

Article

Not peer-reviewed version

---

# Citation Analysis of Global Research on Data Literacy

---

[Williams Ezinwa Nwagwu](#) \*

Posted Date: 20 August 2024

doi: 10.20944/preprints202408.1285.v1

Keywords: Citation patterns; Scholarly landscape; Data literacy; Scholarly attention; Co-citation analysis



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Article*

# Citation Analysis of Global Research on Data Literacy

Williams Ezinwa Nwagwu

University of Ibadan; we.nwagwu@ui.edu.ng

**Abstract:** This study analyzes the scholarly landscape of data literacy through citation and co-citation analyses of documents, sources, and authors. Using Scopus data and VOSviewer, the study identifies significant contributions and thematic trends. A minimum criterion of 10 citations per document was applied, filtering the dataset to 205 documents, with a focus on 81 interlinked documents. Citation analysis covered document, source, and author metrics, while co-citation analysis examined cited authors, sources, and references. The study found 997 documents on data literacy, narrowed down to 205 significant ones, with 81 interlinked documents showing a high average citation rate. Key sources included the ACM International Conference Proceeding Series and Teachers College Record, and prominent authors like Ellen B. Mandinach and Kim Schildkamp emerged as central figures. Data literacy research spans fields like education, sociology, and information science, highlighting its interdisciplinary nature. The study's focus on citation metrics may introduce selection bias, emphasizing widely cited works. Future research could explore less-cited but influential works and broader datasets to mitigate biases. Policymakers can use these insights to integrate data literacy into educational curricula and design targeted professional development programs. Promoting interdisciplinary collaboration and supporting open access to scholarly literature can enhance data literacy initiatives. This study provides a comprehensive citation and co-citation analysis of data literacy research, offering valuable insights into key contributions and thematic trends, informing policy and practice, and underscoring the importance of data literacy in contemporary education and society.

**Keywords:** citation patterns; scholarly landscape; data literacy; scholarly attention; co-citation analysis

---

## 1. Introduction and Literature Oversight

Data literacy is the ability to read, understand, create, and communicate data as information. It is a subset of both visual, information and other literacies, and is an important skill for knowledge workers, consumers, and in modern and traditional cultures (Wang and Strong 1996, Levitan and Verhulst 2016, Mandinach and Gummer 2016, O'Connor 2021). Data Literacy comprises several interconnected components and dimensions, integrating a blend of skills, knowledge, and attitudes essential for comprehensive proficiency. Data Literacy involves a range of technical skills, including proficiency in data analysis, statistical reasoning, data visualisation, and proficiency in using relevant software or programming languages for data manipulation and interpretation. It encompasses an understanding of data concepts, such as different data types (quantitative, qualitative, mixed methods), data sources, data collection methods, and an awareness of ethical considerations related to data handling and usage (Verhulst 2016, Nwagwu 2024).

Data Literacy encompasses attitudes conducive to a data-driven mindset, including curiosity, critical thinking, skepticism, and an appreciation for the role of data in decision-making. It involves the willingness to explore and question data, acknowledging both its potential and limitations. These components collectively form the foundation of Data Literacy, equipping individuals with the ability to not only comprehend and interpret data but also to critically evaluate its relevance, make informed judgments, and effectively communicate insights derived from data to diverse audiences. As the digital landscape continues to evolve, Data Literacy remains an evolving concept, demanding

continual adaptation and acquisition of new skills to navigate and harness the potential of data in an ever-changing world (D'Ignazio 2017, Stanton et al. 2017).

The importance of Data Literacy in modern society cannot be overstated, as it serves as a cornerstone for informed decision-making, transformative problem-solving, innovation, and ethical considerations across diverse domains. Its significance lies in its capacity to empower individuals and organisations to harness the potential of data in a rapidly evolving digital landscape. Data Literacy is indispensable in an era where data has proliferated across every aspect of society. The ability to navigate, interpret, and derive insights from this abundance of information is critical for individuals and organisations alike. In a world inundated with data, being data-literate isn't just advantageous; it is an imperative skill necessary for success in various spheres. Data Literacy forms the bedrock of informed decision-making. It enables individuals to base their choices on evidence and insights derived from data rather than intuition or guesswork. Decisions driven by data are often more accurate, effective, and strategic (Osasona, et al. 2024)).

Gould (2017) touches upon the importance of Data Literacy in understanding probabilities and making decisions based on statistical reasoning. He discusses how a grasp of Data Literacy aids in distinguishing chance occurrences from significant trends. Furthermore, he argues that Data Literacy is foundational in preparing individuals for the demands of a data-driven job market and the importance of informed decision-making therein. Saltu-Rivas et al. (2022) highlights the crucial role of Data Literacy in fostering informed decision-making processes. She emphasises how enhancing Data Literacy skills empowers individuals to utilise data effectively for better decision outcomes. Morrow (2021) discusses the importance of Data Literacy in organisational contexts. He illustrates how improved Data Literacy aids in better communication and decision-making within businesses and institutions. Also, Heiser et al. (2023) explores the significance of Data Literacy in managerial roles. She emphasises how Data Literacy equips managers with the competence to make informed decisions based on data-driven insights.

In the age of big data, ethical considerations surrounding data usage are paramount. Data Literacy includes an understanding of ethical principles related to data privacy, security, bias, and responsible data handling. It empowers individuals to navigate these ethical complexities, ensuring that data is used ethically and responsibly. Data Literacy empowers individuals and organisations across diverse fields. In healthcare, it aids in diagnosing illnesses, predicting outbreaks, and personalising treatments. In business, it drives marketing strategies, operational efficiencies, and customer insights. In education, it improves student performance analysis and personalised learning (Ghodoosi 2023). Policymakers leverage Data Literacy to craft evidence-based policies, while scientists use it to advance research and innovation. Ultimately, Data Literacy empowers individuals and organisations to extract meaningful insights from data, enabling them to make better decisions, solve complex problems, foster innovation, and navigate ethical considerations effectively. As data continues to permeate every facet of society, the importance of Data Literacy will only continue to grow, shaping a more informed, innovative, and ethically conscious world. These opinions are supported by several authors (Aldboush and Ferdous 2023).

Debruyne et al. (2022) explores the ethical implications of algorithms and data-driven decision-making, emphasizing the need for ethical considerations within Data Literacy in various societal contexts. Loukides, Mason and Patil (2018) emphasises the ethical responsibilities surrounding data usage. He advocates for a broader understanding of Data Literacy that includes ethical considerations in handling data, emphasizing the ethical dimensions of Data Literacy. Nissenbaum (2009) delves into the ethical dimensions of Data Literacy, stressing the importance of understanding the ethical implications of data use and its societal impact, especially concerning privacy and social implications.

The impact of Data Literacy spans across various domains, revolutionizing approaches, and driving transformative changes. Let us delve into some compelling instances and success stories that highlight the profound effects of enhanced Data Literacy in diverse fields like healthcare, business, education, and policymaking. Data Literacy has been a game-changer in healthcare, significantly influencing patient care and public health initiatives. Imagine the power of predictive analytics, where healthcare professionals can foresee potential health risks for individuals based on data

patterns, enabling early interventions and personalised treatments. During outbreaks or pandemics, data-driven insights aid in predicting disease trends, allocating resources effectively, and formulating targeted public health strategies. For instance, analysing epidemiological data assists in identifying areas susceptible to outbreaks, guiding authorities in implementing preventive measures.

In the corporate world, Data Literacy empowers organisations to make informed decisions, optimise operations, and drive innovation. Companies leveraging data analytics gain deep insights into consumer behaviour, enabling them to tailor marketing strategies, optimise supply chains, and develop products aligned with market demands. Consider the impact of data-driven decision-making in retail, where inventory management based on analytics minimises excess stock, maximises sales opportunities, and enhances overall efficiency, thereby reducing costs and boosting profitability. Data Literacy in education is reshaping learning paradigms. Educators armed with Data Literacy skills analyses student performance data to personalise learning experiences. This enables tailored teaching methodologies, identifies learning gaps, and provides targeted interventions to ensure every student's needs are met. Institutions utilising data-driven insights at a systemic level enhance curriculum design, resource allocation, and policy formulation to foster a conducive learning environment.

In governance, Data Literacy fuels evidence-based policymaking. Governments leverage data analytics to devise more effective policies in various sectors. Whether it is designing healthcare policies for improved service delivery, urban planning based on demographic trends, or optimising resource allocation in public sectors, data-driven insights guide policymakers in making informed decisions. Monitoring policy effectiveness through data analysis allows for adaptive policymaking, enhancing governance and citizen welfare. Success stories abound across these domains, showcasing tangible impacts of improved Data Literacy. From reducing hospital readmission rates through predictive analytics in healthcare to optimising inventory management for cost-efficiency in business, and from personalised learning experiences in education to evidence-based policymaking in governance, Data Literacy's transformative effects are evident. In each success story, Data Literacy acts as the catalyst, empowering individuals and organisations to harness the potential of data, make informed choices, drive innovation, and ultimately, bring about positive, measurable changes. As Data Literacy continues to evolve and permeate various sectors, its transformative impact will remain a driving force in shaping a more data-driven, efficient, and innovative world (Ongena 2023, Jiang 2023, Olszewski and Abukhdier 2023, Bartholo, Koslinski and de Castro 2022).

The concept of Data Literacy has emerged as a fundamental skill that transcends boundaries and holds immense significance across diverse domains. It represents more than just an understanding of numbers and statistics; rather, it embodies a multifaceted competency involving skills, knowledge, and attitudes necessary to navigate the ever-expanding sea of data. The evolution of Data Literacy has been remarkable. Initially rooted in statistical analysis and data interpretation, it has evolved to encompass a broader spectrum of proficiencies. It now demands technical skills in data manipulation, visualisation, and an understanding of ethical considerations surrounding data usage. This evolution reflects the dynamic nature of data itself and the need for individuals to adapt to emerging technologies and methodologies (Vance, Glimp, Pieplow, Garrity and Melbourne 2022).

Across healthcare, business, education, policymaking, and numerous other fields, Data Literacy plays a pivotal role. In healthcare, it enables predictive analysis for personalised treatments and proactive healthcare interventions, thereby improving patient outcomes. In businesses, Data Literacy empowers organisations to make informed decisions, optimise operations, and foster innovation by leveraging consumer insights and market trends. Within education, it revolutionises teaching methodologies, allowing educators to personalise learning experiences and enhance student outcomes. In policymaking, it drives evidence-based decisions, leading to more effective policies and governance (Pins et al. 2022).

Assessing Data Literacy remains a challenge due to its multifaceted nature. Evaluating skills, knowledge, and attitudes related to data involves subjective elements, making standardised assessments difficult (Santos, & Pedro & Mattar, 2021)). Furthermore, the rapid evolution of technology requires continuous refinement of assessment tools to ensure relevance and accuracy.



Success stories in these domains underscore the transformative impact of enhanced Data Literacy. From predicting disease outbreaks to optimising supply chains and tailoring educational approaches, Data Literacy enables innovation, efficiency, and informed decision-making. Looking ahead, there is a need for concerted efforts to promote Data Literacy. This entails integrating it into educational curricula, refining assessment methodologies, fostering ethical data practices, and encouraging cross-disciplinary collaboration. The ongoing cultivation of a data-literate society is crucial to navigating the complexities of our data-centric world, driving innovation, and ensuring responsible and informed decision-making across all facets of society. As the data landscape continues to evolve, the significance of Data Literacy remains steadfast, shaping a future where individuals and organisations harness data's potential for societal advancement and positive change (Wilkerson, Lanouette and Shareff 2021).

Research in data literacy is widely acknowledged as crucial for the advancement of educational programs and pedagogies. However, there exists uncertainty regarding the scope of data literacy and its optimal integration into educational curricula. While consensus exists on the necessity of data literacy among the general population, educators and policymakers often lack clarity on its specific components, resulting in sporadic efforts to incorporate it into educational standards. To address this, a clearer conceptualization of data literacy is needed, encompassing the knowledge and cognitive skills required for interpreting and evaluating data across diverse contexts, including personal decision-making, civic engagement, and scientific inquiry (Gehrke, Kistler, Lübke, Markgraf, Krol and Sauer 2021).

This clarification will facilitate the development of assessments, teaching materials, and educational standards tailored to fostering data literacy skills among students of all ages. Furthermore, research in data literacy aims to provide instructional support across disciplines, enabling students to effectively understand and utilise data in their academic pursuits and daily lives. As access to high-quality data becomes increasingly prevalent in various fields, proficiency in data manipulation and interpretation is essential for informed decision-making in education, business, and public policy (Phadkule 2022).

Nwagwu (2024) has x-rayed the literature on data literacy, observing, among others, that there is a growing interest on research in the area. How do we examine the connections between different pieces of work, such as articles or books already carried out on data literacy? Imagine that one is looking at a big web of knowledge on literacy, and each citation is considered to be link between two points. When someone writes a research paper or a book, they often reference other works they used for their research. Citation analysis offers a way to look at these references to see which works are being referenced the most, which ones are influencing others the most, and how ideas flow between different authors and topics. Researchers use citation analysis to understand trends in research, identify key authors and publications in a field, and track the impact of their own work or others.

Citation analysis involves the evaluation of scholarly works based on their citations, examining patterns and trends to understand influence, impact, and scholarly communication dynamics (Garfield, 1972). It is commonly employed in bibliometrics and scientometrics to assess research productivity and impact (Moed 2005). Citation analysis, a methodical exploration of citations within scholarly literature, offers a quantitative lens into the multifaceted realm of research communication and impact. By scrutinizing citation patterns, including frequency and context, researchers glean valuable insights into the influence and significance of individual works, authors, journals, and research fields (Abramo, D'Angelo & Murgia, 2019). Researchers harness citation analysis for myriad purposes. Firstly, they assess impact by scrutinizing the citations garnered by specific papers, authors, or journals, thereby gauging their resonance within the scholarly community (Bordons, Fernández & Gómez, 2002). Evidently, citation analysis uncovers trends in research topics, interdisciplinary collaborations, and burgeoning fields, aiding in the identification of evolving scholarly landscapes (Bornmann & Leydesdorff 2014).

Furthermore, citation analysis serves as a pivotal tool for evaluating authors and institutions, empowering funding agencies and academic entities to gauge productivity and impact with precision (Larivière et al., 2016). By mapping intricate knowledge networks through citation networks,

researchers delineate the propagation of ideas and identify pivotal contributors and influential nodes, enriching our understanding of research dynamics (Wang et al. 2019). Additionally, citation analysis facilitates the ranking of academic journals based on impact and prestige, thereby informing strategic decisions regarding research dissemination and publication avenues (Waltman & van Eck, 2012). Overall, citation analysis offers indispensable insights into the scholarly landscape, equipping stakeholders with the requisite intelligence for informed decisions pertaining to research priorities, collaborations, and resource allocation (Alonso, Cabrerizo, Herrera-Viedma & Herrera 2009).

Related to citation analysis is co-citation analysis. Co-citation refers to the frequency with which two works or more are cited together by other authors in their own publications. It is a measure of the relatedness or similarity of the content of two documents based on the citations they receive. For example, if Author A and Author B are frequently cited together in other works, it suggests a strong connection between their research and ideas (Small July 1973, Hjørland and Nicolaisen, 2005). This could indicate that they are working on similar topics or that their work complements each other in some way. Studying co-citation patterns can reveal intellectual connections between authors, identify influential works, and uncover emerging trends or research areas within a field. It is often used in bibliometric analysis to map the intellectual structure of a discipline or to identify key players in a particular research area. Co-citation of cited authors refers to the phenomenon where two or more authors are cited together in the same paper by another author (Gipp and Beel, 2009, Small and Klavans, 2013).

Co-citation is essentially a measure of how often two documents are cited together by other documents. When at least one other document references the same two documents, they are considered co-cited. The frequency of these co-citations indicates the strength of their relationship and suggests they are likely related in meaning. Similar to bibliographic coupling, co-citation is a way to gauge semantic similarity between documents through citation analysis. Imagine there is a diagram illustrating this concept.

Research in data literacy is widely acknowledged as crucial for the advancement of educational programs and pedagogies. However, there exists uncertainty regarding the scope of data literacy and its optimal integration into educational curricula. While consensus exists on the necessity of data literacy among the general population, educators and policymakers often lack clarity on its specific components, resulting in sporadic efforts to incorporate it into educational standards. To address this, a clearer conceptualization of data literacy is needed, encompassing the knowledge and cognitive skills required for interpreting and evaluating data across diverse contexts, including personal decision-making, civic engagement, and scientific inquiry. Specifically, understanding the connections between different pieces of work, such as articles or books, through citation analysis and co-citation analysis, is crucial for advancing our understanding of data literacy.

Citation analysis offers a quantitative lens into the multifaceted realm of research communication and impact, enabling researchers to assess the influence and significance of individual works, authors, journals, and research fields. Co-citation analysis, on the other hand, provides insights into the relatedness or similarity of the content of two documents based on the citations they receive, helping to identify intellectual connections between authors, influential works, and emerging trends within a field. Despite the importance of citation analysis and co-citation analysis in advancing our understanding of data literacy, there remain challenges in evaluating data literacy skills and integrating findings from citation analysis and co-citation analysis into educational practices effectively.

## 2. Methodology

Citation analysis requires access to the database(s)/index(es) where the publications of documents in the area of the subject, authors, institutions and/or countries whose citation is being conducted. This enables the counting of the number of times an article is cited by other works to measure the impact of a publication or author. The most popular databases include Google Scholar, Web of Science and Scopus. This assessment is concerned with global research on data literacy as indexed by Scopus, a database of Elsevier Publishers. There have been expressed concerns about the

selective policies of Web of Science regarding how these policies affect Africa and the need to democratize indexation of global research evidence (Nwagwu 2006, Nwagwu 2010, Asubiaro 2022). On this basis, Scopus was used for this analysis.

Scopus, a leading academic database, offers a vast array of scholarly literature across numerous disciplines, including journals, conference proceedings, books, and patents. Researchers can access it through institutional or direct subscriptions, enabling them to explore specific topics through targeted searches. Upon identifying relevant publications, researchers extract citation data from Scopus, which provides valuable metrics such as citation frequency and total citations. This data forms the basis for detailed analysis, where researchers trace citation patterns, explore trends over time, and identify influential works. Visualizations aid in interpreting and communicating findings, which ultimately contribute to a deeper understanding of scholarly impact. Conclusions drawn from this analysis inform future research directions and are shared through reports, articles, or presentations, with proper acknowledgment of Scopus as the source of the data. Thus, Scopus facilitates a journey of academic exploration, illuminating the pathways of knowledge dissemination and discovery (Burnham 2006, Baas et al. 2020).

### *2.1. Data Retrieval*

We specified the syntax "DATA LITERACY" in the Article, Title, Keyword column Scopus covering the period 2023. While it is obvious that there may be studies that have addressed this subject using other syntax, "Data Literacy" will yield an appropriately representative sample of keywords in the area. Any studies on "Data Literacy" that does not use the syntax must be addressing the subject from a very tangential perspective, and may not contribute strongly to the motif of this study. All irrelevant items were deleted, resulting to 997 documents. On the Scopus interface, we selected year, author, document type, and subject area as visualised in the "Analyze Results" resource of Scopus. The data for year was transferred to MS Excel and visualised in that environment. In the resulting interface, we selected "Abstract ad Keywords" option, and thereafter exported and saved as a CSV file.

### *2.2. Data Analysis*

Data was analysed using Vosviewer.

In Vosviewer we selected Citation under Type of Analysis, and then as Unit of Analysis, selected authors, documents, sources, organisations and countries respectively, one after the other. The counting method preferred was the Fractional counting, a preference explained in Nwagwu (2023) and Nwagwu (2024). Thereafter we selected Co-citation and Cited references, Cited sources and Cited authors, one after the other. The result was displayed by Cluster, Links, Citations, Norm. citations and Pub. Year. As usual, Vosviewer provides maps and tables.

## **3. Results**

### *3.1. Citation by Documents (2005-2023)*

A total 997 documents were written on data literacy, a mean of 52 documents per year. We placed a minimum number of 10 citations per document and this resulted to 205 documents out of which only 81 were connected to other documents and were the basis for this analysis. The 81 documents produced 2768 citations or a mean of 34 citations per document. Prado is 2013 "Incorporating Data

**Table 1.** Citation by Top Thirty Documents (2005-2023).

	<b>Label</b>	<b>Cluster</b>	<b>Links</b>	<b>Citations</b>	<b>Norm. citations</b>	<b>Pub. Year</b>
1	Prado (2013)	3	8	161	3.0377	2013
2	Gray (2018)	1	2	131	10.1307	2018
3	Pangrazio (2019)	8	10	123	9.8743	2019
4	Schildkamp (2015)	2	2	114	3.1947	2015
5	Schildkamp (2019)	2	2	89	7.1449	2019
6	Hoogland (2016)	2	9	83	3.5485	2016
7	Koltay (2017b)	5	9	79	3.5333	2017
8	Koltay (2015b)	1	17	76	2.1298	2015
9	d'ignazio (2017)	4	4	76	3.3992	2017
10	Gould (2017)	7	5	74	3.3097	2017
11	Carmi (2020)	8	1	65	7.18	2020
12	Kippers (2018)	2	5	64	4.9493	2018
13	Koltay (2016b)	1	6	60	2.5652	2016
14	Mandinach (2021a)	2	3	56	8.1654	2021
15	Pangrazio (2020)	6	3	55	6.0753	2020
16	Reeves (2015)	9	1	52	1.4572	2015
17	Macmillan (2014)	1	2	49	4.6838	2014
18	Cowie (2017)	6	2	48	2.1468	2017
19	Raffaghelli (2020a)	7	5	44	4.8603	2020
20	Stephenson (2007)	5	8	41	1.9524	2007
21	Ebbeler (2017)	2	3	40	1.789	2017
22	Stornaiuolo (2020)	10	3	37	4.0871	2020
23	Wolff (2019)	6	5	36	2.8901	2019
24	Koltay (2019)	1	5	36	2.8901	2019
25	Athanases (2012)	9	1	34	0.2163	2012
26	Bowler (2017)	6	6	33	1.4759	2017
27	Maybe (2015)	3	7	32	0.8968	2015
28	Federer (2016)	1	1	32	1.3681	2016
29	Pothier (2020)	5	2	31	3.4243	2020
30	Lee (2021)	10	1	31	4.5201	2021

Literacy into Information Literacy Programmes: Core competences and content" in *Libri* has been cited (161) more than all papers on the subject matter. Prado is paper was based on the observation that rise of the importance of data in society necessitates libraries' integration of data literacy into their information programs. The paper proposed a framework of core competencies to address this need, facilitating the development of resources and guiding further research.



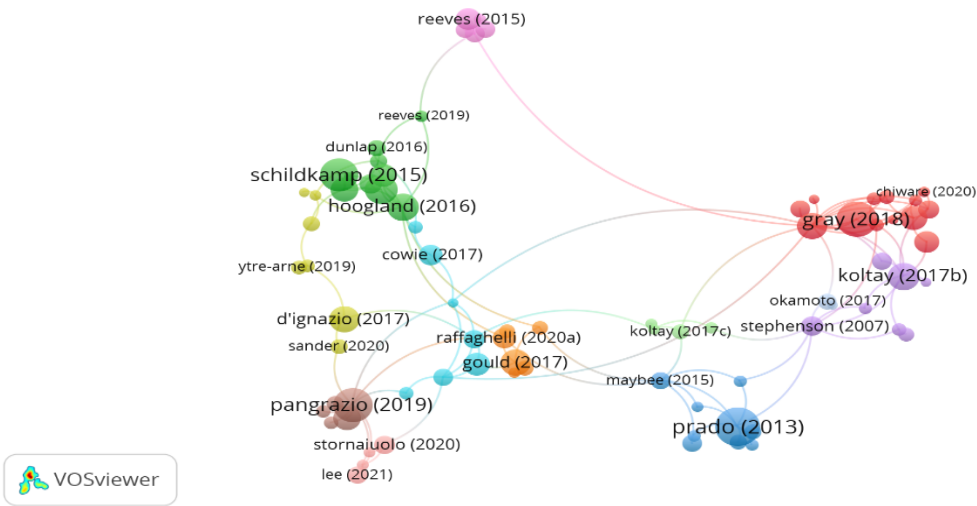


Figure 1. Citation by Top Thirty Documents (2005-2023).

Flowing is Gray, Gerlitz and Bounegru is (2018) “Data infrastructure literacy” in *Big Data and Society* In this paper, the author reminisces on a report from the UN that makes the case for “global data literacy” in order to realise the opportunities afforded by the “data revolution”. The document has been cited 131 times. Prado is paper has both higher links (8) and cluster (3) than Gray is 2 and 1 respectively.

3.2. Citation by Sources (2005-2023)

The total number of sources indexed is 546; for minimum number of documents per source placed at 2 resulted to 156 documents out of which only 105 were linked, and were used in the analysis. We sorted the data according to documents first, and then by citations. Citations are usually indexed with reference to the documents being cited; hence sorting by documents yields some great insights about citations by sources. By this token, it can be seen that *ACM International Conference Proceeding Series* published the highest number of documents on the subject matter (47) and this documents were cited 260 times during the period. *Communications in Computer and Information Science* published 23 papers that were cited 79 times.

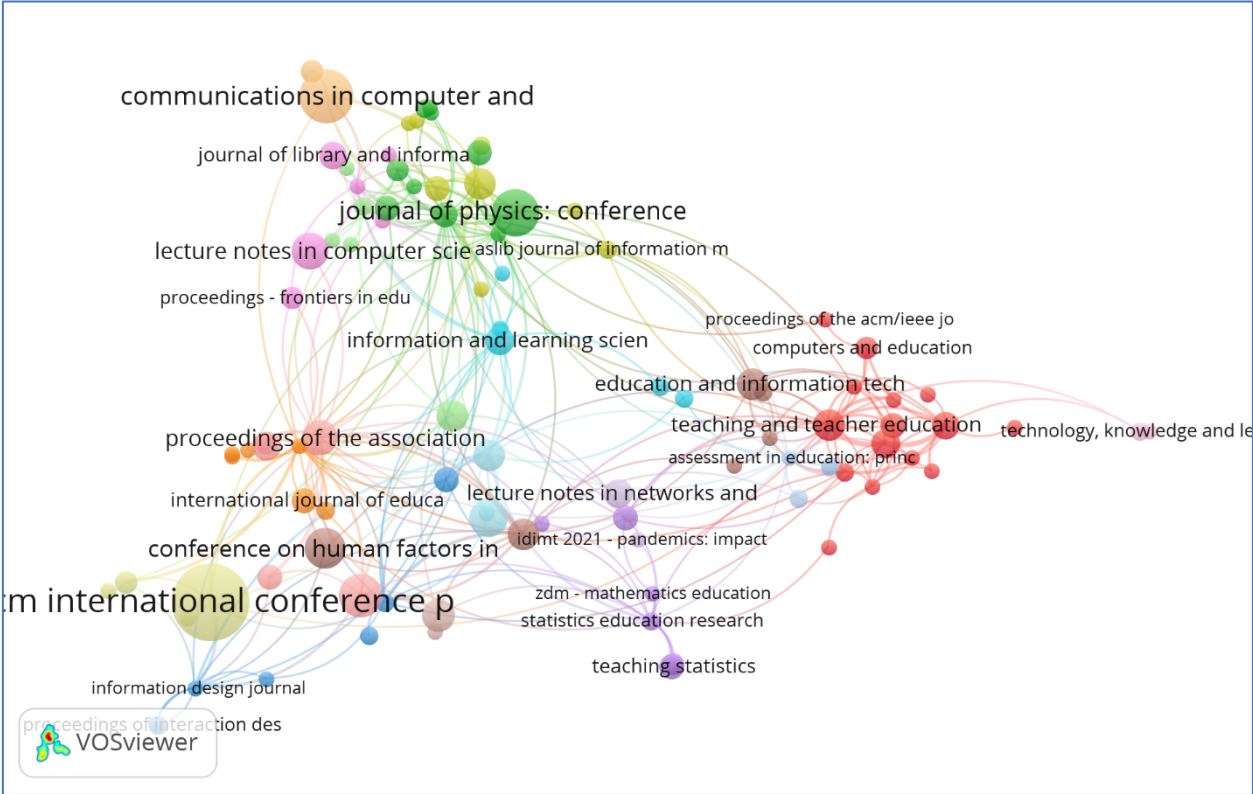


Figure 1. Citation by Sources.

*The Teachers College Record*, (TCR) “... a journal of research, analysis, and commentary in the field of education that has been published continuously since 1900 by Teachers College, Columbia University, ranks the first. The journal published only six articles, once in 2012, and five times in 2015 – the journal, with only one cluster has been cited 429 times altogether. A single cluster and numerous citations, typically indicates focused article publication that draws considerable attention from researchers in a specific field, cluster here denoting a specialised area or narrow subject domain within the journal is broader coverage. *Teaching and Teacher Education* is the next journal after TCR. It has featured eight documents on the subject, and has a single cluster and 389 citations. But it has higher links (16) and Total Link Strength (30) than TCR (13) and 23 respectively. *Educational Researcher*, obviously has a wider focus and has the third highest number of citations (270). By number of documents published in the sources, *ACM International Conference Proceeding Series* and *Communications in Computer and Information Science*, 47 and 23 respectively. *ACM International Conference Proceeding Series* also accounted for the fourth highest number of citations.

Table 2. Citation by Top Thirty Sources.

Label	Cluster	Li nk strength	Total link strength	Docu ment s	Cita tion s	Norm. citation s	Avg. pub. year	Avg. cita tio ns	Avg. norm. citations
1ACM International Conference Proceeding Series	13	8	11	47	260	27.1754	2020.064	5.5319	0.5782
2communications in computer and information science	16	6	10	23	79	5.0262	2018.696	3.4348	0.2185
3Journal of Physics: Conference Series	2	5	6	18	45	4.951	2020	2.5	0.2751
4Proceedings of International Conference of the Learning Sciences, ICLS	10	8	12	15	16	4.8611	2021.067	1.0667	0.3241
5Conference on Human Factors in Computing Systems – Proceedings	8	4	4	13	239	25.534	2019.385	18.3846	1.9642
6CEUR Workshop Proceedings	15	5	6	12	41	3.5447	2019.5	3.4167	0.2954
7Lecture Notes in Computer Science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)	9	2	2	11	21	1.8881	2019.455	1.9091	0.1716

8Proceedings of the Association for Information Science and Technology	10	25	29	10	88	5.9689	2018.8	8.8	0.5969
9Computer-Supported Collaborative Learning Conference, CSCL	17	7	8	9	20	2.0777	2019.22 2	2.2222	0.2309
10Teaching and Teacher Education	1	16	30	8	389	20.7427	2018	48.625	2.5928
11British Journal of Educational Technology	8	14	20	8	74	22.5432	2021.5	9.25	2.8179
12Journal of Business and Finance Librarianship	4	10	17	8	54	17.1122 5	2020.62 5	6.75	2.139
13Education and Information Technologies	8	8	9	8	43	40.4072	2022.755	3.75	5.0509
14Journal of Media Literacy Education	15	15	21	8	40	4.4184	2020	5	0.5523
15Higher Education Dynamics	11	8	10	8	1	1.8713	2023	0.125	0.2339
16Studies in Educational Evaluation	1	8	20	7	179	20.0918 7	2019.8525 74	5.71	2.8703
17Information and Learning Science	6	18	20	7	21	4.5491	2022	3	0.6499
18Journal of Map and Geography Libraries	10	1	1	7	18	1.8239 4	2019.71 4	2.5714	0.2606
19Lecture Notes in Networks and Systems	14	2	2	7	0	0	2022.57 1	0	0
20Teachers College Record	1	13	23	6	429	11.2856	2014.5	71.5	1.8809
21Journal of Library and Information Science in Agriculture	9	1	1	6	2	0.2916	2021	0.3333	0.0486
22Big Data and Society	2	3	3	5	186	17.0758	2019.2	37.2	3.4152
23Journal of Documentation	2	30	42	5	128	7.1709	2018.6	25.6	1.4342
24Journal of Academic Librarianship	4	8	15	5	107	13.753	2019.6	21.4	2.7506
25International Journal of Educational Technology in Higher Education	7	3	3	5	88	16.7639	2021.2	17.6	3.3528
26information communication and society	3	8	8	5	62	8.1744	2020.2	12.4	1.6349
27Action in Teacher Education	1	8	10	5	36	2.0608	2018	7.2	0.4122
28Teaching Statistics	5	1	1	5	29	9.4049	2021.8	5.8	1.881
29Education Sciences	5	9	10	5	25	2.8322	2021	5	0.5664
30Library Philosophy and Practice	2	2	2	5	4	0.3815	2020	0.8	0.0763

We physically searched for the website of the sources/journals to identify their disciplinary affiliations. We found that Education 24, Library and Information Science (7), Psychology (1), Computer Science/Information Technology (14), Social Sciences (7), Mathematics/Statistics (4) Multidisciplinary (3), General Science (1), Humanities (2) and Various Disciplines/ Interdisciplinary (10).

Citation by Authors (2005-2023)

For a minimum threshold of document per author placed at 3 and number of citations per author placed at 1, we obtained 2351 authors suitable for the analysis. Please see table xx and Figure xx. The table provides an overview of a citation study on data literacy by authors. It displays the crucial metrics such as cluster affiliation, linkages, total link strength, document count, citations received, and normalised citations. These metrics collectively offer valuable insights into the network dynamics and scholarly impact within the domain of data literacy research. Each author is cluster affiliation signifies their thematic association, while linkages and total link strength quantify the interconnectedness and prominence within their respective clusters. The document count reflects the scholarly output of each author, whereas the citations received and normalised citations shed light

on their influence and recognition within the scholarly community. This comprehensive analysis aids in understanding the landscape of data literacy research, identifying key contributors, and discerning patterns of scholarly dissemination and impact.

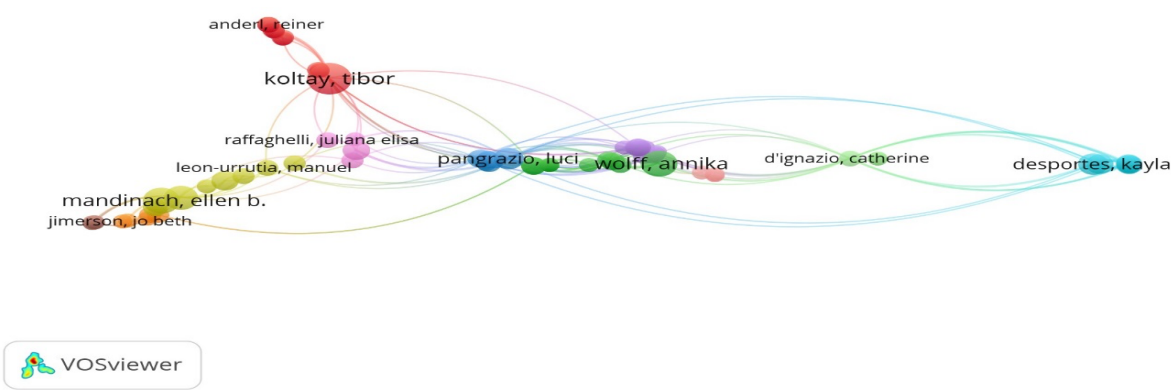


Figure 3. Citation by Authors (2005-2023).

In the realm of data literacy research, several authors stand out for their significant contributions, as evidenced by their citations, link strength, and publication output. Among these standout performers is Ellen B. Mandinach, whose work has garnered a remarkable total link strength of 10 and an impressive 776 citations across 31 documents, reflecting her profound influence and prolific output in the field. Kim Schildkamp also emerges as a prominent figure, with a substantial link strength of 27 and 510 citations distributed over 9 documents. Her research has evidently made a substantial impact within the community. Tabor Koltay distinguishes himself with 20 links and 343 citations spread across 15 documents, demonstrating a consistent and significant presence in data literacy scholarship. Luci Pangrazio, despite a lower number of documents, commands attention with an outstanding norm.citations score of 29.355, indicating high influence relative to

Table 3. Citation by Top Thirty Authors (2005-2023).

	Label	Clust er	Lin ks	Total link strength	Docume nts	Citatio ns	Norm. citations
1	Mandinach, Ellen B.	4	1	2	10	776	31.397
2	Schildkamp, Kim	4	17	27	9	510	36.300
3	Koltay, Tabor	1	20	25	15	343	16.873
4	Pangrazio, luci	3	33	58	7	304	29.355
5	Selwyn, Neil	3	20	29	5	245	21.001
6	Jimerson, Jo Beth	7	1	1	3	111	4.274
7	d'Ignazio, Catherine	11	21	31	4	97	9.573
8	Reeves, Todd D.	4	2	4	5	91	4.393
9	Wolff, Annika	2	11	13	10	88	5.626
10	Carmi, Elinor	3	2	4	4	84	9.710
11	Yates, Simeon J.	3	2	4	3	83	9.134



1 2	Markham, Annette N	11	1	1	3	74	6.923
1 3	Stewart, Bonnie	9	4	5	4	70	7.704
1 4	Cowie, Bronwen	7	10	14	4	54	2.914
1 5	Schultheis, Elizabeth H.	2	3	3	4	54	7.424
1 6	Acker, Amelia	5	11	14	5	53	2.860
1 7	Bowler, Leanne	5	11	14	5	53	2.860
1 8	Lee, victor r.	5	9	12	4	52	12.763
1 9	Schneider, rené	1	1	1	4	52	1.233
2 0	Wilkerson, michelle hoda	5	7	8	4	52	7.3697
2 1	Raffaghelli, juliana elisa	9	4	4	4	51	10.9321
2 2	Dasgupta, sayamindu	2	11	11	4	46	2.5879
2 3	Raffaghelli, juliana e.	9	7	12	6	45	6.7135
2 4	Chi, yu	5	11	15	3	42	2.1719
2 5	Jeng, wei	5	11	14	3	42	2.1719
2 6	Condon, patricia b.	1	6	10	3	37	8.1592
2 7	Knight, simon	5	6	9	4	34	4.7121
2 8	Mccosker, anthony	3	2	5	3	32	5.145
2 9	Manca, stefania	9	5	6	3	29	8.4752
3 0	Nguyen, dennis	4	4	5	4	27	8.2391

her publication output. Neil Selwyn and Jo Beth Jimerson also deserve recognition for their contributions. While Selwyn exhibits a strong influence with 245 citations across 5 documents, Jimerson is work, though fewer in number, is notable for its focused impact, as evidenced by 111

citations across only 3 documents. Catherine d’Ignazio, Todd D. Reeves, and Annika Wolff are among the notable performers, each with distinctive strengths in linkages and citations, reflecting their significant roles in advancing the discourse on data literacy.

3.3. Co-Citation of Cited Authors

Co-citation of cited authors refers to the phenomenon where two or more authors are cited together in the same paper by another author. There were 47361 cited authors on the subject matter. A total 118 met a threshold of 10 citations per author. Table 4 and Figure 4 represents a co-citation analysis of cited authors. Several authors stand out for their significant co-citation relationships, indicating the extent to which their work is cited together by other authors. Ellen B. Mandinach leads the pack with an impressive 156 links, reflecting a total link strength of 316.4462. Her work has been cited in conjunction with others 358 times, showcasing her pivotal role in shaping the discourse on data literacy. Kim Schildkamp closely follows with 137 links and a total link strength of 281.3285, reflecting 334 co-citations. Her research is evidently influential and frequently referenced alongside other prominent figures in the field.

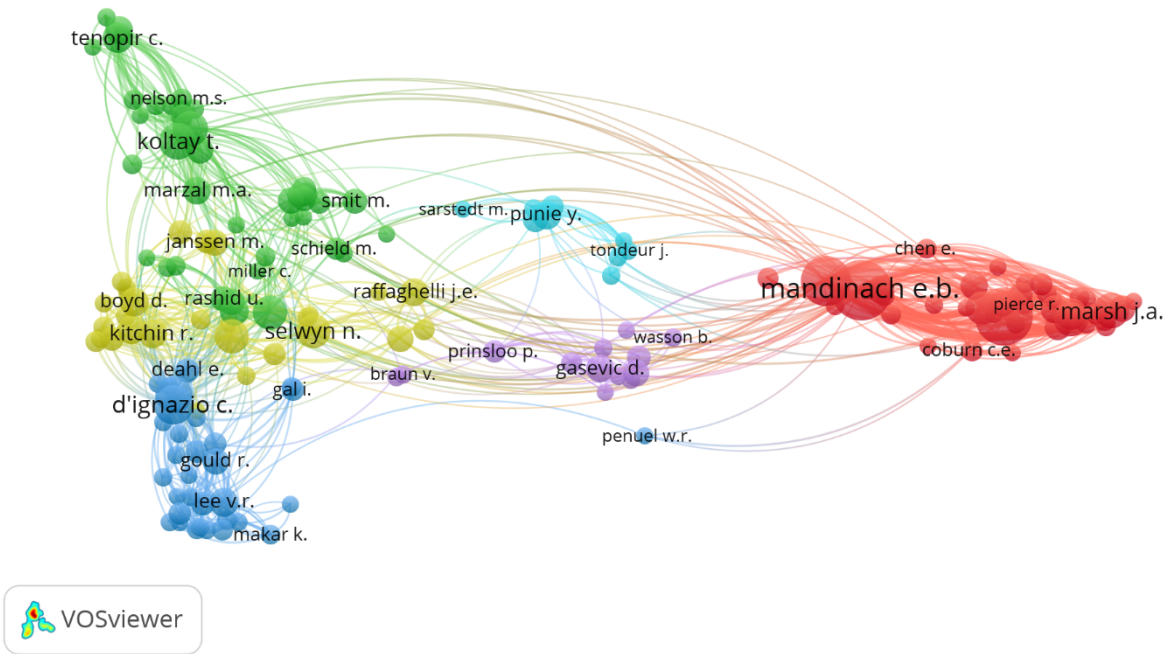


Figure 4. Co-citation of Cited Authors.

Table 4. Co-citation by Top Thirty Co-cited Authors.

	Label	Cluster	Links	Total link strength	Citations
1	Mandinach E.B.	1	156	316.4462	358
2	Schildkamp K.	1	137	281.3285	334
3	Gummer E.S.	1	156	233.3139	253
4	D’ignazio C.	3	130	145.3356	160
5	Bhargava R.	3	135	128.8378	136
6	Koltay T.	2	125	117.2822	132
7	Carlson J.	2	121	120.7855	129

8	Marsh J.A.	1	82	118.2453	124
9	Selwyn N.	4	136	110.5652	121
10	Pangrazio I.	4	135	110.1652	119
11	Datnow A.	1	90	106.8738	113
12	wolff a.	2	149	106.232	113
13	poortman c.l.	1	131	105.2675	111
14	wayman j.c.	1	100	93.5117	98
15	tenopir c.	2	57	79.4804	92
16	kitchin r.	4	124	64.6402	88
17	lee v.r.	3	88	74.6421	85
18	kortuem g.	2	138	79.1342	83
19	Boyd d.	4	130	73.4887	82
20	Crawford k.	4	137	72.8432	82
21	Punie y.	6	113	72.7417	81
22	Reeves t.d.	1	116	73.3235	79
23	Raffaghelli j.e.	4	134	67.4998	78
24	Allard s.	2	55	66.4713	77
25	Mandinach e.	1	133	75.2902	77
26	Gasevic d.	5	103	66.4572	76
27	Jimerson j.b.	1	105	74.1937	76
28	Marzal m.a.	2	119	72.0797	76
29	Gould r.	3	122	66.8782	73
30	Janssen m.	4	112	60.0328	72

E.S. Gummer also commands attention with 156 links and a total link strength of 233.3139, indicating substantial co-citation relationships resulting in 253 citations. Catherine d'Ignazio and R. Bhargava emerge as notable contributors, each with strong co-citation linkages and significant citation counts, demonstrating their impact and integration within the broader scholarly community. Tabor Koltay and J. Carlson are among the key players, with their work being frequently co-cited alongside others, reflecting their significant contributions to the field. Neil Selwyn and Luci Pangrazio also deserve recognition for their substantial co-citation relationships, indicative of their influence and interconnectedness within the research landscape. These authors

represent the core of the data literacy research community, characterised by their extensive co-citation relationships and the collective impact of their contributions on advancing the field is knowledge and understanding.

#### 3.4. Co-Citation by Cited Reference

There were 35410 cited references; at a threshold of 10 citations per cited reference yielded 64 items. Table 5 and Figure 5 show that Mandinach and Gummer is research, published in 2013, delves into implementing data literacy in educator preparation. This work, falling into Cluster 2, stands out with 38 co-citations, indicating its widespread reference by other scholars. The total link strength of 30 suggests strong connections with other works in the cluster. With 37 citations, it is evidently a seminal piece, underlining its pivotal role in discussions surrounding educator training in data literacy. D'Ignazio and Klein's "Data Feminism," a recent publication from 2020 within Cluster 1, has attracted attention with 28 co-citations and 24 total link strength. This work highlights the intersection

of gender studies and data literacy, a topic gaining prominence. Its 29 citations affirm its relevance and influence in shaping conversations about social perspectives in data literacy.

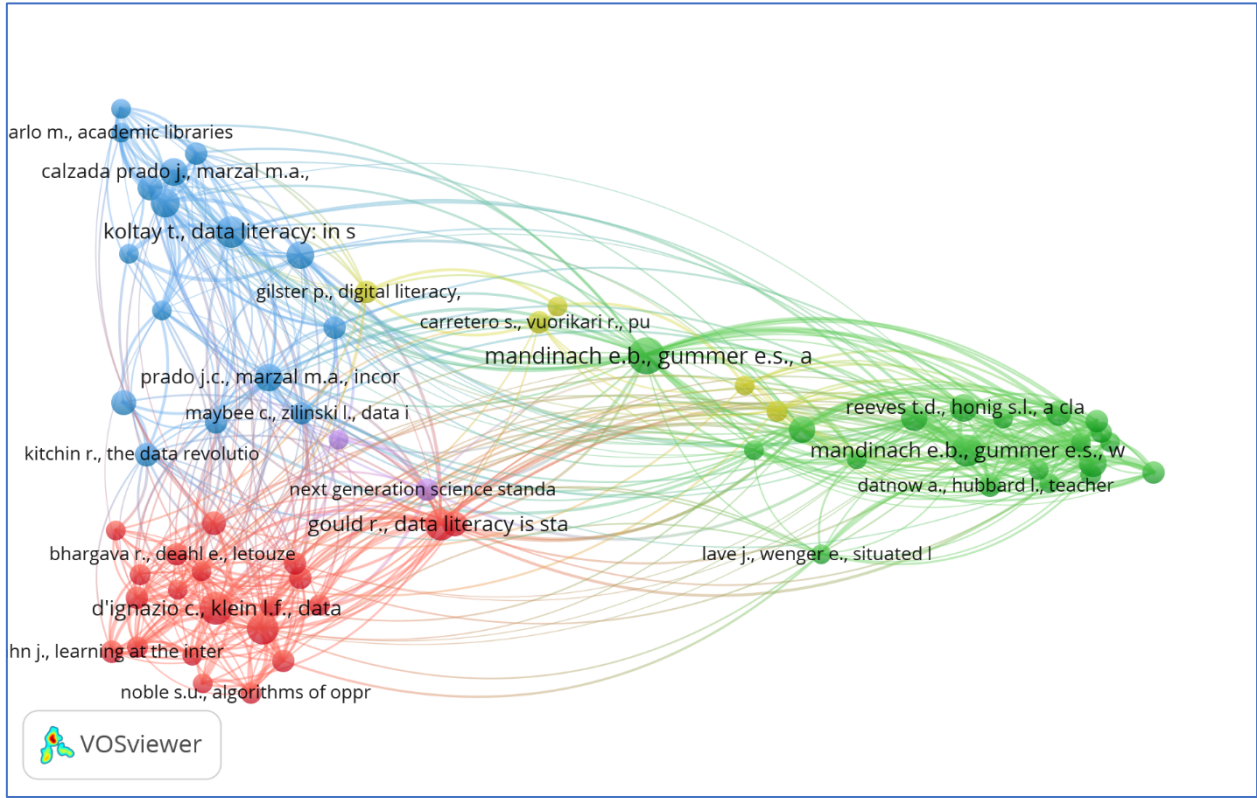


Figure 5. Co-citation by Cited References.

Table 5. Co-citation by Top Thirty Cited References.

	Label	Cl ust er	Li n ks	Total link strengt h	Cit atio ns
1	Mandinach E.B., Gummer E.S., A systemic view of implementing data literacy in educator preparation, <i>Educational Researcher</i> , 42, 1, pp. 30-37, (2013)	2	38	30	37
2	d'Ignazio C., Klein I.F., <i>Data Feminism</i> , (2020)	1	28	24	29
3	Koltay T., Data literacy: in search of a name and identity, <i>Journal of Documentation</i> , 71, 2, pp. 401-415, (2015)	3	24	20	27
4	Mandinach E.B., Gummer E.S., What does it mean for teachers to be data literate: laying out the skills, knowledge, and dispositions, teaching and teacher education, 60, pp. 366-376, (2016)	2	30	26	27
5	Gould R., Data literacy is statistical literacy, <i>statistics education research journal</i> , 16, 1, pp. 22-25, (2017)	1	31	23.3333	26
6	o'Neil c., weapons of math destruction: how big data increases inequality and threatens democracy, (2016)	1	30	22	26
7	Carlson J., Fosmire M., Miller C.C., Nelson M.S., determining data information literacy needs: a study of students and research faculty, <i>portal: libraries and the academy</i> , 11, 2, pp. 629-657, (2011)	3	23	19	22



8	Calzada prado j., marzal m.a., incorporating data literacy into information literacy programs: core competencies and contents, libri, 63, 2, pp. 123-134, (2013)	3	18	17	21
9	Mandinach e.b., a perfect time for data use: using data-driven decision making to inform practice, educational psychologist, 47, 2, pp. 71-85, (2012)	2	22	19	21
1 0	Prado j.c., marzal m.a., incorporating data literacy into information literacy programs: core competencies and contents, libri, 63, 2, pp. 123-134, (2013)	3	27	18	21
1 1	Koltay t., data literacy for researchers and data librarians, journal of librarianship and information science, 49, 1, pp. 3-14, (2017)	3	22	15	20
1 2	Reeves t.d., honig s.l., a classroom data literacy intervention for pre-service teachers, teaching and teacher education, 50, pp. 90-101, (2015)	2	28	20	20
1 3	Gummer e.s., mandinach e.b., building a conceptual framework for data literacy, teachers college record, 117, 4, pp. 1-22, (2015)	2	30	16	18
1 4	Kippers w.b., poortman c.l., schildkamp k., visscher a.j., data literacy: what do educators learn and struggle with during a data use intervention?, studies in educational evaluation, 56, pp. 21-31, (2018)	2	29	18	18
1 5	Braun v., clarke v., using thematic analysis in psychology, qualitative research in psychology, 3, 2, pp. 77-101, (2006)	1	19	11	17
1 6	Information literacy competency standards for higher education, (2000)	3	17	16	17
1 7	Janssen m., charalabidis y., zuiderwijk a., benefits, adoption barriers and myths of open data and open government, information systems management, 29, 4, pp. 258-268, (2012)	3	5	6	16
1 8	Marsh j.a., interventions promoting educators' use of data: research insights and gaps, teachers college record, 114, 11, pp. 1-48, (2012)	2	18	16	16
1 9	Boyd d., crawford k., critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon, information, communication & society, 15, 5, pp. 662-679, (2012)	1	20	11	15
2 0	Kitchin r., the data revolution: big data, open data, data infrastructures and their consequences, (2014)	3	17	11	15
2 1	Bhargava r., deahl e., letouze e., noonan a., sangokoya d., shoup n., beyond data literacy: reinventing community engagement and empowerment in the age of data, (2015)	1	17	11	14
2 2	Datnow a., hubbard l., teacher capacity for and beliefs about data-driven decision making: a literature review of international research, journal of educational change, 17, 1, pp. 7-28, (2016)	2	21	12	14
2 3	gilster p., digital literacy, (1997)	4	14	7	14
2 4	Kahn j., learning at the intersection of self and society: the family geobiography as a context for data science education, journal of the learning sciences, 29, 1, pp. 57-80, (2020)	1	13	11	14
2 5	Maybe c., zilinski l., data informed learning: a next phase data literacy framework for higher education, proceedings of the association for information science and technology, 52, 1, pp. 1-4, (2015)	3	20	11	14

2 6	Means b., padilla c., gallagher l., use of education data at the local level: from accountability to instructional improvement, (2010)	2	15	13	14
2 7	Next generation science standards: for states, by states, (2013)	5	11	9	14
2 8	Stornaiuolo a., authoring data stories in a media makerspace: adolescents developing critical data literacies, <i>Journal of The Learning Sciences</i> , 29, 1, pp. 81-103, (2020)	1	26	12	14
2 9	Zuboff is., the age of surveillance capitalism: the fight for a human future at the new frontier of power, (2019)	1	18	9	14
3 0	Benjamin r., race after technology: abolitionist tools for the new jim code, (2019)	1	18	13	13

Koltay is article from 2015 explores the concept of data literacy identity, contributing to Cluster 3. With 24 co-citations and 20 total link strength, it reflects a robust connection within the cluster. Having garnered 27 citations, this work is evidently influential in discussions about defining and understanding data literacy. Gould is exploration of data literacy as statistical literacy, published in 2017 and affiliated with Cluster 1, draws significant attention with 31 co-citations and approximately 23.33 total link strength. With 26 citations, it underscores the critical relationship between data literacy and statistical understanding, highlighting its importance in educational discourse.

O'Neil is book "Weapons of Math Destruction" from 2016, situated in Cluster 1, resonates strongly with 30 co-citations and a total link strength of 22. With 26 citations, it sheds light on the societal implications of big data, emphasizing the ethical considerations inherent in data usage. This work contributes significantly to discussions about the broader societal impacts of data literacy. Carlson et al. is study on determining data information literacy needs, published in 2011 and associated with Cluster 3, garners attention with 23 co-citations and 19 total link strength. Its 22 citations underline its relevance in understanding the data literacy requirements of students and research faculty, particularly in academic settings. Calzada Prado and Marzal is work on incorporating data literacy into information literacy programs, appearing in Cluster 3 and published in 2013, is cocited with 18 other works, indicating its integration within the cluster. With 17 total link strength and 21 citations, this study contributes significantly to discussions about integrating data literacy into broader educational frameworks.

Mandinach is research on using data-driven decision making to inform practice, published in 2012 and associated with Cluster 2, draws attention with 22 co-citations and 19 total link strength. With 21 citations, it underscores the importance of leveraging data in educational decision-making processes, highlighting its relevance for educators and policymakers. Prado and Marzal is work on incorporating data literacy into information literacy programs, published in 2013 and associated with Cluster 3, stands out with 27 co-citations, 18 total link strength, and 21 citations. This indicates its pivotal role in shaping discussions about integrating data literacy into educational curricula and programs.

3.5. Co-Citation by Cited Sources

A total of 17706 cited sources were identified, and 173 sources met the threshold of 20 citations per source. Table 6 at Figure 6 speak to the detail. Teaching and Teacher Education: Belongs to cluster 5, with 130 links, a total link strength of 245.704, and 286 citations, indicating its significance in educational research and pedagogy. Teachers College Record: Also part of cluster 5, with 145 links, a total link strength of 239.297, and 279 citations, highlighting its influence in educational policy and practice. Big Data & Society: Falls under cluster 2, with 144 links, a total link strength of 157.076, and 197 citations, suggesting its importance in the intersection of big data and societal impacts. Educational Researcher: Belongs to cluster 5, with 161 links, a total link strength of 181.1752, and 188 citations, indicating its role as a leading publication in educational research. Journal of the Learning

Sciences: Part of cluster 3, with 122 links, a total link strength of 144.878, and 177 citations, emphasizing its significance in the study of learning processes and educational technologies.

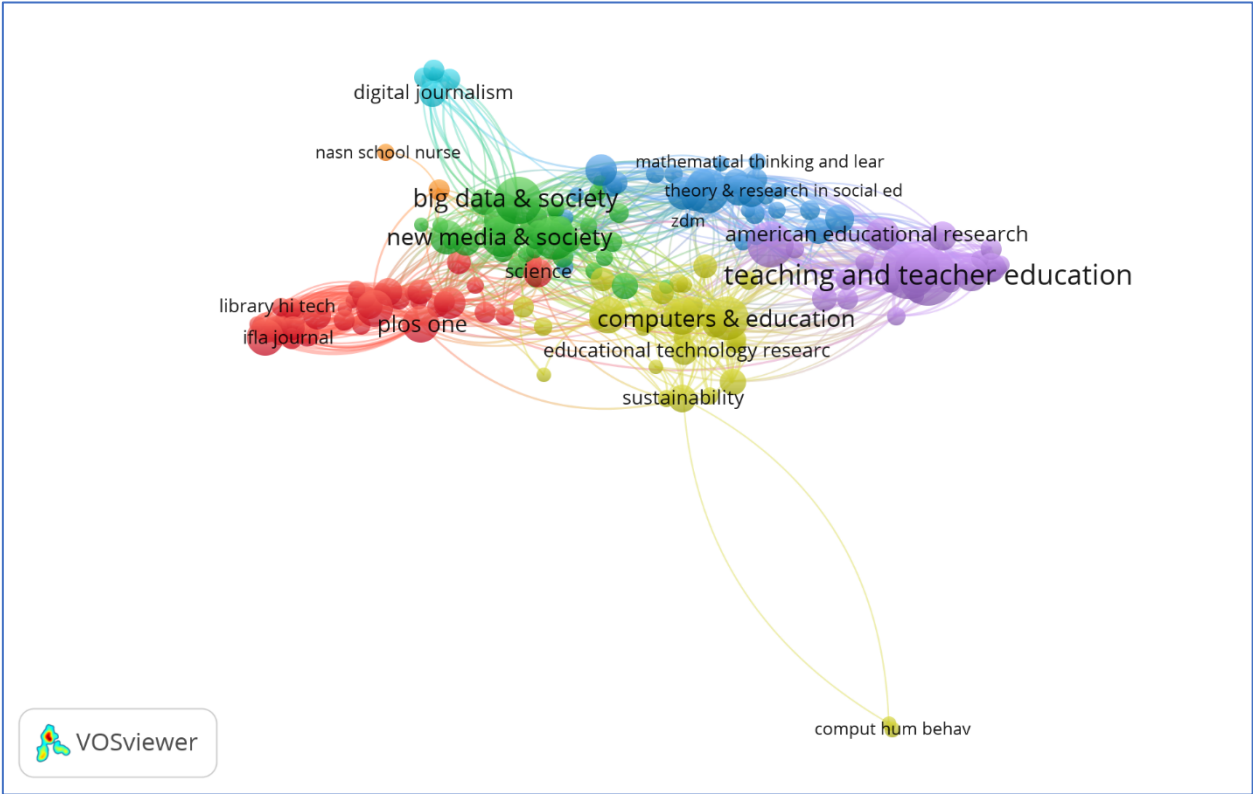


Figure 6. Co-citation by cited sources.

Table 6. Co-citation by cited sources.

	Label	Cluste r	Link s	Total link strength	Citation s
1	Teaching and Teacher Education	5	130	245.7039	286
2	Teachers College Record	5	145	239.2967	279
3	Big Data & Society	2	144	157.0756	197
4	Educational Researcher	5	161	181.1752	188
5	Journal of the Learning Sciences	3	122	144.8794	177
6	Studies in Educational Evaluation	5	137	144.1676	174
7	Computers & Education	4	147	141.4396	164
8	The Journal of Community Informatics	2	154	143.8328	162
9	New Media & Society	2	153	131.1638	145
10	Plos One	1	131	93.2724	123
11	British Journal of Educational Technology	4	148	109.6554	121
12	Computers in Human Behavior	4	147	109.5222	120

1 3	Journal of Documentation	1	144	111.9254	120
1 4	Journal of Escience Librarianship	1	81	90.4297	117
1 5	Government Information Quarterly	2	101	64.2789	106
1 6	School Effectiveness and School Improvement	5	80	85.5202	101
1 7	Statistics Education Research Journal	3	135	81.0099	95
1 8	American Educational Research Journal	5	118	88.2994	92
1 9	Libri	1	147	83.3571	90
2 0	IEEE Transactions on Visualization and Computer Graphics	3	104	60.8033	89
2 1	Information, Communication & Society	2	141	81.5082	89
2 2	Cognition And Instruction	3	111	77.2383	83
2 3	International Journal of Digital Curation	1	75	69.7547	79
2 4	American Journal of Education	5	90	68.1408	75
2 5	Journal of Research in Science Teaching	3	107	63.4963	74
2 6	Science	1	134	65.8598	74
2 7	Educational Psychologist	3	106	70.7714	73
2 8	Journal of Teacher Education	5	83	67.297	72
2 9	Educational Studies in Mathematics	3	115	58.116	71
3 0	First Monday	2	122	64.4278	70

Studies in Educational Evaluation: Also associated with cluster 5, with 137 links, a total link strength of 144.168, and 174 citations, indicating its role in evaluating educational practices and policies. Computers & Education: Falls under cluster 4, with 147 links, a total link strength of 141.4396, and 164 citations, highlighting its importance in the intersection of technology and education. The Journal of Community Informatics: Belongs to cluster 2, with 154 links, a total link strength of 143.834 and 162 citations, suggesting its role in community-based research and



informatics. *New Media & Society*: Also part of cluster 2, with 153 links, a total link strength of 131.164, and 145 citations, indicating its significance in the study of new media and digital cultures.

*PLOS ONE* is in cluster 1, with 131 links, a total link strength of 93.272 and 123 citations, highlighting its role as an open-access multidisciplinary journal. *British Journal of Educational Technology* belongs to cluster 4, with 148 links, a total link strength of 109.655, and 121 citations, emphasizing its importance in educational technology research and practice. *Computers in Human Behaviour* is in cluster 4, with 147 links, a total link strength of 109.522, and 120 citations, indicating its significance in the study of human-computer interaction in educational contexts. *Journal of Documentation*: Associated with cluster 1, with 144 links, a total link strength of 111.925, and 120 citations, highlighting its role in the study of information science and documentation. *Journal of eScience Librarianship*: Also part of cluster 1, with 81 links, a total link strength of 90.430 and 117 citations, indicating its significance in the field of library and information science, particularly in *eScience*. *Government Information Quarterly* falls under cluster 2, with 101 links, a total link strength of 64.279, and 106 citations, highlighting its role in the study of government information policies and practices.

*School Effectiveness and School Improvement* belongs to cluster 5, with 80 links, a total link strength of 85.520, and 101 citations, indicating its significance in research on school effectiveness and improvement strategies. *Statistics Education Research Journal* is part of cluster 3, with 135 links, a total link strength of 81.0099, and 95 citations, emphasizing its role in the advancement of statistical education research. *American Educational Research Journal* also associated with cluster 5, with 118 links, a total link strength of 88.299, and 92 citations, highlighting its significance in educational research across various domains. *Libri* falls under cluster 1, with 147 links, a total link strength of 83.357, and 90 citations, indicating its importance in the study of libraries and information science. *IEEE Transactions on Visualization and Computer Graphics* belongs to cluster 3, with 104 links, a total link strength of 60.803, and 89 citations, suggesting its role in advancing research in visualization and computer graphics in education.

*Information, Communication & Society*: Part of cluster 2, with 141 links, a total link strength of 81.508, and 89 citations, highlighting its importance in the study of information and communication technologies in society. *Cognition and Instruction*: Falls under cluster 3, with 111 links, a total link strength of 77.238, and 83 citations, emphasizing its role in the advancement of cognitive science and instructional design. *International Journal of Digital Curation* belongs to cluster 1, with 75 links, a total link strength of 69.755, and 79 citations, indicating its significance in the study of digital curation practices and standards. *American Journal of Education*: Also associated with cluster 5, with 90 links, a total link strength of 68.1408, and 75 citations,

highlighting its importance in educational policy and practice. *Journal of Research in Science Teaching* falls under cluster 3, with 107 links, a total link strength of 63.496, and 74 citations, emphasizing its role in the advancement of science education research. *Science*: Part of cluster 1, with 134 links, a total link strength of 65.860, and 74 citations, indicating its significance as a leading scientific source

#### 4. Discussion of Findings

This study undertook a citation analysis of research on data literacy by documents, sources and authors; and a co-citation of cited authors, cited sources and cited references based on data collected from Scopus and analysed with Vosviewer. The analysis of citation by documents reveals a picture of the scholarly landscape surrounding data literacy. Initially, the study identifies a substantial volume of literature produced on the subject, totaling 997 documents over the observed period, with an average of 52 documents per year. However, to ensure a focus on significant contributions, a criterion of a minimum of 10 citations per document is applied. This filters the dataset down to 205 documents that meet the threshold, indicating a subset of scholarly works that have garnered substantial attention within the academic community. Further refinement of the dataset reveals that out of the 205 documents meeting the citation threshold, only 81 are interlinked with other documents. These interconnected documents form the basis for the subsequent analysis. This

observation suggests that while numerous documents exist on data literacy, only a fraction of them are deeply engaged with and cited within the scholarly discourse, indicating their centrality in shaping the conversation on this topic (Small 1973, Small and Kalavan 2013)

Within this subset of interconnected documents, a remarkable average of 34 citations per document is observed, totaling 2768 citations across the 81 papers. This high citation rate underscores the significance and impact of these select works within the field of data literacy. Notably, Prado's (2013) paper, "Incorporating Data Literacy into Information Literacy Programmes: Core Competences and Content," emerges as a standout, amassing 161 citations, surpassing all other papers in terms of citation count. Prado's work is noteworthy for its emphasis on the growing importance of data in contemporary society and its advocacy for the integration of data literacy into information programs. By proposing a comprehensive framework of core competencies, Prado provides not only a roadmap for educators and policymakers but also a foundation for further research and development in the field.

The analysis of citation by sources provides insight into the key journals, conference proceedings and other sources shaping the discourse on data literacy. Among the top sources, *the ACM International Conference Proceeding Series* emerges as the most prolific, publishing 47 documents on the subject and accumulating 260 citations. This indicates the significant contribution of academic conferences in disseminating research and fostering scholarly dialogue within the field. Additionally, the study highlights the *Teachers College Record* as a standout source, despite its relatively low publication output of only six articles. TCR's remarkable citation count of 429 underscores its substantial impact within the field of education and its role as a leading platform for scholarly discourse on data literacy. The disciplinary affiliations of these sources vary, encompassing diverse fields such as Education, Library and Information Science, Psychology, Computer Science/Information Technology, Social Sciences, Mathematics/Statistics, Humanities, and Multidisciplinary subjects. This interdisciplinary nature of data literacy sources underscores the broad relevance and significance of the area across various domains of knowledge.

The study also examined authors' citations during 2005 to 2023. For the analysis, a minimum threshold of three documents per author and one citation per author was set, resulting in 2351 authors suitable for examination. Each author's cluster affiliation indicates their thematic association within the field, while linkages and total link strength quantify their interconnectedness and prominence within their respective clusters. The document count reflects the scholarly output of each author, whereas citations received and normalized citations shed light on their influence and recognition within the scholarly community. The analysis identifies standout authors based on their citation counts, link strength, and publication output. Notable figures include Ellen B. Mandinach, Kim Schildkamp, Tabor Koltay, and Luci Pangrazio, among others. These authors have demonstrated significant contributions to the field, as evidenced by their citations and linkages. Additionally, the analysis highlights the network dynamics within the domain of data literacy research, providing insights into the patterns of scholarly dissemination and impact. It aids in understanding the landscape of data literacy research, identifying key contributors, and discerning thematic trends and patterns of scholarly influence.

The analysis of Co-citation of Cited Authors reveals several prominent authors whose works are extensively cited together, indicating the interconnectedness of their ideas and the influence of their contributions. Ellen B. Mandinach, with 156 links and a total link strength of 316.4462, emerges as a central figure in shaping discussions on data literacy. Her research, particularly on implementing data literacy in educator preparation, has garnered significant attention, as reflected in the high number of co-citations. Similarly, Kim Schildkamp's and E.S. Gummer's works are frequently cited alongside Mandinach's, highlighting their collective impact on advancing the field. Furthermore, authors like Catherine d'Ignazio and R. Bhargava, Tabor Koltay, and J. Carlson also feature prominently in co-citation networks, underscoring their significant contributions to the discourse on data literacy. These authors represent a diverse range of perspectives, from exploring the intersection of gender studies with data literacy to investigating the practical implications of data-driven decision-making in educational settings.

The analysis of co-cited references sheds light on seminal works that have shaped the trajectory of data literacy research. Mandinach and Gummer's research on implementing data literacy in educator preparation stands out, with 38 co-citations and a total link strength of 30, indicating its foundational role in the field. Similarly, d'Ignazio and Klein's exploration of "Data Feminism" and Koltay's examination of data literacy identity contribute fresh insights, reflecting the evolving nature of data literacy discourse. The analysis highlights the interdisciplinary nature of data literacy research, with works spanning education, sociology, and information science. Gould's exploration of data literacy as statistical literacy and O'Neil's examination of the societal implications of big data underscore the multifaceted dimensions of the field. These co-cited references not only contribute to theoretical frameworks but also offer practical insights for educators, policymakers, and researchers grappling with the challenges of data literacy.

The co-citation analysis of cited sources reveals key publications that serve as pillars in data literacy research. *Teaching and Teacher Education* and *Teachers College Record* emerge as central hubs, showcasing their pivotal role in disseminating scholarly discourse on data literacy. These journals provide platforms for researchers to exchange ideas, disseminate findings, and engage in critical debates shaping the field. Additionally, interdisciplinary journals such as *Big Data & Society* and *New Media & Society* highlight the intersection of data literacy with broader societal trends, including the impact of technology on information dissemination and digital cultures. By analyzing the co-citation patterns of these sources, researchers can gain insights into the evolving landscape of data literacy research and identify emerging trends and themes that warrant further investigation (Trujillo and Long 2018).

## 5. Conclusions

A significant volume of literature on data literacy exists, with 997 documents produced over the observed period. A stringent criterion of a minimum of 10 citations per document was applied to focus on significant contributions, resulting in a subset of 205 documents meeting the threshold. Only 81 out of the 205 documents were found to be interconnected with other documents, suggesting their centrality in shaping the scholarly discourse on data literacy. The ACM International Conference Proceeding Series emerged as the most prolific source, publishing 47 documents on data literacy and accumulating 260 citations. Teachers College Record (TCR) stood out as a leading source despite publishing only six articles, with a remarkable citation count of 429, highlighting its substantial impact within the field of education.

Ellen B. Mandinach emerged as a central figure in data literacy discussions, with her research on implementing data literacy in educator preparation receiving significant attention. Authors like Catherine d'Ignazio, R. Bhargava, Tabor Koltay, and J. Carlson also featured prominently in co-citation networks, indicating their significant contributions to the discourse on data literacy. Seminal works by Mandinach and Gummer on implementing data literacy in educator preparation stood out, indicating their foundational role in the field. The interdisciplinary nature of data literacy research was evident, with works spanning education, sociology, and information science, contributing to both theoretical frameworks and practical insights. *Teaching and Teacher Education* and *Teachers College Record* emerged as central hubs in disseminating scholarly discourse on data literacy, showcasing their pivotal role in shaping the field. Interdisciplinary journals such as *Big Data & Society* and *New Media & Society* highlighted the intersection of data literacy with broader societal trends, providing insights into the evolving landscape of research in the field.

The focus of the study on documents with a specified minimum number of citations per document may introduce selection bias by favoring widely cited works over potentially valuable but less recognized contributions. Additionally, limitations in the dataset's scope could lead to an incomplete representation of the scholarly landscape. The analysis' temporal scope and reliance on predefined criteria for identifying relationships between documents may further introduce biases, impacting the interpretation of findings and the generalizability of results beyond the specific dataset and methodology employed.

## 6. Implications for Policy and Practice

*Informing Educational Policies:* By recognizing the significant impact of data literacy in contemporary society, policymakers can use insights from this study to advocate for the integration of data literacy into educational curricula at all levels. Also, by acknowledging the pivotal role of data literacy in preparing students for the digital age, policymakers can prioritize initiatives aimed at enhancing data literacy skills among learners.

*Professional Development for Educators:* Policymakers and educational institutions can use the findings of this study to design targeted professional development programs for teachers, equipping them with the necessary skills and knowledge to integrate data literacy into their teaching practices effectively. By investing in teacher training and support, policymakers can ensure that educators are equipped to address the evolving demands of data-driven education.

*Promoting Interdisciplinary Collaboration:* Given the interdisciplinary nature of data literacy research, policymakers can encourage collaboration across various fields, including education, information science, psychology, and computer science. By fostering interdisciplinary partnerships and initiatives, policymakers can facilitate the exchange of ideas, resources, and best practices, ultimately enhancing the quality and effectiveness of data literacy initiatives.

*Supporting Open Access and Collaboration:* Policymakers can support initiatives aimed at promoting open access to scholarly literature and fostering collaboration among researchers. By facilitating access to research findings and encouraging collaboration among scholars, policymakers can accelerate the pace of innovation and knowledge dissemination in the field of data literacy.

## References

1. Abramo, G., D'Angelo, C.A., & Murgia, G. (2019). Evaluating research: From informed peer review to bibliometrics. *Research Evaluation*, 28(3): 174-181.
2. Aldboush, H.H.H., and Ferdous, M. (2023). "Building Trust in Fintech: An Analysis of Ethical and Privacy Considerations in the Intersection of Big Data, AI, and Customer Trust" *International Journal of Financial Studies* 11(3): 90. <https://doi.org/10.3390/ijfs11030090>.
3. Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2009). H-index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics*, 3(4), 273-289.
4. Asubiaro, T. & Elueze, I. (2022). Evidence-Based Biomedical Research in Sub-Saharan Africa: How Library and Information Science Professionals Contribute to Systematic Reviews and Meta-Analyses. *Journal of the Medical Library Association* DOI: 110. 72-80. 10.5195/jmla.2022.1249.
5. Baas, J. & Schotten, M. & Plume, A. & Côté, G. & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies* 1. 1-10. 10.1162/qss\_a\_00019.
6. Bartholo, T.L., Koslinski, M.C., Tymms, P.B., & Castro, D.L. (2022). Learning loss and learning inequality during the Covid-19 pandemic. *Ensaio: Avaliação e Políticas Públicas em Educação*. 31 (119) • <https://doi.org/10.1590/S0104-40362022003003776>.
7. Beel, J. and B. Gipp. (2009). Google Scholar's Ranking Algorithm: The Impact of Articles' Age (An Empirical Study). In S. Latifi, ed., *Proceedings of the 6th International Conference on Information Technology: New Generations (ITNG'09)*: 160-164, Las Vegas: IEEE. Available on <http://www.sciplore.org>.
8. Bordons, M., Fernández, M., & Gómez, I. (2002). Advantages and limitations in the use of impact factor measures for the assessment of research performance. *Scientometrics*, 53(2), 195-206.
9. Bornmann, L., & Leydesdorff, L. (2014). *Scientometrics in a changing research landscape*. Springer.
10. Burnham, Judy. (2006). Scopus database: A review. *Biomedical Digital Libraries* 3(1). DOI: 10.1186/1742-5581-3-1.
11. Debruyne, C., Grehan, L., Hurley, M., Kearns, A.M., & O'Neill, C. (2022). One year of DALIDA Data Literacy Workshops for Adults: A Report. Companion Proceedings of the Web Conference 2022. (WWW '22 Companion), April 25-29, 2022, Virtual Event, Lyon, France. ACM, New York, NY, USA, 5 pages. DOI:10.1145/3487553.3524694
12. D'Ignazio, C. (2017). Creative data literacy: Bridging the gap between the data-haves and data-have-nots. *Information Design Journal* 23. 6-18. 10.1075/idj.23.1.03dig.
13. D'Ignazio, C. and Klein, L.F. (2020). *Data Feminism*, The MIT Press, 328pp
14. Garfield E (1972 Nov 3). Citation analysis as a tool in journal evaluation. *Science*. 178(4060):471-9. doi: 10.1126/science.178.4060.471. PMID: 5079701.
15. Gehrke, M. & Kistler, T. & Lübke, K. & Markgraf, N. & Krol, B. & Sauer, S. (2021). Statistics education from a data-centric perspective. *Teaching Statistics* 43, S201-S215. 10.1111/test.12264.



16. Ghodoosi, B., West, T., Li, Q., Torrisi-Steele, G., & Dey, S. (2023). A systematic literature review of data literacy education. *Journal of Business & Finance Librarianship*, 28(2), 112–127. <https://doi.org/10.1080/08963568.2023.2171552>.
17. Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal* 16. 22-25. 10.52041/serj.v16i1.209.
18. Heiser R.E.; Dello Stritto M.E.; Brown A.S.; Croft B. (2023). Amplifying Student and Administrator Perspectives on Equity and Bias in Learning Analytics: Alone Together in Higher Education. *Journal of Learning Analytics* 10(1):8-23. DOI: 10.18608/jla.2023.7775.
19. Hjørland, B. & Nicolaisen, J. (2005). Bradford's Law of Scattering: Ambiguities in the Concept of "Subject". *Lecture Notes in Computer Science* 3507. 96-106. 10.1007/11495222\_9.
20. Larivière, V., Kiermer, V., MacCallum, C. J., McNutt, M., Patterson, M., Pulverer, B., Swaminathan, S. (2016). A simple proposal for the publication of journal citation distributions. *BioRxiv*, 062109.
21. Levitan, L.C., & Verhulst, B. (2016). Conformity in groups: The effects of others' views on expressed attitudes and attitude change. *Political Behavior*, 38(2), 277–315. [Doi:10.1007/s11109-015-9312-x](https://doi.org/10.1007/s11109-015-9312-x).
22. Loukides M, Mason H, and Patil D.J. (2018). *Ethics and Data Science*, O'Reilly
23. Mandinach E.B. and Gummer E.S. (2016) Data Literacy for Educators: Making It Count in Teacher Preparation and Practice. *Harvard Educational Review* 86 (4): 607–610. DOI: 10.17763/1943-5045-86.4.607a.
24. Moed, H. (2005). Citation Analysis in Research Evaluation. 10.1007/1-4020-3714-7.
25. Morrow, J. (2021). *Be Data Literate: The Data Literacy Skills Everyone Needs To Succeed* 1st Edition, Kindle Edition.
26. Nissenbaum, H. (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Bibliovault OAI Repository, the University of Chicago Press.
27. Nwagwu, W.E. (2010). Cybernating the academe: Centralised Scholarly Ranking and Visibility of Scholars in the Developed World. *Journal of Information Science* Vol. 36. No. 2: 228-241.
28. Nwagwu W.E. (2023). Nature and characteristics of global attention to research on article processing charges, *The Journal of Academic Librarianship*, Volume 49, Issue 6, 102808, Doi: 10.1016/j.acalib.2023.102808.
29. Nwagwu, W. E. (2024). Mapping the field of global research on data literacy: Key and emerging issues and the library connection. *IFLA Journal*, 0(0). <https://doi.org/10.1177/03400352241257669>
30. Olszewski, J., & Abukhadier, M. (2023). BCStat: A Sustainable Approach to "CollaborationStat". *State and Local Government Review*, 55(3), 187-191. <https://doi.org/10.1177/0160323X231200219>.
31. Osasona, F.; Amoo, O.; Atadoga, A.; Abrahams, T.; Farayola, O. & Ayinla, B. (2024). Reviewing the ethical implications of AI in decision making processes. *International Journal of Management & Entrepreneurship Research* 6. 322-335. 10.51594/ijmer.v6i2.773.
32. Phadkule, P. (2022). Data Literacy for Public Policymaking – Lessons from Covid-19 Crisis in India. *Journal of Asian Public Policy*, 1–15. <https://doi.org/10.1080/17516234.2022.2143085>
33. Pins, D., Jakobi, T., Boden, A., Alizadeh, F., & Wulf, V. (2021). Alexa, We Need to Talk: A Data Literacy Approach on Voice Assistants. *Proceedings of the 2021 ACM Designing Interactive Systems Conference*.
34. Santos, C. & Pedro, N. & Mattar, J. (2021). Digital competence of higher education professors: analysis of academic and institutional factors. *Obra Digital* 67-92. 10.25029/od.2021.311.21.
35. Small H (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *JASIST* July/August 24(4): 265-269. DOI:10.1002/asi.4630240406.
36. Small, H.G., & Klavans, R. (2013). Identifying Scientific Breakthroughs by Combining Co-citation Analysis and Citation Context. In: Ed Noyons, Patrick Ngulube, Jacqueline Leta (Eds), *Proceedings of ISSI 2011 – the 13th International Conference of the International Society for Scientometrics and Informetrics*, Vol. 2, Durban, South Africa, 4-7 July 2011, ISSI, Leiden University and University of Zululand, 2011, 783-793.
37. Stanton, N. A., Salmon, P. M., Walker, G. H., Salas, E., & Hancock, P. A. (2017). State-of-science: situation awareness in individuals, teams and systems. *Ergonomics* 60(4), 449–466. [DOI:10.1080/00140139.2017.1278796](https://doi.org/10.1080/00140139.2017.1278796).
38. Strong, D; Lee, Y & Wang, R. (2002). Data Quality in Context. *Communications of the ACM* 40. DOI: 10.1145/253769.253804.
39. O'Connor A. (2021, Oct. 4). Data Literacy: What It Is and Why It Matters, *Enterprise Times*.
40. Trujillo, C.M., and Long, T.M. (2018). Document co-citation analysis to enhance transdisciplinary research. *Science Advances* 34(1):e1701130. doi: 10.1126/sciadv.1701130. PMID: 29308433.
41. Vance, E.A., Glimp, D.R., Pieplow, N.D., Garrity, J.M., & Melbourne, B.A. (2022). Integrating the humanities into data science education. *Statistics Education Research Journal* 21(2). Available at <https://par.nsf.gov/biblio/10357864-integrating-humanities-data-science-education>, accessed May 8 2024.
42. Waltman, L., & van Eck, N.J. (2012). The inconsistency of the h-index. *Journal of the American Society for Information Science and Technology*, 63(2), 406-415.

43. Wang, J., Veugelers, R., & Stephan, P. (2019). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy*, 48(1), 140-153.
44. Wilkerson, M., Lanouette, K.A., & Shareff, R. (2021). Exploring variability during data preparation: a way to connect data, chance, and context when working with complex public datasets. *Mathematical Thinking and Learning*, 24, 312 - 330.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.