibly and effectively diminish interpretation bias. *Journal of Clinical Epidemiology*, 67(7), 769–772.

Javed, H., Bagh, T., & Razzaq, S. (2017). Herding effects, over confidence, availability bias and representativeness as behavioral determinants of perceived investment performance: An empirical evidence from Pakistan Stock Exchange. *Journal of Global Economics*, 6(1), 1–13.

Jones, K.M.L. (2019). "Just because you can doesn't mean you should": Practitioner perceptions of learning analytics ethics. *Portal: Libraries and the Academy*, 19(3), 407–428.

Review

Jones, K.M.L., Asher, A., Goben, A., Perry, M.R., Salo, D., Briney, K.A., & Robertshaw, M.B. (2020). "We're being tracked at all times": Student perspectives of their privacy in relation to learning analytics in higher education. *Journal of the Association for Information Science and Technology*, 71(9), 1044–1059.

Jones, K.M.L., & Salo, D. (2018). Learning analytics and the academic library: Professional ethics commitments at a crossroads. *College and Research Libraries*, 79(3), 304–323.

Kerr, K.A., Norris, T., & Stockdale, R. (2008). The strategic management of data quality in healthcare. *Health Informatics Journal*, 14(4), 259–266.

Kilgour, F.G. (1969). The economic goal of library automation. *College & Research Libraries*, 30(4), 307–311.

Kinksman, B. (1957). Proper and improper use of statistics in geophysics. *Tellus*, 9(3), 408–418.

Kiviet, J.F. (1995). On bias, inconsistency, and efficacy of various estimators in dynamic panel data models. *Journal of Econometrics*, 68(1), 53–78.

Klein, L.R. (1953). *A textbook of econometrics*. White Plains, NY: Row, Peterson, and Company.

Koene, A. (2017) Algorithmic bias: Addressing growing concerns. *IEEE Technology and Society Magazine, June 2017*, 31–32.

Koltay, T. (2017). Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science*, 49(1), 3–14. https://doi.org/10.1177/0961000615616450

Kostrewski, B.J., & Oppenheim, C. (1979). Ethics in information science. *Journal of Information Science*, 1(5), 277–283.

Kuiler, E.W., & McNeely, C.L. (2020). Knowledge formulation in the health domain: A semiotics-powered approach to data analytics and democratization. In Batarseh, F.A., & Yang, R., *Data democracy: At the nexus of artificial intelligence, software development, and knowledge engineering*. Cambridge, MA: Academic Press.

Laskowski, C. (2021). Structuring better services for unstructured data: Academic libraries are key to an ethical research data future with big data. *Journal of Academic Librarianship*, 47(4), 102335.

Liang, F., Yu, W., An, D., Yang, Q., Fu, X., & Zhao, W. (2018). A survey on big data market: Pricing, trading, and protection. *IEEE Access*, 6, 15132–15154.

Lund, B.D. (2022). The Art of (Data) Storytelling: Hip Hop Innovation and Bringing a Social Justice Mindset to Data Science and Visualization. *The International Journal of Information, Diversity, & Inclusion (IJIDI)*, 6(1/2), 31–41.

Lund, B.D., Wang, T., Shamsi, A., Abdullahi, J., Awojobi, E.A., Borgohain, D.J., …, & Yusuf, A.O. (2021). Barriers to scholarly publishing among library and information science researchers: International perspectives. *Information Development*. https://doi.org/10.1177/02666669211052522

Lynn, P., & Jowell, R. (1996). How might opinion polls be improved?: The case for probability sampling. *Journal of the Royal Statistical Society: Series A*, 159(1), 21–28.

Ma, J., & Lund, B.D. (2020). The evolution of LIS research topics and methods from 2006 to 2018: A content analysis. *Proceedings of the Association for Information Science and Technology*, 57(1), e241.

Ma, J., & Lund, B.D. (2021). The evolution and shift of research topics and methods in library and information science. *Journal of the Association for Information Science and Technology*, 72(8), 1059–1074.

Mamede, S., van Gog, T., van den Berge, K., Rikers, R.M.J., van Saase, J.L., van Guldener, C., & Schmidt, H.G. (2010). Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *Journal of the American Medical Association*, 304(11), 1198–1203.

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

Mark, M.M., Eyssell, K.M., & Campbell, B. (1999). The ethics of data collection and analysis. *New Directions for Evaluation*, 82, 47–56.

Martinez-Mesa, J., Gonzalez-Chica, D.A., Duquia, R.P., Bonamigo, R.R., & Bastos, J.L. (2016). Sampling: How to select participants in my research study? *Anais Brasileiros de Dermatologia*, 91(3), 326–330.

Millsap, R.E., & Everson, H.T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement*, 17(4), 297–334.

Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data and Society, July-December 2016*, 1–21.

Mittelstadt, B.D., & Floridi, L. (2015). The ethics of big data: Current and foreseeable issues in biomedical contexts. *Science and Engineering Ethics*, 22, 303–341.

Nicholson, N.N., & Bartlett, E. (1962). Who uses university libraries? *College and Research Libraries*, 23(3), 217–259.

Nissenbaum, H. (2004). Privacy as conxtual integrity. *Washington Law Review*, 79, 119–158.

Noble, S.U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York, NY: New York University Press.

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M., …, & Staab, S. (2020). Bias in data-driven artificial intelligence systems: An introductory survey. *WIREs Data Mining and Knowledge Discovery*, 10, paper e1356. https://doi.org/10.1002/widm.1356

Ortiz-Repiso, V., Greenberg, J., & Calzada-Prado, J. (2018). A cross-institutional analysis of data-related curricula in information science programs. *Journal of Information Science*, 44(6), 768–784.

Osborne, N.K.P., Woods, S., Kieser, J., & Zajac, R. (2014). Does contextual information bias bitemark comparisons? *Science and Justice*, 54(4), 267–273.

Paré, G., Trudel, M.C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information and Management*, 52(2), 183–199.

Pather, S., & Gomez, R. (2010). Public access ICT: A south-south comparative analysis of libraries, telecentres and cybercafes in South Africa and Brazil. *AMCIS 2010 Proceedings*, article 526. https://aisel.aisnet.org/amcis2010/526

Pederson, E.R., Neighbors, C., Tidwell, J., & Lostutter, T.W. (2011). Do undergraduate student research participants read psychological consent forms? *Ethics and Behavior*, 21(4), 332–350.

Pinfield, S., Cox, A.M., & Smith, J. (2014). Research data management and libraries: Relationships, activities, drivers and influencers. *PLoS One*, 9(12), e114734.

Poole, A.H. (2021). Leading the way: A new model for data science education. *Proceedings of the Association for Information Science and Technology*, 58(1), 525–531.

Pope, A. (1975). Bradford's law and the periodical literature of information science. *Journal of the American Society for Information Science*, 26, 207–213.

Review

Prado, J.C., & Marzal, M.A. (2013). Incorporting data literacy into information literacy programs: Core competencies and contents. *Libri*, 63(2), 123–134. https://doi.org/10.1515/libri-2013-0010

Price, W., & Nicholson, I. (2019). Medical AI and contextual bias. *Harvard Journal of Law and Technology*, 33, 65–116.

Prindle, S., & Loos, A. (2017). Information ethics and academic libraires: Data privacy in the era of big data. *Journal of Information Ethics*, 26(2), 22–33.

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, 1(1), 51–59.

Ranganathan, S.R. (1931). *The five laws of library science*. London, UK: Edward Goldston, LTD.

Ray, J.M. (ed.) (2014). *Research data management: Practical strategies for information professionals.* West Lafayette, IN: Purdue University.

Read, M. (2008). Libraries and repositories. *New Reivew of Academic Librarianship*, 14(1/2), 71–78.

Richards, N.M., & King, J.H. (2014). Big data ethics. *Wake Forest Law Review*, 49, 393–432.

Roberts, S.T. (2016). Commerical content moderation: Digital laborers' dirty work. In Noble, S.U. and Tynes, B. (eds.), *The intersectional Internet*. New York, NY: Peter Lang.

Roeschley, A., & Khader, M. (2020). Defining data ethics in library and information science. *iConference Proceedings, 2020*. Retrieved from http://hdl.handle.net/2142/106536

Rolfe, H. (2017). Inequality, social mobility and the new economy: Introduction. *National Institute Economic Review*, 240(1), R1–R4.

Rowley, J. (2007). The wisdom hierarchy: Representation of the DIKW hierarchy. *Journal of Information Science*, 33(2), 163–180.

Rubel, A. (2014). Libraries, electronic resources, and privacy: The case for positive intellectual freedom. *The Library Quarterly*, 84(2), 183–208.

Rubel, A., & Jones, K.M.L. (2016). Student privacy in learning analytics: An information ethics perspective. *The Information Society*, 32(2), 143–159.

Rubin, R. (2017). *Foundations of library and information science*. Chicago, IL: American Library Association.

Rüegg, J., Gries, C., Bond-Lamberty, B., Bowen, G.J., Felzer, B.S., McIntyre, N.E., ... & Weathers, K.C. (2014). Completing the data life cycle: using information management in macrosystems ecology research. *Frontiers in Ecology and the Environment*, 12(1), 24–30.

Sanfilippo, M.R., Shvartzshnaider, Y., Reyes, I., Nissenbaum, H., & Egelman, S. (2020). Disaster privacy/privacy disaster. *Journal of the Association for Information Science and Technology*, 71(9), 1002–1014.

Saracevic, T. (1999). Information Science. *Journal of the American Society for Information Science*, 50(12), 1051–1063

Schradie, J. (2020). The great equalizer reproduces inequality: How the digital divide is a class power divide. In Eidlin, B., & McCarthy, M.A. (ed.), *Rethinking class and social difference* (vol. 27). Bingley, UK: Emerald Publishing Limited.

Selwyn, N. (2020). Re-imagining learning analytics... a case for starting again? *Internet and Higher Education*, 46, article 100745.

Semeler, A.R., & Pinto, A.L. (2020). Librarianship in the age of data science: Data librarianship venn diagram. *International Conference on Data and Information in Online Environments*, 2020, 118–130. https://doi.org/10.1007/978-3-030-50072-6_10

Semeler, A.R., Pinto, A.L., & Rozados, H.B.F. (2017). Data science in data librarianship: Core competencies of a data librarian. *Journal of Librarianship and Information Science*, 51(3), 771–780.

Severson, H.H., & Ary, D.V. (1983). Sampling bias due to consent procedures with adolescents. *Addictive Behaviors*, 8(4), 433–437.

Shachaf, P. (2005). A global perspective on library association codes of ethics. *Library and Information Science Research*, 27(4), 513–533.

Shankar, K., Jeng, W., Thomer, A., Weber, N., & Yoon, A. (2020). Data curation as collective action during COVID-19. *Journal of the Association for Information Science and Technology*, 72(3), 280–284.

Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423.

Shen, Y. (2015). Strategic planning for a data-driven, shared-access research enterprise: Virginia Tech research data assessment and landscape study. *Proceedings of the Association for Information Science and Technology*, 52, 1–4.

Shilton, K. (2012). Participatory personal data: An emerging research challenge for the information sciences. *Journal of the American Society for Information Science and Technology*, 63(10), 1905–1915.

Shiri, A. (2016). Exploring information ethics: A metadata analytics approach. *Journal of Information Ethics*, 25(1), 17–37.

Si, L., Zhuang, X., Xing, W., & Guo, W. (2013). The cultivation of scientific data specialists: Development of LIS education oriented to e-science service requirements. *Library Hi Tech*, 31(4), 700–724.

Siguenza-Guzman, L., Saquicela, V., Avila-Ordonez, E., Vandewalle, J., & Cattrysse, D. (2015). Literature review of data mining applications in academic libraries. *The Journal of Academic Librarianship*, 41(4), 499–510.

Šimundić, A. (2013). Bias in research. *Biochemia Medica*, 23(1), 12–15.

Smith, J., & Noble, H. (2014). Bias in research. *Evidence-based Nursing*, 17(4), 100–101.

Smith, L.C. (1981). Citation analysis. *Library Trends*, 30(1), 83–106.

Song, I., & Zhu, Y. (2017). Big data and data science: Opportunities and challenges of iSchools. *Journal of Data and Information Science*, 2(3), 1–18.

Spector, W.S. (1956). *Handbook of biological data*. Washington, D.C.: United States' Department of Energy.

Spink, A., Wolfram, D., Jansen, B.J., & Saracevic, T. (2001). Searching the web: The public and their queries. *Journal of the American Society for Information Science and Technology*, 52(3), 226–234.

Špiranec, S., Kos, D., & George, M. (2019). Searching for critical dimensions in data literacy. *Information Research*, 24(4), paper colis 1922. https://InformationR.net/ir/24-4/colis/colis 1922.html

Stobierski, T. (2021). *8 steps in the data life cycle*. Retrieved from https://online.hbs.edu/blog/post/data-life-cycle

Taylor, R.S. (1966). Professional aspects of information science and technology. In Cuadra, C. (ed), *Annual Review of Information Science and Technology* (volume 1). New York, NY: John Wiley and Sons.

Tenopir, C., Hughes, D., Allard, S., Frame, M., Birch, B., Sandusky, R.J., Langseth, M.L., & Lundeen, A. (2015). Research data services in academic libraries: Data intensive roles for the future? *Journal of eScience Librarianship*, 4(2). https://doi.org/10.7191/jeslib.2015.1085

Tenopir, C., Rice, N.M., Allard, S., Baird, L., Borycz, J., Christian, L., Grant, B., Olendorf, R., & Sandusky, R.J. (2020). Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PloS One*, 15(3), article e0229003.

Tenopir, C., Sandusky, R.J., Allard, S., & Birch, B. (2014). Research data management services in academic research libraries and perceptions of librarians. *Library and Information Science Research*, 36(2), 84–90.

Tenopir, C., Talja, S., Horstmann, W., Late, E., Hughes, D., Pollock, D., Schmidt, B., Baird, L., Sandusky, R.J., & Allard, S. (2017). Research data services in European academic research libraries. *LIBER Quarterly*, 27(1), 23–44.

Thelwall, M. (2004). *Link analysis: An information science approach.* Amsterdam, NL: Elsevier.

Tremblay, M.C., Deckard, G.J., & Klein, R. (2016). Health informatics and analytics: Building a program to integrate business analytics across clinical and administrative disciplines. *Journal of the American Medical Informatics Association*, 23(4), 824–828.

Trepanier, C., Shiri, A., & Samek, T. (2019). An examination of IFLA and Data Science Association ethical codes. *IFLA Journal*, 45(4), 289–301.

Tukey, J.W. (1962). The future of data analysis. *The Annals of Mathematical Statistics*, 33(1), 1–67.

Turing, A.M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Journal of Math*, 58, 230–265.

van der Aalst, W.M.P., Bichler, M., & Heinzl, A. (2017). Responsible data science. *Business and Information Systems Engineering*, 59, 311–313.

Virkus, S., & Garoufallou, E. (2019). Data science from a library and information science perspective. *Data Technologies and Applications*, 53(4), 422–441.

Wallace, D.P. (1985). The use of statistical methods in library and information science. *Journal of the Association for Information Science and Technology*, 36(6), 402–411.

Waller, M.A., & Fawcett, S.E. (2013). Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management. *Journal of Business Logistics*, 34(2), 77–84.

Walter, S. (2008). The library as ecosystem. *Library Journal*, 133(16), 28–32.

Wang, L. (2018). Twinning data science with information science in schools of library and information science. *Journal of Documentation*, 74(6), 1243–1257.

Wang, Y., & Lin, C. (2019). A survey of data science programs and courses in the iSchools. *Proceedings of the ASIST Annual Meeting*, 56, 801–802.

Washington Durr, A.K. (2020). A text analysis of data science career opportunities and US iSchool curriculum. *Journal of Education for Library and Information Science*, 61(2), 270–293.

Westgard, J.O., & Hunt, M.R. (1973). Use and interpretation of common statistical tests in method-comparison studies. *Clinical Chemistry*, 19(1), 49–57.

White, H.D. (1977). *Machine-readable social science data*. Englewood, CO: Information Handling Services.

Whittaker, M., Alper, M., Bennett, C.L., Hendren, S., Kaziunas, L., Mills, M., …, & West, S.M. (2019). *Disability, bias, and AI*. New York, NY: AI Now.

Wilholt, T. (2009). Bias and values in scientific research. *Studies in History and Philosophy of Science: Part A*, 40(1), 92–101.

Wood, S.E. (2017). Police body cameras and professional responsibility: Public records and private evidence. *Preservation, Digital Technology, and Culture*, 46(1), 41–51.

Yarger, L., Payton, F.C., & Neupane, B. (2020). Algorithmic equity in the hiring of underrepresented IT job candidates. *Online Information Review*, 44(2), 383–395.

Zhang, P., & Benjamin, R.I. (2007). Understanding information related fields: A conceptual framework. *Journal of the Association for Information Science and Technology*, 58(13), 1934–1947.

Zimmer, M. (2008). Preface: Critical perspectives on web 2.0. *First Monday*, 13(3). https://firstmonday.org/ojs/index.php/fm/article/download/2137/1943

Zwitter, A. (2014). Big data ethics. *Big Data and Society*, 1(2), 1–6.