



*Information Research, Special Issue: Proceedings of the 15th ISIC - The Information Behaviour Conference, Aalborg, Denmark, August 26-29, 2024*

# Trends in data literacy, 2018-2023: a review of the literature

*Leanne Bowler and Charlie Shaw*

DOI: <https://doi.org/10.47989/ir292822>

## Abstract

**Introduction.** A review of the literature on data literacy from 2018 to mid-2023.

**Method.** A scoping review of the data literacy literature, in order to provide an overview of a topic, clarify key concepts, and identifying knowledge gaps.

**Analysis.** A detailed search strategy was applied to three data bases: SCOPUS, the ACM Digital Library, and the IEEE Xplore, followed by a qualitative analysis using a coding matrix of eight categories.

**Results.** Trends and gaps in the research were identified. The volume of research on data literacy is growing. Most data literacy research is in the context of formal learning environments. The focus of data literacy tends to be on digital data, interactions with big data systems, and the datafication of life. Critical data literacy approaches were identified in one third of the sample papers - a small but emerging stance to data literacy education. Most empirical studies about data literacy occurred in the Global North.

**Conclusion.** This paper contributes to our understanding of how data and data literacy are currently interpreted in the research literature, as well as the types of topics, skills, and data practices explored. It reveals developing trajectories and proposes future steps for building out the field of data literacy.

## Introduction

**This paper presents** preliminary research analysis from a scoping review of the English-language research literature on data literacy from 2018 to mid-2023. The purpose of the literature review is to reveal trends in the emerging field of data literacy. This paper reveals developing trajectories, contributes to themes related to the meaning of data and data literacy, as expressed in the research literature, and proposes future steps for building out the field of data literacy.

## Data and data literacy

There is no consensus definition of *data* because data is situated, taking its meaning from its context and the perspective of its beholder (Borgman, 2015, p. 18). As well, the dividing line between data and information is also not always clear, and, as Ford, in his book *Introduction to Information Behaviour* (2015) notes, 'the classification of a particular stimulus as data or information is relative and depends on a person's perception of a meaningful pattern in the data. Thus, one person's data might be another's information and vice versa' (p. 13). This blurring of boundaries has led to a call for a more unified approach to teaching data and information literacy in library education (Chiewphasa and Sisk, 2022). From this perspective, the ISIC realm of *information* behaviour and practices might rightly be expanded to include *data*.

The contextuality of how we use the term *data* is a conundrum for data literacy educators. Is it about statistics, computation, research evidence, or just life? Is data digital, numeric, or organic? In this complex soup of meanings, how do we plan for data literacy? As a result, definitions of data literacy vary across the literature and continue to evolve as a concept. At its most basic level, data literacy is about the tactics and procedures needed to find, process, organize, sort, and summarize datasets throughout the data life cycle (D'Ignazio and Bhargava, 2016). In the field of Information Science, the definition of data literacy includes skills necessary for the curation and preservation of data (Lyon and Brenner, 2015). Some interpretations lie closer to data science, viewing quantitative reasoning, numeracy,

statistical analysis, and computation as elemental to data literacy (for example, Schield, 2004), while others are more humanistic, and include dispositions that facilitate the ability to critique data practices and find meaning in data beyond statistical and mathematical arguments (Deahl, 2014; Finzer, 2013; Gray et al., 2018; Tygel and Kirsch, 2015; ). Arguments for a critical data literacy that raises awareness about the power embedded in data and associated social justice issues add yet another layer (Bilstrup et al., 2022; Chiewphasa and Sisk, 2022; Fotopoulou, 2021 ).

To understand the trajectory and landscape of the scholarly conversation about data literacy, we looked at a wide selection of the literature, using the technique of the scoping review. In this short paper, we report our preliminary findings, with further exploration of our data expected in the future.

## Methodology

A scoping review was our chosen method, due to the varied, multidisciplinary definitions and implementations of the term 'data literacy'. Scoping reviews gather existing research on a topic and summarize it. Such reviews are useful for providing an initial overview of a topic, for clarifying key concepts and definitions, identifying knowledge gaps, and can be the precursor to a larger systemic review (Arksey and O'Malley, 2005, Munn et al., 2018, Oliver et al., 2023).

We aimed to achieve an overview of the existing peer-reviewed literature on data literacy - specifically, how *data* and *data literacy* are defined; current research gaps, and the key concepts, competencies, skills, and knowledge being targeted. Due to the overlap between data and information, we included literature about information literacy education if it incorporated aspects of data. Results were deemed out of scope if they exclusively focused on math or computer science.

The literature review was conducted in July and August of 2023. The search was restricted to English language, scholarly, peer-reviewed documents published in 2018 or later. *Data literacy* was our sole search term, allowing us to focus on the similarities and variations between

definitions and conceptualizations of data literacy. As well, we sought results where full text was available, so as to be able to confirm and qualify content. We included empirical studies, practice guides, and position papers.

To encourage multidisciplinary results, the database SCOPUS was chosen as an initial source because SCOPUS collates results from a number of different academic databases without the duplicated results and non-peer reviewed sources often seen with Google Scholar. Two more industry-specific databases were then searched: the Association for Computer Machinery (ACM) Digital Library (DL) and the Institute of Electrical and Electronic Engineers (IEEE) Xplore. ACM DL and IEEE Xplore were chosen due to their prevalence in the accessible and relevant full texts from our SCOPUS search. All three databases were subject to the same search term and limitations.

The first, and most fruitful, search was done on SCOPUS and retrieved 228 search results. Upon review we found that 60 of these results (26%) were out of scope due to language or disciplinary focus. Another 85 results (37%) had inaccessible full texts due to institutional restrictions. This left us with 83 full texts (36%) retained and indexed within our review. The additional two databases indexed resulted in far fewer relevant results, as many were already retrieved during our SCOPUS search. ACM DL retrieved an additional 29 in-scope results and IEEE Xplore retrieved an additional four in-scope results; however, when duplicates were accounted for only 17 new texts were added to our review. We concluded the search with 100 results that were in-scope, with accessible full-texts. Results are summarized in Table 1 below:

Databases searched. Search term: "Data Literacy"	Initial # of results retrieved	Final results
SCOPUS	228	83
ACM Digital Library	35	10
IEEE Xplore	12	7
TOTAL:		100

**Table 1.** Results of search for literature about data literacy

The authors created a coding framework for analyzing the content of the retrieved papers that included eight overarching themes (age, learning environment, data type, aspects of data literacy (i.e., specific qualities), data concepts explored, country of publication, children and youth, and whether authors had provided a definition for data literacy) and 52 associated codes. Themes and codes arose from the authors' prior research in the area of data literacy, including earlier reviews of the

literature (for example, Acker and Bowler, 2018; Bowler et al., 2017, Bowler et al., 2022). Further interaction with the sample papers in this study allowed for sixteen additional thematic codes to be added to the broader themes. As well, a new theme was added for "Country", to which we applied the International Organization for Standardization (ISO) country codes. In Table 2 below, we summarize the coding scheme and provide some examples of the codes.

Themes	Initial codes (Deductive analysis)	Additional codes (Inductive analysis)
Age	7	2
Country	ISO 3166-1 country codes	-
Learning Environment	7 (e.g., library, school, higher education, etc.)	1
Data Type	10 (e.g., census, GIS, health, etc.)	5
Article defines data literacy? Y/N	2	-
If yes, definition of data literacy	Text	-
Aspect(s) of data literacy	8 (e.g., skills, teaching plans, assessment, etc.)	3
Data concept(s) explored	11 (e.g., privacy, data justice, Big Data, data bias, etc.)	5
Type of engagement with data (Children and Youth only)	7 (e.g., playful, social, etc.)	0
	52	16

**Table 2.** Summary of the Coding Scheme

## Results

The scoping review covered literature published from 2018 to mid-2023, the majority of which was published in 2020 or later (89%), indicating growing interest in data literacy over time. All papers in our sampling were peer-reviewed. Most data literacy research and teaching occurred in formal learning environments, with higher education being the most popular (48%), followed by the K-12 environment (i.e., primary and secondary school) (31%).

The majority of literature we reviewed focuses on the data literacy of adults of unspecified age, often either teachers, researchers, higher education students, or the general public. We found no literature that focused specifically on

elderly adults or pre-school learning environments and young children – pointing to a gap in research and practice. It is interesting to note that these two populations are not associated with workforce training and thus reveal an economic imperative to data literacy, rather than seeing it as a life-wide skill needed for the 21<sup>st</sup> century.

In terms of types of data explored through data literacy education, digital data - meaning data generated through the use of digital platforms and social media - was the most common type of data (51%). Digital data themes were also explored within a broader category we labelled *Big Data Literacy* (a term borrowed from Sander, 2020), which, in addition to digital data recording our online behaviour, also touches on

machine learning/artificial intelligence and the etiquette and best practices around everyday interactions with big data systems – in short, a systemic approach to the datafication of life. Civic and government data was also used as a data source in 15% of the sample papers (of those, US Census data in 2 cases), while geographic information system (GIS) data was used in 5% of the papers, and research data management in 27%.

We categorized the data concepts within our results, finding that big data/datafication to be the most frequent data concept (49%), followed by data bias, data justice, and open data (24%, 23%, and 22% respectively). Note that data concepts could overlap within one article.

Our review also sought to identify instances of critical data literacy, where data was explored through a social equity and justice lens, either by way of an explicit statement in a paper or implicitly, through the type of concepts explored (i.e., data bias, data justice, data ethics, and data rights). We saw that 30% of the sample papers took a critical approach to data literacy in our samples, suggesting that basic technical skills may no longer be sufficient in data literacy education.

Within our sample of 100 papers, we were able to identify a country source in sixty-two cases (two of which were coded for two countries). Seventy-four percent of papers coded for country source were from the Global North (US, Germany, UK, Ireland, Spain, Finland, Poland, Canada, Austria, Denmark, Iceland, Netherlands, Belgium, Slovenia, Estonia, Lithuania, and Switzerland), with 17% from South Asian and Asia-Pacific nations (including ten percent from China) and five percent from African nations. While we acknowledge that our search within the English-language literature probably skewed these results, we also note that the distribution of research from around the world demonstrates a global interest in data literacy, from countries with diverse forms of government, gross domestic product (GDP), population sizes, and industrial concerns.

## Discussion

A review of the literature on data literacy revealed current perceptions of data and data

literacy. The research and practice on data literacy, as reported in the literature, has focused on educational environments at the higher education level and K-12 (primary and secondary school). While students in higher education tend to learn statistical and research-data management skills, we found evidence at the K-12 level of creative and fun pedagogies around data literacy. Gamifying data (Legaki et al., 2022) and making data *playable* (Werning, 2020) are just two examples.

The literature we reviewed suggests that, while schools have moved forward in delivering data literacy education, there is limited evidence of data literacy programming and activities within public libraries, one example being the *Data Literacy with, by, and for Youth* project (Bowler et al, 2022a, 2022b, 2023). Anecdotally, a scan of library websites does show some activity in this area so the issue may be that librarians aren't publishing their accounts. Increasing data literacy efforts in public libraries could fill in gaps in terms of reaching seniors and pre-school children. More research and practice in the area of informal learning about data literacy is needed.

Our review revealed continued definitional issues around data literacy. From hard computational and statistical skills (categorized by some authors as data literacy when perhaps their work falls under the remit of data science) to more humanistic, socio-technical, and psychological approaches focusing on self-awareness of oneself as a data subject. Data literacy means many things to many people. Are we speaking the same language? Perhaps, as Fotopoulou (2021) points out, we should not look for a unified approach, declaring that it is '*unhelpful to talk about a single form of data literacy*' (p.1), but rather, a pluralistic approach to data literacies.

## Conclusion

We have much left still to uncover in our scoping review and the results presented here offer only a partial view. Future work will explore in greater depth how *data literacy* and associated data concepts are defined. While we make no claim to this being a comprehensive review, these preliminary results do point to

areas where further analysis is needed, in order to fill out the data literacy picture. For example, a variety of types of data, from digital data to government census data, were reflected in our sample. Each data type presents a distinct set of skills and practices and therefore, different approaches to both research and instruction. Furthermore, most empirical studies about data literacy that we reviewed occurred in the Global North, suggesting the need for global perspectives.

Further research should build typologies of data (similar to those for information, such as Bates (2006) and Buckland (1991), and many later works) and correlate the typology to data literacy method. A large majority of relevant literature on data literacy in our sample was published in 2020 or later, indicating increasing interest and need for more systematic approaches to understanding trajectories in both research and practice in data literacy.

## About the authors

**Leanne Bowler** is a Professor at the School of Information, Pratt Institute, in New York City. Her research and teaching focuses on young peoples' critical interactions with information and data, their technology practices, STEM learning, and how family, teachers, and out-of-school organizations such as libraries and museums can support young people's competencies in a socio-technical world. She can be reached at [lbowler@pratt.edu](mailto:lbowler@pratt.edu)

**Charlie Chayyim Shaw** is a Master of Science in Library and Information Science (MSLIS) student and Graduate Research Assistant at Pratt Institute's School of Information. His research interests include digital imaging, digital preservation and curation, and critical archival theory, particularly in the context of historical materials related to trans identity and drag performance. He can be reached at [cshaw@pratt.edu](mailto:cshaw@pratt.edu)

## References

- Acker, A., & Bowler, L. (2017). What is your data silhouette? Raising teen awareness of their data traces in social media. *#SMSociety17: Proceedings of the 8th International Conference on Social Media & Society*, 1-5. Association for Computing Machinery.
- Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19-32. <https://doi.org/10.1080/1364557032000119616>
- Bates, M. J. (2006). Fundamental forms of information. *Journal of the American Society for Information and Technology*, 57(8), 1033-1045.
- Bilstrup, K.E.K., Kaspersen, M.H., Lunding, M.S., Schaper, M.M., Van Mechelen, M., Tamashiro, M.A., Smith, R.C., Iversen, O.S. and Petersen, M.G., (2022, June). Supporting critical data literacy in K-9 education: three principles for enriching pupils' relationship to data. In *Proceedings of the 21st Annual ACM Interaction Design and Children Conference* (pp. 225-236).
- Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. Cambridge, MA: MIT Press.
- Bowler, L., Acker, A., Jeng, W., & Chi, Y. (2017). "It lives all around us": Aspects of data literacy in teen's lives. *Proceedings of the Association for Information Science and Technology*, 54(1), 27-35. <https://doi.org/10.1002/pra2.2017.14505401004>
- Bowler, L., Rosin, M., Lopatovska, I., & Vroom, L. (2022a). Methods of Engaging Teens in Conversations about Personal Digital Data: Public Library Context. In *Filipiak, D. & Kalir, J. H.*

(Eds.). (2022). *Proceedings of the 2022 Connected Learning Summit*. Pittsburgh, PA: ETC Press. July 28-29, 2022. pp. 9-17.

Bowler, L., Rosin, M., Lopatovska, I., & Vroom, L. (2022b). Co-designing data labs at the public library: Data literacy with, for, and by teens. *iConference 2022: Information for a Better World: Shaping the Global Future*, February 28 - March 4, 2022.

Bowler, L., Lopatovska, I., & Rosin, M.S. (2023). Teen-adult interactions during the co-design of data literacy activities for the public library: insights from a natural language processing analysis of linguistic patterns, *Information and Learning Sciences*, Vol. ahead-of-print No. ahead-of-print. <https://doi.org/10.1108/ILS-06-2023-0076>

Buckland, M. (1991). Information as Thing. *Journal of the American Society for Information Science*. Jun1991, 42(5), 351-360.

Chiewphasa, B. B., & Sisk, M. L. (2022). Leveraging Critical Information Literacy to Develop Social Justice-Minded Data Literacy Competencies. *Journal of Critical Digital Librarianship*, 2(1), 3.

Deahl, E. (2014). *Better the data you know: Developing youth data literacy in schools and informal learning environments*. Available at SSRN 2445621.

D'Ignazio, C., & Bhargava, R. (2016). DataBasic: Design Principles, Tools and Activities for Data Literacy Learners. *The Journal of Community Informatics*, 12(3).

Finzer, W. (2013). The data science education dilemma. *Technology Innovations in Statistics Education*, 7(2). <https://doi.org/10.5070/T572013891>

Ford, N. (2015). *Introduction to Information Behaviour*. London: Facet Publishing.

Fotopoulou, A. (2021). Conceptualising critical data literacies for civil society organisations: Agency, care, and social responsibility. *Information, Communication & Society*, 24(11), 1640-1657. <https://doi.org/10.1080/1369118X.2020.1716041>

Gray, J., Gerlitz, C., & Bounegru, L. (2018). Data infrastructure literacy. *Big Data & Society*, 5(2), 2053951718786316. <https://doi.org/10.1177/2053951718786316>

Legaki, N.-Z., Thibault, M., & Hamari, J. (2022). Gamified educational software for data literacy—A research through design approach to GANDALF. *Proceedings of the 17th International Conference on the Foundations of Digital Games*, 1-4.

Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher*, 42(1), 30-37. <https://doi.org/10.3102/0013189X12459803>

Munn, Z., Peters, M. D., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC medical research methodology*, 18, 1-7.

Oliver, G., Cranefield, J., Lilley, S., & Lewellen, M. (2023). Data Cultures: a scoping literature review. *Information Research*, 28(1), 3-29.

Shields, M. (2004). Information Literacy, Statistical Literacy, Data Literacy. *IASSIST Quarterly*, 28(2-3), 6. <https://doi.org/10.29173/iq790>

Sander, I. (2020). What is critical big data literacy and how can it be implemented? *Internet Policy Review*, 9(2). <https://doi.org/10.14763/2020.2.1479>

Tygel, A., & Kirsch, R. (2015). Contributions of Paulo Freire for a critical data literacy. In *Proceedings of Web Science 2015 Workshop on Data Literacy* (pp. 318-34).

Werning, S. (2020). Making Data Playable: A Game Co-Creation Method to Promote Creative Data Literacy. *Journal of Media Literacy Education*, 12(3), 88-101.