

Defining Data Ethics in Library and Information Science

Ana Roeschley¹[0000-0002-4748-9216] and Malak Khader¹[0000-0002-6908-6223]

¹ University of North Texas, Denton, TX 76207, USA
Ana.Roeschley@unt.edu

Abstract. In the library and information sciences (LIS), data ethics is an area of increasing focus. The purpose of this study is to answer these questions and comprehensively define data ethics in the LIS fields based on the diverse body of literature on the topic. Through an integrative literature review, we found four overarching themes in LIS literature on data ethics: privacy, research ethics, ethical ecosystems, and control. Additionally, these four themes gave us an opportunity to create a comprehensive definition of data ethics in the library and information science fields.

Keywords: Data Ethics, Privacy, Research Ethics, Ethical Ecosystems, Data Control.

1 Introduction

In the library and information sciences (LIS), data ethics is an area of increasing focus. However, while there is plentiful literature on data ethics in library and information science, data ethics is a broad inter-disciplinary topic and there is a need for an understanding of the topic's scope in the LIS fields. What is precisely meant when the phrase "data ethics" is used in LIS literature? What are the main themes within the scholarly literature on data ethics in the library and information sciences? The purpose of this study is to answer these questions and comprehensively define data ethics in the LIS fields based on the diverse body of literature on the topic.

Through an integrative literature review, we found four overarching themes in LIS literature on data ethics: privacy, research ethics, ethical ecosystems, and control. These findings allow us to contextualize key concerns for the field including data use and misuse, ethical behavior, protocols for data use, and conceptualizations of information societies. Additionally, these four themes gave us an opportunity to create a comprehensive definition of data ethics in the library and information science fields.

2 Background

The amount of data that is easily accessible continues to increase each day. With it, the amount of data about individuals that is made accessible through their everyday lives also continues to rapidly increase. When the Cambridge Analytica scandal broke in 2018, it became evident how easily consultant groups like Cambridge Analytica and

others could use and misuse vast amounts of personal data for profit [1]. In the case of Cambridge Analytica, private data was used in the hopes of swaying a major election [2]. While this misuse of personal data brought serious questions regarding the importance of ethical data use to light, LIS researchers and practitioners have long grappled with the ethical ramifications of data use and abuse [3], [4], [5], [6]. However, to this day, there has been no clear stated definition of data ethics within the library and information fields.

3 Method

We used “data ethics” as the search term in the databases Web of Science, Academic Search Complete, JSTOR, ProQuest, LISA, and Library Literature & Information Science Retrospective. We additionally delimited our search to sources that were English language and scholarly and peer-reviewed articles from the library and information science discipline. We systematically went through the titles, abstracts, and keywords of every article to ensure that data ethics were relevant to each article. Articles were then identified and selected if both of the keywords “data” and “ethics” appeared in the title, abstract, or subject terms. After compiling all the articles that fit the criteria, we were able to gather a total of 350 applicable articles. We did not limit our results by a certain time period, however the first instance of the term “data ethics” in the literature available to us was found in 1976. It’s important to note that our delimitation to scholarly and peer reviewed literature excludes landmark data ethics pieces like the HEW Report. Once we had our final list of articles compiled, we used grounded theory coding [7] to analyze the articles’ abstracts in order to determine the common ethical issues the authors presented in each article.

4 Results

After categorizing the issues described in the literature into groups, we found that there were four major themes: privacy, research ethics, ethical ecosystems, and control. We found that these themes have evolved over time and reflect the ethical challenges faced by LIS scholars and practitioners. The literature was synthesized and a comprehensive definition of data ethics within LIS contexts which is explained in the article discussion.

4.1 Theme 1: Privacy

Privacy is unsurprisingly a central concern for LIS scholars researching data ethics. The right to privacy is a core value in both the library and information fields at large and is reflected in professional codes of ethics. In the United States, the American Library Association’s (ALA) Code of Ethics third tenant is as follows, “We protect each library user’s right to privacy and confidentiality with respect to information sought or received and resources consulted, borrowed, acquired or transmitted” [8]. Similarly, in the international information science field, the Association of Information Science and Technology’s (ASIS&T) Professional Guidelines state that it is an information scientist’s

responsibility to “uphold each user’s, provider’s, or employer’s right to privacy and confidentiality” [9]. Yet, while ethical provisions and guidelines for privacy have been put into place throughout the LIS professions, concerns surrounding privacy are prevalent in literature on data ethics in LIS [10], [5], [6], [11]. With rising cybersecurity concerns surrounding the USA PATRIOT Act and proliferation of private third-party vendors, LIS literature has seen a number of post-2001 papers on how patron privacy can be protected [12], [13], [14].

4.2 Theme 2: Research Ethics

Research ethics is another interrelated area of concern that has arisen in the wake of technological advances. While the advent of freely available online data has brought recent attention to intellectual property violations, this has long been an issue of importance for LIS scholars. Our findings show that academic integrity is an important area of focus within the research ethics theme [15], [16], [17]. How researchers use, misuse, represent, and misrepresent their research data is a significant cause of study within the research ethics theme. Plagiarism and faking of data within scholarly communication is a concerning trend that implicates not just authors but journal editors as well [18].

Additionally, research ethics regarding the use of data in LIS include the ethical ramifications of using existing data that has been gathered through unethical means, ethically archiving sensitive data used for research, and informed consent in the era of social media [18], [20]. Throughout the literature in this area of LIS data ethics, we found articles addressing research methods, challenges to research, and ethical failures in research. This theme highlights articles by LIS scholars who are working towards understanding and improving how research in the field can be done more ethically. As Hernon states, this work is done to “advance the discussion of research and to expand the research base in library and information science to deal better with a host of issues, such as library services for important populations such as people of color” and people with disabilities [21].

4.3 Theme 3: Ethical Ecosystems

An ethical ecosystem is an environment in which one would expect to see a common standard for ethical behavior. This usually consists of groups working within the same virtual or physical setting. Ecosystems include, but are not limited to, professional environments such as business, medical, academic, and government sectors. They also include virtual communities, like social media platforms [22], [23]. Recently, there has been a strong focus on the ethical behaviors of social media websites [24]. Many social media platforms use application programming interfaces (APIs) that share information with outside parties [25]. Ethical ecosystems exist in many forms, whether they are explicitly defined through codes of ethics or tacitly understood by their members [26], [27]. These enforcements of ethical ecosystems can create an ethical “checks and balance” system among each other to ensure that every party maintaining the minimum level of ethical standards.

4.4 Theme 4: Control

The fourth theme discovered is “control”. For the purposes of this project, control addresses the security of data and legal limitations regarding data. Here we see the issues of governance, copyright, intellectual property, intellectual freedom, storage and sharing of information, government data, data protection, cybersecurity, big data, and data manipulation and integrity. The rise of big data has led to many innovations, including some that could help experts track health populations and understand information about crisis behavior change. Some of these innovations offer great value and insight; however, there is major concern with these innovations because large amounts of personal data are being captured, controlled, and stored without clear standards or understandings of individual rights to personal data. Ensuring that personal data is properly protected, used, and shared is a priority [28], [29]. Concerns with control, specifically data control, are heavily influenced by the notion of privacy, especially when considering big data. Organizations, corporations, and other entities are using private, personal information about those who access their services and are turning them into points of data for their benefit [30].

5 Defining Data Ethics in Library and Information Science Contexts

Through the exploration of ethical concerns regarding privacy, research ethics, data control, and the establishment of ethical ecosystems, LIS researchers also question how data ethics are perceived in different contexts [31], [32], [33]. LIS researchers not only interrogate data ethics, but they also ask how cultural, technological, and social contexts affect perceptions and concerns regarding data ethics. Through these examinations of data ethics within specific contexts, the LIS literature on the topic strives towards nuanced understandings of both individuals and the information worlds they occupy.

In the 1990’s LIS scholars [3], [4], [34], [35] asked how understandings and applications of principles regarding data ethics need to be adapted for an online environment. Similarly, Williams moved this conversation forward as the new media environment began to proliferate LIS research and education in the early 2000’s. Through a heuristic self-study, Williams posited that, “If teachers or researchers, or anyone for that matter, find themselves operating unethically, perhaps they need some form of inner work” [36].

Contextual and nuanced understandings of data ethics continue to be placed at the heart of the literature in the present era as well. Similar to Williams [36], Battley [27] also advocates inner work and reflexivity. Battley describes the ethical challenges faced by archival scholars and practitioners who are working towards and researching community engagement and participation. She states that “research methodologies themselves can be problematic. The ethno-centric, culturally charged nature of Western academic research has been highlighted by researchers within communities with a history

of colonization” [27]. Battley argues that participatory research and reflexivity are positive steps towards a more ethical research process.

Continuing the vein of examining how context affects understandings of data ethics, Ess [37] explores how ethical pluralism can help us understand how culture impacts perceptions of ethics. As he explains, we need to understand both Eastern and Western viewpoints on topics like privacy to better understand how to approach these complex topics with nuance. Ess argues that how we view data ethics is tied to the societal contexts within which we exist. Ethical pluralism can help library and information scholars and professionals in both understanding and working with communities that have been previously been marginalized, underrepresented, and misrepresented by the LIS field. Gardiner, McDonald, Byrne, and Thorpe [38] explain why a more pluralistic approach is necessary in the LIS field as they discuss the tensions between established Western colonial approaches to intellectual property and the approaches taken by Indigenous communities. They explain that Western approaches to scholarly communication and Western approaches to regulating intellectual property are often in direct conflict with the wishes and practice of Indigenous communities. Thus, Gardiner et al. argue that context must be taken into consideration when LIS professionals work with data sets created by Indigenous communities [38].

In addition to societal contexts, personal contexts affect how data ethics are interpreted. In the exploration of privacy, Mai argues that understandings of information philosophy and ethical approaches to information should be pragmatic. Mai explains that the context of “the ‘privacy situation’ shapes our understanding of privacy given the specifics of the situation” [39]. This finding is supported by Camarero, Antón, and Rodríguez who in a study on e-book piracy found that “readers’ personal traits may also shape their attitude towards piracy and towards the price of e-books” [28].

Personal, societal, cultural, and professional contexts all affect how data ethics are perceived within LIS literature. These contextual perceptions help to inform how we develop and implement codes of ethics. In a review of 35 international LIS codes of ethics and standards, Koehler and Pemberton (2000) found that overall, regardless of context, the protection of user privacy was highlighted across the board as a core ethical value [26]. However, in a study examining the implementation of privacy policies, Magi explains that there is still room “for professional library associations to focus on the practical problems that emerge as librarians try to follow the code of ethics” [40]. Fleischmann, Hui, and Wallace also explore the role of codes of ethics within LIS contexts. They explain that we need to “consider how these codes of ethics can be developed not from the top-down, but rather from the bottom up” [41]. Such a bottom-up approach can aid us in highlighting practical applications of codes of ethics in everyday professional situations.

Through this overview of the literature, we can see that within the LIS realms, data ethics is a complex and nuanced topic. However, a comprehensive definition of data ethics within LIS can be seen as follows: *Data ethics are the moral principles that guide the study and use of data within the LIS fields. Data ethics guide LIS scholars and practitioners to further the development of their work while simultaneously ensuring that individual privacy and confidential data are not compromised as a byproduct of*

this LIS work. Data ethics guide LIS researchers to maintain the integrity of their research through the protection of data and prevention of data manipulation. Data ethics are established to safeguard the spaces, communities, cohorts, and other LIS ecosystems within which data is created, accessed, and studied. Data ethics guide decisions regarding who is in control of data, who should be in control of data, and who should make decisions regarding how data is preserved and accessed in the future. These perceptions of data ethics are understood within their own cultural, social, personal, and professional contexts. Data ethics are enforced within LIS through the establishment and continued refinement of codes of ethics.

6 Conclusion

Our findings show that the LIS literature on data ethics is dominated by articles that examine the realms of privacy, research ethics, data control, and ethical ecosystems. Additionally, LIS literature on data ethics is also concerned with how we approach ethical questions within the LIS fields. We found a number of articles that work to define how LIS scholars and practitioners understand and implement data ethics. Synthesizing these works with the established four categories can bring us towards a comprehensive definition of data ethics in LIS.

References

1. Isaak, J., Hanna, M. J.: User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. *Computer*, 51(8), 56-59 (2018).
2. Berghel, H.: Malice Domestic: The Cambridge Analytica Dystopia. *Computer*, (5), 84-89 (2018).
3. Collier, G.: Information privacy. *Information Management & Computer Security*, 3(1), 41-45 (1995).
4. Forcht, K. A., Thomas, D. S.: Information compilation and disbursement: moral, legal and ethical considerations. *Information Management & Computer Security*, 2(2), 23-28 (1994).
5. Kaske, N. K.: Research methodologies and transaction log analysis: Issues, questions, and a proposed model. *Library Hi Tech*, 11(2), 79-86 (1993).
6. Marx, G. T.: Ethics for the new surveillance. *The Information Society*, 14(3), 171-185 (1998).
7. Charmaz, K.: *Constructing grounded theory*. 2nd edn. Sage, London (2014).
8. American Library Association Homepage, <http://www.ala.org>, last accessed 2019/12/14.
9. Association for Information Science and Technology Homepage, <https://www.asist.org>, last accessed 2019/12/14.
10. Hodel-Widmer, T. B.: Designing databases that enhance people's privacy without hindering organizations: Towards informational self-determination. *Ethics and Information Technology*, 8(1), 3 (2006).
11. Mittelstadt, B.: From individual to group privacy in big data analytics. *Philosophy & Technology*, 30(4), 475-494 (2017).
12. Breeding, M.: Issues and technologies related to privacy and security. *Library Technology Reports*, 52(4), 5 (2016).

13. Dixon, P.: Ethical issues implicit in library authentication and access management: Risks and best practices. *Journal of Library Administration*, 47(3-4), 141-162 (2008).
14. Salo, D., Kharfen, S.: Ain't nobody's business if I do (read serials). *The Serials Librarian*, 70(1-4), 55 (2016).
15. Auer, N. J., Krupar, E. M.: Mouse click plagiarism: The role of technology in plagiarism and the librarian's role in combating it. *Library Trends*, 49(3), 415-415 (2001).
16. Dadzie, P. S.: Rethinking information ethics education in Ghana: Is it adequate? *International Information and Library Review*, 43(2), 63-69 (2011).
17. Thomas, M.: University student and faculty opinions on academic integrity are informed by social practices or personal values. *Evidence Based Library and Information Practice*, 4(3), 49 (2009).
18. Thornton, M. A., Stewart, O. J., Rupp, D. E., Rogelberg, S. G.: Catalyzing ethical behavior among journal editors in the organizational sciences and beyond. *Journal of Information Ethics*, 23(2), 9 (2014).
19. Bramstedt, K. A., Xu, J.: China: A case study regarding transplant publishing issues. *Journal of Information Ethics*, 17(2), 12-22 (2008).
20. Waskul, D.: Considering the electronic participant: Some polemical observations on the ethics of on-line research. *The Information Society*, 12(2), 129-140 (1996).
21. Hernon, P.: (2000). Editorial: Survey research: Time for some changes. *The Journal of Academic Librarianship*, 26(2), 83.
22. Holley, E. G.: Librarians, 1876-1976. *Library Trends*, 25(1), 177 (1976).
23. Ulvik, S., Salvesen, G.: Ethical reference practice. *New Library World*, 108(7/8), 342-353 (2007).
24. Zimmer, M., Proferes, N.J.: A topology of twitter research: Disciplines, methods, and ethics. *Aslib Journal of Information Management*, 66(3), 250-261 (2014).
25. Lomborg, S., Bechmann, A.: Using APIs for data collection on social media. *The Information Society*, 30(4), 256-265 (2014).
26. Koehler, W. C., Pemberton, J. M.: A search for core values: Towards a model code of ethics for information professionals. *Journal of Information Ethics*, 9(1), 26 (2000).
27. Battley, B.: Co-producing archival research with communication, reflexivity and friendship: Crossing the three-wire bridge. *Archival Science*, 17(4), 371-391 (2017).
28. Camarero, C., Antón, C., Rodríguez, J.: Technological and ethical antecedents of e-book piracy and price acceptance. *The Electronic Library*, 32(4), 542-566 (2014).
29. Portmess, L., Tower, S.: Data barns, ambient intelligence and cloud computing: The tacit epistemology and linguistic representation of Big Data. *Ethics and Information Technology*, 17(1), 1-9 (2015).
30. Mai, J.: Big data privacy: The datafication of personal information. *The Information Society*, 32(3), 192-199 (2016).
31. Capurro, R.: Towards an ontological foundation of information ethics. *Ethics and information technology*, 8(4), 175-186 (2006).
32. King, S. A.: Researching Internet communities: Proposed ethical guidelines for the reporting of results. *The Information Society*, 12(2), 119-128 (1996).
33. Winston, M. D. Ethical leadership and ethical decision making: A meta-analysis of research related to ethics education. *Library & Information Science Research*, 29(2), 230-251 (2007).
34. Allen, C.: What's wrong with the 'golden rule'? Conundrums of conducting ethical research in cyberspace. *The Information Society*, 12(2), 175-188 (1996).
35. Mathieu, R. G., Woodard, R. L.: Data integrity and the Internet: implications for management. *Information Management & Computer Security*, 3(2), 3-7 (1995).
36. Williams, M.: Ethics in Research. *Australasian Journal of Information Systems*, 8(2), 68-74 (2001).

37. Ess, C.: Ethical pluralism and global information ethics. *Ethics and Information Technology*, 8(4), 215-226 (2006).
38. Gardiner, G., McDonald, J., Byrne, A., Thorpe, K.: Respect, trust and engagement: creating an Australian indigenous data archive. *Collection Building*, 30(4), 148-152 (2011).
39. Mai, J. E.: Personal information as communicative acts. *Ethics and Information Technology*, 18(1), 51-57 (2016).
40. Magi, T.: A study of US library directors' confidence and practice regarding patron confidentiality. *Library Management*, 29(8/9), 746-756 (2008).
41. Fleischmann, K. R., Hui, C., Wallace, W. A.: The societal responsibilities of computational modelers: Human values and professional codes of ethics. *Journal of the Association for Information Science and Technology*, 68(3), 543-552 (2017).