



Scholarly Literature in Digital Humanities: A Comparative Analysis of Open Access and Non-Open Access Publications

Sandra M K

Research Scholar

Department of Library and Information Science

Tumkur University, Karnataka. India

sandrasaj98@gmail.com

Rupesh Kumar A

Assistant Professor

Department of Library and Information Science

Tumkur University, Karnataka. India

a.rupeshkumar@gmail.com

ABSTRACT

Digital Humanities (DH), as a research area, has attracted the attention of scholars across the globe. The present study attempts to trace the publication and citation trends in DH literature with a comparative analysis of Open Access (OA) and Non-Open Access (Non-OA) publications using a dataset of 3,731 publications collected from Scopus. DH literature is characterized by a steady growth from 1971 to 2024. The highest number of publications have been produced during 2022 (475). Although a consistent increase in OA publications can be witnessed since 2009, a significant portion of DH literature (65%) is non-open access. In terms of citations, OA publications show greater potential. Mann Whitney U test shows that the mean rank of citation counts for OA publications (2036.47) is higher than that of non-OA publications (1772.56). The result is statistically significant ($p\text{-value}=.000$). Green OA is the most preferred OA channel, with 27% of OA publications. USA is the top contributor to DH literature with 23% of publications, while Belgium tops the list with 62% open access publication output. Computer Science (48%) is the most predominant subject area in DH. Articles and conference papers constitute 86% of the total literature, indicating authors' preference for scholarly communication.

Keywords: Digital Humanities, Open Access, Non-open Access, Comparative analysis, Publication trends, Citation Trends

Received: 1 September 2024, Revised 6 November 2024, Accepted 18 November 2024

Copyright: with Authors

1. Introduction

Digital Humanities (DH) is evolving with the intersection and collaboration of multiple subject areas. It originates from “humanities computing,” which deals with using computer technology in research, teaching, and learning in the subject areas associated with humanities disciplines like history, philosophy, language, linguistics, and other related fields. Its scope and coverage cannot be easily defined as it is widely scattered among different subject areas. The history of digital humanities can be traced back to 1946, when a computational text analysis was used to prepare *Index Thomisticus* (Smith, 2002). With the emergence of “digital humanities,” traditional research methods in the social science and humanities disciplines are replaced with digital tools and techniques, which enable researchers and investigators to reach a more precise visualization of data.

Research trends and the impact of any novel topic are usually measured using bibliometric analysis. As an emerging trend, digital humanists track the growth of DH from different aspects like its interdisciplinarity, collaborations and intersections, affiliation, publication count, etc. The present study analyses the growth of openaccess (OA) and non-openaccess (non-OA) digital humanities literature. Many articles are disseminated through OA journals (Poulin and Tomaszewski, 2014). Knowledge can be created by drawing inferences from what is known. So, it is important to provide access to already existing information to enhance further research and practically implement the research output. A comparative study of the citation count is also done between the OA and non-OA DH research outputs to understand the acceptance of DH.

2. Literature Review

Digital Humanities, as a research discipline, has attracted the attention of scholars across the globe. This is evident by the growing number of publications on DH. Several studies have pointed out the increasing number of publications in digital humanities research (Tang et.al., 2017; Wang, 2018; Basak, 2022). This may be attributed to the transformative nature of the subject itself. An early study by Dalbello (2011) explored digital humanities’ historical development and transformative impact in different disciplines through a conceptual analysis of textual data published in the 1980s. Later, Liu (2013) investigated the potential of digital humanities to transform humanities research using computational technologies and discovered the complexities and evolving nature of the subject. The interdisciplinarity of DH evoked the interest of researchers from various disciplines to examine the growth of scholarly literature, authorship patterns, variations of regional research outputs, subject inclinations, etc. using bibliometric analysis. Luhmann and Burghardt (2021) studied the interdisciplinarity in digital humanities and argued that it was a discipline in its own right.

Gupta and Chakravarty (2021) found social sciences to be the predominant subject area of DH research. This was supported by a later study (Chansanam et.al., 2022) which revealed that computer science and arts and humanities were the next preferred areas of research in DH. Interestingly, similar findings were observed in analysing scholarly literature on digital humanities and libraries. Social Science, Computer Science, and Arts and Humanities were the top three subject areas contributing to research in digital humanities and libraries. Further, the USA, China, and UK contributed the highest number of documents.(Makwana & Gadhavi, 2023)

Open-access publications increase visibility, accessibility, and readership for scientific communication.(Solomon, 2014.)Thus, many researchers have studied the ability of open-access publications to attract citations. Several studies have found that the citation advantage of open-access publications is greater than that of non-open-access publications.(Chua et al., 2017; Dorta-González & Dorta-González, 2023; Koler-Povh et al., 2014)

An exhaustive literature search and review revealed scant studies on DH scholarly literature comparing open-access and non-open-access publications. This indicated a potential research gap, hence the present study.

3. Methodology

Scopus database was used for the present study because of its coverage and ease of data extraction. Scopus indexes the most significant number of journals in all different fields compared to the Web of Science (Li et al. 2010). Data collection was done on June 18, 2024. Using the advanced search feature, data was gathered using the query AUTHKEY (digital AND humanities) OR INDEXTERMS (digital AND humanities). The query resulted in a dataset of 3,731 records. The data was extracted into a Microsoft Excel file and classified based on year of publication, document type, language, subject area, and open access type for further analysis. The publications selected ranged from 1971 to 2024. No publications were found during 1973-1976, 1979-1983, 1985, 1987-1988, 1991-1993, 1995, and 2000. So, these 17 years were excluded.

The focus of the investigation was comparing the productivity and impact of DH scholarly literature in terms of OA and Non-OA publications. For this purpose, OA publications were considered under the following four categories as given in Scopus(*Changes to Scopus Open Access (OA) Document Tagging* | *Elsevier Scopus Blog*, n.d.)

- **Gold OA:** The final published version of the research output will be freely and permanently accessible through the publisher's website, and an APC (article processing charge) is usually applicable.
- **Green OA:** Authors make the pre-print or post-print of an article accessible through an institutional repository. APC is not applicable.
- **Bronze OA:** Making articles freely accessible immediately after publication for a short period. This is mainly for the promotional purpose of the published article.
- **Hybrid Gold OA:** A subscription-based journal that allows authors to choose publishing open access. Authors usually pay an APC to these kinds of journals to make their articles open-access.

Research Questions

The present study aimed to address the following questions:

- What is the trend in publications and citation patterns among OA and non-OADH literature?
- Which countries contribute to DH literature predominantly?
- Which subject areas produce high research output in DH literature?
- What are the preferred channels of scholarly communication in DH?

4. Data Analysis

As explained earlier, the data collected from the Scopus database was tabulated and analysed to address the

research questions. The following section presents the analysis and results.

4.1 Publication and Citation Trends

To understand the growth of literature on a particular subject, it is essential to trace the publication and citation trends. Digital humanities literature is characterized by a steady growth from 1971 to 2024—however, the interest in publishing DH literature on open access picked up much later. Since 2009, a consistent increase in OA publications has been witnessed. In terms of total publications, a growing trend may be observed. The highest number of publications have been produced during 2022 (475). However, a significant portion of DH literature (65%) is non-open access.

In terms of citations, there is no consistent trend noticed. OA publications (1321) have garnered 12290 citations, averaging 9.30 citations per publication, while non-OA publications (2410) have attracted 9323 citations with a mean citation of 3.87. Thus, it may be noted that OA publications show greater citation potential. Figure 1 shows the publication and citation trends in DH literature. Huge spikes in citation trends of OA publications may be observed during 2011 and 2014.

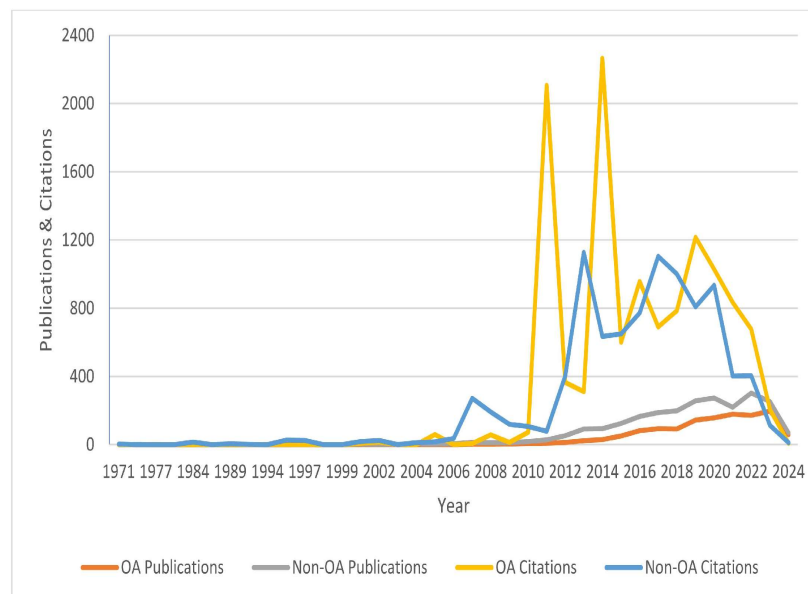


Figure 1: Publication and Citation Trends of Digital Humanities Literature

Open-access publishing is useful for all research fields (Morillo, 2020). Like any other researcher, digital humanists also contribute to open-access literature. The data retrieved from the Scopus database is mainly categorized into four types of open-access publishing and their combinations. 30% (404 publications) of the total open access DH literature is gold, 29% (395 publications) is green, 12% (165 publications) is hybrid gold, and 9% (122 publications) is bronze open access. Open-access publishing caught more attention than non-open-access publishing. Zhang (2006) notes that citations received by open-access publications are higher than those received by non-open-access publications. Open access publications attracted 12290 citations, while non-open access publications received 9323 citations. That is, 60% of the total citations were gained by open-access publications. Among these open-access publications, a higher number of citations were attracted by green open-access publications (5933 citations, 27%).

Year	No. of OA Publications	No. of Non-OA Publications	Total Publications	OA Citations	Non-OA Citations	Total Citations	Year	No. of OA Publications	No. of Non-OA Publications	Total Publications	OA Citations	Non-OA Citations	Total Citations
1971	-	1	1	-	4	4	2007	2	14	16	7	272	279
1972	-	2	2	-	0	0	2008	2	12	14	57	192	249
1977	1	-	1	2	0	2	2009	3	12	15	14	119	133
1978	-	1	1	-	1	1	2010	7	18	25	72	107	179
1984	-	1	1	-	15	15	2011	9	28	37	2106	79	2185
1986	-	1	1	-	0	0	2012	14	52	66	365	402	767
1989	-	1	1	-	5	5	2013	23	92	115	311	1127	1438
1990	-	1	1	-	2	2	2014	29	93	122	2266	635	2901
1994	1	-	1	2	0	2	2015	51	125	176	600	650	1250
1996	-	1	1	-	26	26	2016	82	166	248	957	772	1729
1997	-	2	2	-	25	25	2017	93	188	281	690	1104	1794
1998	-	1	1	-	1	1	2018	92	199	291	786	1002	1788
1999	-	1	1	-	0	0	2019	145	257	402	1215	809	2024
2001	1	4	5	8	19	27	2020	157	274	431	1029	934	1963
2002	1	2	3	15	25	40	2021	178	220	398	835	403	1238
2003	-	2	2	-	0	0	2022	172	303	475	678	404	1082
2004	-	5	5	-	12	12	2023	196	252	448	206	113	319
2005	1	4	5	59	17	76	2024*	60	69	129	9	11	20
2006	1	6	7	1	36	37	Total	1321	2410	3731	12290	9323	21613

*up to 18th June 2024

Table 1: Year wise distribution of open access and non-open access publications and citations

4.2 Open Access Publications in Digital Humanities

Type of OA Publishing	No. of Publications	% of Total	No of Citations	% of Total
Bronze OA	122	3.27	608	2.81
Bronze OA; Green OA	43	1.15	424	1.96
Gold OA	404	10.83	1242	5.75
Gold OA; Green OA	133	3.56	2388	11.05
Green OA	395	10.59	5933	27.45
Green OA; Hybrid Gold OA	59	1.58	836	3.87
Hybrid Gold OA	165	4.42	859	3.97
Total OA	1321	35.41	12290	56.86
Non-OA	2410	64.59	9323	43.14
Total	3731	100.00	21613	100.00

Table 2: Open Access (OA) Publishing in DH research

Statistical Analysis

The citation advantage of OA publications was tested using the Mann-Whitney U-test. The mean rank of citation count for OA publications was 2036.47, and that of non-OA publications was 1772.56 (p-value=0.000). The results indicate that OA publications receive higher citations than non-OA publications. These results were consistent with the results of previous studies by Chua et al. (2017) and Koler-Povh et al. (2014).

4.3 Country-wise distribution of DH literature

Country	OA Publications	Non-OA Publications	Total	Country	OA Publications	Non-OA Publications	Total
United States	258	607	865	Switzerland	40	43	83
Germany	131	270	401	Australia	37	46	83
United Kingdom	181	171	352	Sweden	24	42	66
Italy	103	188	291	Ireland	33	33	66
France	104	121	225	Brazil	29	29	58
Spain	128	85	213	Portugal	19	37	56
China	34	175	209	Taiwan	8	47	55
Undefined	28	143	171	Denmark	23	27	50
Netherlands	83	84	167	Japan	17	31	48
Canada	46	82	128	Belgium	28	17	45
Finland	41	67	108	India	13	30	43
Austria	37	64	101	Greece	15	28	43
Russian Federation	28	56	84				

Table 3. Country-wise distribution of DH literature

Although 93 countries are involved in research related to the digital humanities, 25 countries with the highest contributions to DH are listed in Table 3. The United States of America produced the highest number of publications (865 – 23% of total). This is followed by Germany (10%) and the UK (9%). India is placed in 24th position with a

1 per cent (43 publications) contribution. The first Scopus-indexed Indian contribution was found in 2014. This shows India's recent entry into DH research.

Considering the open and non-open access publications, Belgium published 62 per cent of its scholarly output on open access. Spain (60 per cent), the Czech Republic (53 per cent), the UK (51 per cent), Ireland, Brazil, and South Korea each made up 50 per cent of their total publications open access. However, the USA, being the top publisher of DH literature, only disseminates 29 per cent of its total publication through open-access channels.

4.4 Subject area-wise distribution of DH literature

Subject Area	No. of Open Access Publication	Total No. of Publications	Subject Area	No. of Open Access Publication	Total No. of Publications
Computer Science	563	1825	Multidisciplinary Economics,	20	24
Social Sciences	669	1748	Econometrics and Finance	8	24
Arts and Humanities	578	1422	Energy	10	17
Mathematics	108	344	Health Professions	4	11
Engineering	94	296	Chemistry	9	11
Decision Sciences	52	130	Neuroscience	7	10
Business, Management and Accounting	27	68	Biochemistry, Genetics and Molecular Biology	5	10
Earth and Planetary Sciences	30	64	Nursing	3	5
Medicine	26	52	Pharmacology, Toxicology and Pharmaceutics	2	4
Environmental Science	29	52	Immunology and Microbiology	2	3
Psychology Agricultural and Biological Sciences	24	45	Chemical Engineering	3	3
Sciences	10	41	Veterinary		1
Materials Science	29	35			
Physics and Astronomy	18	27			

Table 4. Subject area-wise distribution of DH literature

DH is rooted in 27 major subject areas, which include science, computer science, and social science disciplines. As a mixture of computer science and social science, or as a way of technically visualizing humanistic things, we can see the dominance of computer science, social science, and the humanities discipline in DH scholarly publications. Most digital humanities scholarly content is indexed under computer science (1825 publications, 48%). And it is immediately followed by social science (46 per cent) and the arts and humanities (38 per cent). All other subjects contribute less than 10 per cent of the total publications.

Considering the open-access publishing rate, Computer science, social science, and humanities are the prominent contributors to DH research. Still, they only produce 30 per cent, 38 per cent, and 40 per cent of their total publications as open access, respectively. At the same time, science disciplines such as material science (82 per cent), chemistry (81 per cent), neuroscience (77 per cent), physics and astronomy (66 per cent), and medicine (50 per cent) contribute a minimum of 50 per cent of the total DH literature to open access publishing channels.

4.5 Document type-wise Distribution of DH literature

Document Type	No.of Non-OA - Publications	No.of Non-OA Publications	Total No.of - Publications	Citations received by OA publications	Citations received by Non-OA publications	Total Citations
Article	882	986	1868	9308	4782	14090
Book	26	17	43	210	138	348
Book chapter Conference	34	171	205	94	314	408
paper Conference	271	1101	1372	1718	3714	5432
review		1	1			0
Data paper	5		5	9		9
Editorial	8	9	17	71	4	75
Erratum	1		1	1		1
Letter	2	1	3		1	1
Note	6	11	17	16	19	35
Retracted	1		1			0
Review	85	113	198	863	351	1214
Total	1321	2410	3731	12290	9323	21613

Table 5. Document type-wise distribution of DH literature

The major channel of scholarly communication on DH is articles. 50 per cent (1868 publications) of the total publications are articles, and 36 per cent (1372 publications) are conference papers. All other documents, including book chapters (5 per cent), reviews (5 per cent), books (1 per cent), conference reviews, data papers, editorials, erratum, letters, and notes, together contribute only 491 publications (13 per cent).

Here, 50% of the total publications are peer-reviewed articles, but only 44% of them are freely available. Only 19% of conference papers are open to all. As articles and conference papers contribute more to the total DH literature, the lack of free access will reduce awareness about the growth of DH.

Table 4 shows that the number of citations attracted by the OA DH publications is much greater than that of the non-open OA publications. 47 per cent of the OA review papers and peer-reviewed articles contribute 71 per cent and 66 per cent to the total citations, respectively. This higher ratio may be due to the easy visibility of open-access publications.

4.6 Top cited Papers in DH

Title	Year	Cited by	Language of Original Document	Document Type	Access Type
Quantitative analysis of culture using millions of digitized books	2011	1940	English	Article	Open Access
Big Data, new epistemologies and paradigm shifts	2014	1423	English	Article	Open Access
Learning cultural heritage by serious games	2014	387	English	Review	Open Access
The INCEPTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation	2018	232	English	Conference paper	
Towards a sociology of computational and algorithmic journalism	2013	223	English	Article	
Programmed method: Developing a toolset for capturing and analyzing tweets	2014	187	English	Article	
Machine Learning for Cultural Heritage: A Survey	2020	163	English	Article	Open Access
Cultural shift or linguistic drift? Comparing two computational measures of semantic change	2016	154	English	Conference paper	Open Access
Digital storytelling as a signature pedagogy for the new humanities	2008	146	English	Article	
On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges	2015	137	English	Conference paper	

Table 6. Top cited papers in DH

Table 6 lists the top 10 cited papers. Five of these papers are open-access publications. The highly cited DH paper was published in 2011 and received 1940 citations. All the top cited papers are in English, and six of them are peer-reviewed articles.

5. Inferences

Digital humanities, earlier known as humanities computing, is emerging as a new research trend among the scholarly community. In 2009, The Chronicle of Higher Education called digital humanities “the first ‘next big thing’ in a long time,” underlying the implications of digital technologies in transforming humanities research, teaching, and learning. Digital humanities is considered an interdisciplinary subject in which researchers from various fields constantly contribute, and LIS ranks third in the number of authors contributing to digital humanities research (Amanullah, 2022). Thus, the present study investigated the growth of the digital humanities research trend through OA and non-OA publishing channels.

Digital humanities research dates back to the 1950s, but it has gained more attention among researchers. From 1971 to 2024, steady progress in the number of publications on digital humanities shows its growth and importance. DH is spread over 27 subject areas and 93 countries, conveying its impact on the research world. As an intersection of computer technology with humanities subjects, the computer science discipline shows its dominance in DH research with the highest number of publications. The majority of the digital humanities literature is not openly available to all, and thus it restricts the free use of information. Apart from theoretical concepts, DH paves a strong path for building society by preserving the culture for posterity.

From the present study, it is clear that even if the number of open-access publications is comparatively lower than that of non-open-access publications, open-access resources attract more citations. 56 per cent of the total citations from DH publications are gained from open-access publishing only. This finding aligns with the similar observation in the studies conducted by Sahu *et al.* (2005), and Wang *et al.* (2015).

6. Conclusion

This paper summarizes the research trend in the digital humanities. An attempt was made to evaluate the growth of publications in the digital humanities, along with the citations received by open and non-open-access publications. The scattering of DH publications from 1971 to 2024 in different countries and subject areas is also explored. This study found that the majority of digital humanities research outcomes are not available in the public domain, which impedes the creation of new knowledge in the discipline. Digital humanists contribute more of their intellectual output to non-open-access sources. As DH is a way of seeking help from novel technology to solve humanistic problems, easy access to DH must enhance these problem-solving skills, thus strengthening the efficiency of the information society. The number of open and non-open DH publications is increasing daily, and both kinds of publications attract citations. However, the number of citations received by open-access literature is higher than that of non-open-access publishing. DH is growing, and more intense study is needed to analyse its characteristics.

References

[1] Amanullah, A. (2023). An investigation in the interdisciplinary nature of digital humanities: A bibliometric

analysis. *DESIDOC Journal of Library & Information Technology*, 43(4).

[2] Basak, M., Roy, S. B. (2022). Mapping the literature on digital humanities: A bibliometric study using Scopus data. *DESIDOC Journal of Library & Information Technology*, 42(6), 354.

[3] Changes to Scopus Open Access (OA) document tagging | Elsevier Scopus Blog. (n.d.). Retrieved June 26, 2024, from <https://blog.scopus.com/posts/changes-to-scopus-open-access-oa-document-tagging>

[4] Chansanam, W., Ahmad, A. R., Li, C. (2022). Contemporary and future research of digital humanities: A scientometric analysis. *Bulletin of Electrical Engineering and Informatics*, 11(2), 1143–1156.

[5] Chua, S., Qureshi, A. M., Krishnan, V., Pai, D. R., Kamal, L. B., Gunasegaran, S., Afzal, M., Ambawatta, L., Gan, J., Kew, P., Winn, T., Sood, S. (2017). The impact factor of an open access journal does not contribute to an article's citations. *F1000Research*, 6, 208. <https://doi.org/10.12688/f1000research.10892.1>

[6] Dalbello, M. (2011). A genealogy of digital humanities. *Journal of Documentation*, 67(3), 480–506.

[7] Dorta-González, P., Dorta-González, M. I. (2023). The influence of funding on the open access citation advantage. *Journal of Scientometric Research*, 12(1), 68–78. <https://doi.org/10.5530/jscores.12.1.010>

[8] Gupta, N., Chakravarty, R. (2021). Science mapping analysis of digital humanities research: A scientometric study. *Library Philosophy and Practice (e-journal)*, (September), 1–15. <https://digitalcommons.unl.edu/libphilprac/6126/>

[9] Koler-Povh, T., Južniè, P., Turk, G. (2014). Impact of open access on citation of scholarly publications in the field of civil engineering. *Scientometrics*, 98(2), 1033–1045. <https://doi.org/10.1007/s11192-013-1101-x>

[10] Liu, A. (2013). The meaning of the digital humanities. *PMLA*, 128(2), 409–423.

[11] Luhmann, J., Burghardt, M. (2022). Digital humanities—A discipline in its own right? An analysis of the role and position of digital humanities in the academic landscape. *Journal of the Association for Information Science and Technology*, 73(2), 148–171.

[12] Makwana, P. N., Gadhavi, G. G. (2023). Literature survey and analysis of digital humanities and libraries in recent years. *DESIDOC Journal of Library & Information Technology*, 43(4).

[13] Sahu, D. K., Gogtay, N. J., Bavdekar, S. B. (2005). Effect of open access on citation rates for a small biomedical journal. Paper presented at the Fifth International Conference on Peer Review and Biomedical Publication, Chicago, IL.

[14] Smith, M. N. (2002). Computing: What's American literary study got to do with IT? *American Literature*, 74(4), 833–857.

[15] Tang, M. C., Cheng, Y. J., Chen, K. H. (2017). A longitudinal study of intellectual cohesion in digital

humanities using bibliometric analyses. *Scientometrics*, 113, 985–1008.

[16] Wang, Q. (2018). Distribution features and intellectual structures of digital humanities: A bibliometric analysis. *Journal of Documentation*, 74(1), 223–246. <https://doi.org/10.1108/JD-05-2017-0076>

[17] Wang, X., Liu, C., Mao, W., Fang, Z. (2015). The open access advantage considering citation, article usage and social media attention. *Scientometrics*, 103(2), 555–564.

[18] Zhang, Y. (2006). The effect of open access on citation impact: A comparison study based on web citation analysis.