



A Bibliometric Study of World Research Output on Free and Open Source Software Literature during 1960-2016

Dr Preet Kanwal

Department of Computer Science
Sri Guru Gobind Singh College, Sector 26,
Chandigarh, India

Abstract: Bibliometrics play a prominent role in the domain of research evaluation to measure the literature output on any given subject. Bibliometric techniques are used for quantitative analysis of scientific publications based on various parameters like publication counts, authors, types of publications and the citations therein to measure the research outputs of individuals/research teams, institutions, and countries, and to map the development of new disciplines. Similar approach has been adopted in this paper to identify the global literature output on 'Free and Open Source Software' literature based on the research data retrieved from 'Scopus' database source. The study attempts to identify the bibliometric characteristics of the research publications from Scopus database during the period 1960-2016. Bibliometric techniques have been used to analyse the data which has been classified by using spreadsheet software. The results show that the total publication output is 165,751 of which the articles 106108 (64%) are the most prominent publications, most prolific author is A. E. Hassan with 113 publications. The subject area with highest research output is of Medicine (31.72%). 'Plos One' is the highly productive journal with 34354 (20.73%) publications contributing one-fifth of total publications. The most productive institution is University of Washington, Seattle which produced 900 publications. United States tops the list of countries with publication output being 47058 (28.39%) publications.

Keywords: Bibliometrics; Free software; Open Source Software; FOSS

I. INTRODUCTION

Originally, all software was free and came bundled with the hardware provided by the hardware companies. Openness and cooperation in software sharing was followed by academia and research communities during the early years of development of computers in 1950s and 1960s. It was with establishment of Microsoft in mid 1970s that proprietary software requiring product license fee with restrictive copyrights came into being. Free software [1] movement was initiated in mid 1980s followed by Open Source initiative [2] in 1990s. Free and Open Source Software (FOSS) is an umbrella term for any software that can be classified as Free software or Open Source software. It allows sharing of source code to provide the users with freedom to use, study, modify, redistribute the software in any way. The benefits associated with using FOSS instead of proprietary software include decreasing software costs, increasing security, protecting privacy, and giving users more control over their own hardware through customizations and improvements.

Over the years the FOSS movement has gained momentum giving rise to many successful projects. The ability to create new applications quickly and reliably at lower costs have attracted commercial proprietary software companies to invest in FOSS. Nowadays FOSS products are available in almost all areas where originally commercial products were being used. The FOSS development model has gained acceptance worldwide with more and more projects as well as project hosting sites coming up. Since FOSS development model depends upon volunteers who may be geographically distributed and work in different time-zones across language barriers to produce functionally viable projects with or without financial remunerations, it has become a leading research area. Bibliometric techniques are used for quantitative analysis of scientific publications to measure the research outputs of individuals as well as research groups, institutions, and countries, and to map the development of new disciplines [3, 4]. Over the years, considerable literature has been produced in the area of FOSS. The present bibliometric study of research

output on FOSS literature is conducted for quantitative analysis of publications on various parameters.

II. REVIEW OF LITERATURE

Santos performed bibliometric analysis on open innovation research over 10 year period (2003-2013) using the Scopus database. It was observed that open innovation research is mostly focused on the analysis of the U.S.A. and European countries. Analysis by time periods shows an increase on the number of target countries and regions of open innovation research and that the open innovation research is influencing an increasing number of areas apart from business, management and engineering [5].

Khode and Thakkar performed bibliometric analysis of publications on open source appeared in the emerald journals from April 1998 to May 2010. The bibliographical details of each of the 63 publications on open source were downloaded from the emerald website and recorded onto a spread sheet. Results showed that maximum number of articles contributed by single author is 40 (63.49%) and Library Hi Tech and Program electronic library and information systems have first rank among the published literature on open source software respectively [6].

Newby et al applied Lotka's Law to metadata on open source software development from the Linux Software Map (LSM), which documents many open source projects, and Sourceforge, one of the largest resources for open source developers to predict the proportion of authors at different levels of productivity. Authoring patterns found are comparable to prior studies of Lotka's Law for scientific and scholarly publishing. Lotka's Law was found to be effective in understanding software development productivity patterns, and offer promise in predicting aggregate behaviour of open source developers [7].

Maity and Teli limited their bibliometric study within the database of Directory of Open Access Journals and the subject area is Library and Information Science. The articles which were published in English language published from the year 2004 to 2014 were only taken into consideration. After analysis

of research articles productivity on subject subfields of LIS it was found that the field of Information and Communication Technology is more prominent in case of research article productivity [8].

Jayaprakash performed bibliometric analysis of publications on cloud computing taken from the emerald journals from 2000 to 2014. The bibliographical details of each of the 781 publications in various journals of emerald were downloaded from the emerald website and recorded in MS-excel for calculation. It was found that maximum number of articles 287 (37%) contributed by single author and majority of the contribution among the published literature to cloud computing is from Library Hi Tech News and Library Hi Tech journals respectively [9].

Vahid Garousi undertook a bibliometric analysis of the Turkish software engineering (SE) community (researchers and institutions) conducted based on the number of papers published in the software-engineering-related venues taken from Scopus until year 2014. The findings in addition to identifying the top ranked institution and the top-ranked scholar revealed that Turkey produces only about 0.49 % of the worldwide SE knowledge, as measured by the number of papers in Scopus, which is very negligible unfortunately [10].

The literature review reveals that no systematic study has been undertaken for presenting the worldwide growth of FOSS literature. This paper intends to fill the gap related to knowledge of existing FOSS literature and contributes towards providing informative value.

III. STATEMENT OF THE PROBLEM

The research problem is conceived under the title “A Bibliometric Study of World Research Output on Free and Open Source Software Literature during 1960-2016”.

IV. OBJECTIVES

The main objective of this study is to analyze the research output of Free and Open Source Software (FOSS) literature in global contexts as reflected in the publications output during 1960-2016. The study has the following objectives:

- To examine the worldwide research production on FOSS literature.
- To identify the document type of FOSS literature publications
- To identify the most prolific authors producing more than fifty articles in area of FOSS studies
- To identify the journal distribution of research productivity
- To identify the specific area of higher FOSS research
- To identify and rank the institutions involved in research on FOSS and producing related literature.
- To identify research productivity count of the publications on the basis of geographical distribution.

V. RESEARCH METHODOLOGY

The study used Scopus, a multidisciplinary online database, which is the world's leading indexing and abstracting database [11]. The relevant data on FOSS literature was extracted for the period 1960-2016. The advanced search strategy involving query comprising of search terms such as 'open source', 'free software', 'open source software', 'free and open source software', 'libre software', 'foss', 'oss', 'floss', 'f/oss, and 'free and libre open source software' was used to search and download the data using Title, Abstract, and Keywords fields together corresponding to the period

from 1960-2016. A total of 165,751 publications matched the search criteria. Bibliometrics analysis method was used for the study in hand. The bibliographic details for each record including title, year of publication, type of document, author and authors' affiliation, source of publication, subject area and country of input have been collected. Subsequently, all the collected data was compiled and computed using a spreadsheet package. Statistical tools such as frequency distribution and percentage analysis are used for analysis and interpretations.

VI. DATA ANALYSIS

The data has been analyzed on following parameters:

- Year - To analyze the trend in worldwide research production on FOSS.
- Document Type - To analyze the various document types of publications.
- Author - To analyze the prolific authors based on their publication counts.
- Source - To analyze the distribution of research productivity among various sources (journals etc.).
- Subject - To analyze the distribution of literature on FOSS across subjects.
- Institution - To analyze and rank the institutions involved in production of FOSS literature.
- Country - To identify research productivity count of FOSS publications on the basis of geographical distribution.

Analysis based on the above parameters is detailed as below:

A. Year-wise Publication Trends

It is observed from Fig. 1 that the publication of FOSS research seems to be in exponential trend. During the last ten years there is a substantial increase in the publications. This indicates that the awareness and importance of FOSS literature has been in increasing trend.

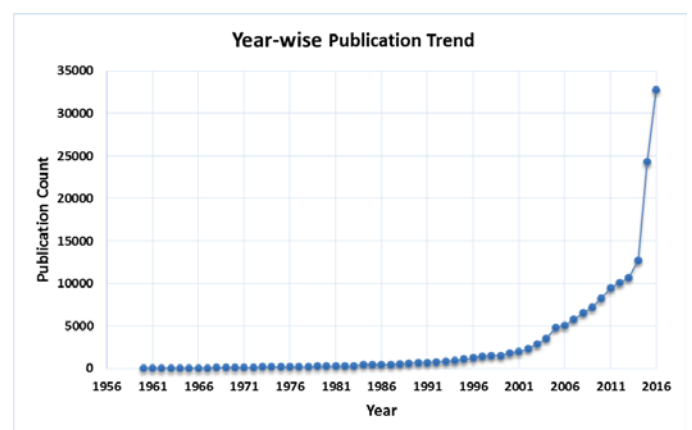


Fig. 1: FOSS Publication Trend (Year-wise)

It is clearly observed from Table I that during the period 1960-2016 a total of 165,751 publications were published at the global level. Table I shows that publication output on FOSS literature grew from 19 publications in 1960 to 32,773 publications in 2016. Although the number of publications increased over the years but they were less than 100 publications per year before 1973. It was only after 1995 that the publication of literature on FOSS gained momentum. The highest number of contributions of 32,773 (19.77%) items is

observed in 2016, followed by 24296 (14.66%) in 2015 and 12673 (7.65%) in 2014.

Table I. Year-Wise Distribution of Research Publications

Sr. No.	Year	Publication Count	Percentage
1	2016	32773	19.77
2	2015	24296	14.66
3	2014	12673	7.65
4	2013	10670	6.44
5	2012	10105	6.1
6	2011	9407	5.68
7	2010	8184	4.94
8	2009	7118	4.29
9	2008	6513	3.93
10	2007	5748	3.47
11	2006	5040	3.04
12	2005	4818	2.91
13	2004	3493	2.11
14	2003	2816	1.7
15	2002	2249	1.36
16	2001	1930	1.16
17	2000	1772	1.07
18	1999	1504	0.91
19	1998	1449	0.87
20	1997	1420	0.86
21	1996	1261	0.76
22	1995	1084	0.65
23	1994	924	0.56
24	1993	786	0.47
25	1992	710	0.43
26	1991	692	0.42
27	1990	664	0.4
28	1989	558	0.34
29	1988	514	0.31
30	1987	437	0.26
31	1986	441	0.27
32	1985	405	0.24
33	1984	387	0.23
34	1983	283	0.17
35	1982	218	0.13
36	1981	245	0.15
37	1980	238	0.14
38	1979	254	0.15
39	1978	199	0.12
40	1977	203	0.12
41	1976	176	0.11
42	1975	200	0.12
43	1974	172	0.1
44	1973	152	0.09
45	1972	94	0.06
46	1971	63	0.04
47	1970	64	0.04
48	1969	71	0.04
49	1968	52	0.03
50	1967	36	0.02
51	1966	39	0.02
52	1965	39	0.02
53	1964	34	0.02
54	1963	25	0.01
55	1962	16	0.01
56	1961	18	0.01
57	1960	19	0.01
	Total	165751	100

B. Publication Distribution Based on Document Type

The Scopus identifies 15 types of documents. Fig. 2 shows the data about spread of FOSS publication output in different types of documents during the period 1960-2016.

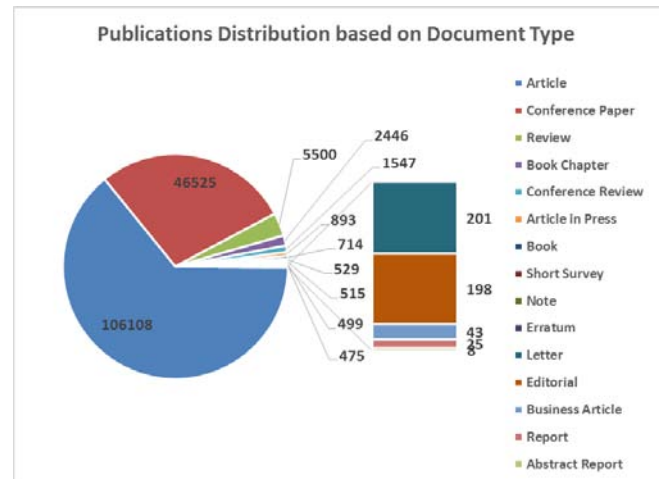


Fig. 2: Publication Distribution based on Document Type

It is observed from data that out of total 165,751 documents on FOSS literature, Article 106108 is the most prominently found document type having 64% share of total publications, followed by Conference Paper 46525 (28.07%), Review 5500 (3.32%), Book Chapter 2446 (1.48%) and Conference Review 1547 (0.93%). Article in Press, Book, Short Survey, Note, Erratum, Letter, Editorial, Business Article, Report and Abstract document types scored low numbers of output i.e., below 0.6% of output for FOSS literature.

C. Author-wise Publication Contribution

The data showed that a total of 158 authors have contributed publications in the area of FOSS literature. Table II presents the frequency distribution of number of authors contributing publications towards literature on FOSS. Maximum 90 authors have contributed in the range of 21-30 publications, followed by 31 authors having contributed less than 20 publications. Only one author has published more than 100 publications during the period of study.

Table II. Author Frequency Distribution

Sr. No.	Range	No. of Authors N=158	Percentage
1	11-20	31	19.62
2	21-30	90	56.96
3	31-40	16	10.13
4	41-50	10	6.33
5	51-60	06	3.8
6	61-70	03	1.9
7	71-80	01	0.63
8	81-90	--	--
9	91-100	--	--
10	>100	01	0.63

Fig. 3 shows data on top productive authors who published more than 50 papers. There are 11 authors who have contributed more than 50 publications each during 1960-2016. The most publishing author is A.E. Hassan with 113 articles, followed by Anon with 73 articles and G. Robles with 67 articles. The data shows that researchers are actively involved in FOSS research. The trend shows a substantial and growing

body of research output in the field of free and open source software literature in future.

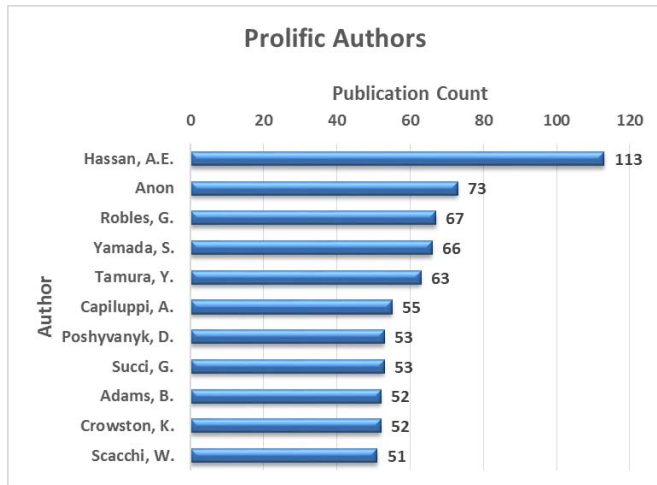


Fig. 3: Most Prolific Authors Contributing More Than Fifty Publications

D. Source-wise Publication Distribution

Table III shows the twenty high productive sources (journals, conference proceedings, lecture notes etc.) published 45752 research papers in FOSS literature, which accounts for nearly 27.62% of the total output during 1969-2017.

Table III. Most Prolific Sources

Rank	Source Title	Publication Output	Percentage
1	Plos One	34354	20.73
2	Lecture Notes In Computer Science Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics	3514	2.12
3	Proceedings of SPIE The International Society For Optical Engineering	1718	1.04
4	BMC Bioinformatics	635	0.38
5	Bioinformatics	620	0.37
6	Astrophysical Journal	456	0.28
7	IFIP Advances In Information And Communication Technology	437	0.26
8	Nucleic Acids Research	398	0.24
9	Proceedings International Conference On Software Engineering	391	0.24
10	Communications In Computer And Information Science	361	0.22
11	Applied Mechanics And Materials	358	0.22
12	Astronomy And Astrophysics	344	0.21
13	Aip Conference Proceedings	310	0.19
14	Advanced Materials Research	304	0.18
15	SAE Technical Papers	298	0.18
16	Journal Of Physics Conference Series	282	0.17
17	Computer Physics Communications	276	0.17
18	Journal Of Bacteriology	242	0.15
19	Studies In Health Technology And Informatics	235	0.14
20	Procedia Computer Science	219	0.13

It is observed that Plos One is the highly productive journal with 34354 (20.73%) publications which accounts for more than one-fifth of the total output. However, the sources with second and third highest output have less than 3% publications namely, Lecture Notes in Computer Science including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics (2.12%) and Proceedings of SPIE The International Society for Optical Engineering (1.04%). The remaining sources have less than 1% of the total publication output. Thus, it is observed that spread of articles across sources is more, however, per source output is meagre. Results show that the journal Plos One is the most preferred source by researchers of free and open source software.

E. Subject-wise Publication Distribution

The data in Table IV shows that the maximum number of FOSS research literature is contributing to the subject of Medicine (31.72%) followed by Biochemistry, Genetics and Molecular Biology (27.44%), Computer Science (26.25%), Agricultural and Biological Sciences (24.96%) and Engineering (21.21%). 14 subjects range from 1-9%. Remaining subjects show less than 1% of publications. 0.22% of publications are in undefined subjects. It is also observed that FOSS research output is prominent in scientific subjects.

Table IV. Subject-wise Publication Output (Top 20)

Rank	Subject-Area	Publication Output	Percentage
1	Medicine	52571	31.72
2	Biochemistry, Genetics and Molecular Biology	45481	27.44
3	Computer Science	43512	26.25
4	Agricultural and Biological Sciences	41370	24.96
5	Engineering	35148	21.21
6	Physics and Astronomy	14188	8.56
7	Mathematics	12533	7.56
8	Earth and Planetary Sciences	11348	6.85
9	Social Sciences	10657	6.43
10	Environmental Science	8345	5.03
11	Materials Science	7279	4.39
12	Chemistry	5358	3.23
13	Energy	4545	2.74
14	Chemical Engineering	3789	2.29
15	Business, Management and Accounting	3363	2.03
16	Arts and Humanities	2844	1.72
17	Decision Sciences	2668	1.61
18	Immunology and Microbiology	2338	1.41
19	Neuroscience	1701	1.03
20	Pharmacology, Toxicology and Pharmaceutics	1596	0.96

F. Institution-wise Publication Distribution

160 institutions are contributing towards producing research output in the area of FOSS literature. Table V presents the top 20 highest contributor institutes. Only 10 institutions have produced more than 700 publications. Maximum contribution to FOSS research output is by University of Washington, Seattle, which has published 900 (0.54%) publications between the years 1960-2016. It is followed by Universidade de Sao Paulo - USP (0.52%) and Chinese Academy of Sciences (0.5%). It has been observed

that at institute level, the contribution per institute is less than 0.55% each.

Table V. Most Productive Institutions

Rank	Institution	Publication Output	Percentage
1	University of Washington, Seattle	900	0.54
2	Universidade de Sao Paulo - USP	856	0.52
3	Chinese Academy of Sciences	835	0.5
4	University of Toronto	803	0.48
5	UC Berkeley	803	0.48
6	Stanford University	794	0.48
7	UCL	766	0.46
8	University of Oxford	754	0.45
9	University of Cambridge	730	0.44
10	Massachusetts Institute of Technology	710	0.43
11	Imperial College London	685	0.41
12	Eidgenossische Technische Hochschule Zurich	680	0.41
13	CNRS Centre National de la Recherche Scientifique	645	0.39
14	The University of British Columbia	634	0.38
15	University of Maryland	612	0.37
16	Consiglio Nazionale delle Ricerche	596	0.36
17	University of Wisconsin Madison	587	0.35
18	University of California, San Diego	570	0.34
19	Tsinghua University	545	0.33
20	Peking University	544	0.33

G. Country-wise Publication Distribution

Table VI presents country-wise distribution of FOSS literature by the topmost twenty countries. United States tops the list of countries producing publication output on FOSS literature with 47058 (28.39%) during the period 1960-2016 followed by China (10.02%) and Germany (8.15%).

Table VI. Most Productive Countries

Rank	Country	Publication Output	Percentage
1	United States	47058	28.39
2	China	16605	10.02
3	Germany	13512	8.15
4	United Kingdom	13193	7.96
5	France	8196	4.94
6	Italy	8136	4.91
7	Canada	7629	4.6
8	Japan	6563	3.96
9	Spain	5989	3.61
10	India	5481	3.31
11	Australia	5367	3.24
12	Netherlands	4513	2.72
13	Brazil	3895	2.35
14	Switzerland	3762	2.27
15	South Korea	3367	2.03
16	Sweden	3273	1.97
17	Taiwan	2550	1.54
18	Belgium	2440	1.47
19	Austria	2089	1.26
20	Poland	1961	1.18

United Kingdom, which comes fourth on the list, produced 7.96% of total output on FOSS literature. It is observed that India is ranked at 10th position with a meagre 3.31%. United

States and China together are producing more than one-third of literature output in FOSS research. The top five countries, namely United States, China, Germany, United Kingdom and France are contributing almost 60% of the total research output in the area of FOSS literature.

VII. FINDINGS

The study was conducted using the Scopus database and other results were based upon the journals and other published material related to FOSS covered therein. The paper does not take into account any research produced beyond the coverage of Scopus database. The major findings of the study are:

- There exist 165751 FOSS related publications during the year 1960-2016.
- The highest number of contributions of 32773 (19.77%) items is observed in 2016.
- Publications in the form of articles are found to be the most prominent type of literature document (64%).
- The most publishing author is A.E. Hassan with 113 publications.
- Plos One is the highly productive source with a share of 20.73% of the total publications.
- Maximum number of publications are in the subject of Medicine (31.72%).
- University of Washington, Seattle produced the highest number of publications 900 (0.54%).
- United States tops the list of countries producing 28.39% publications during the period 1960-2016.
- Nearly 60% of the outputs are provided by top five countries, United States, China, Germany, United Kingdom and France.
- During the last ten years there is a substantial increase in the publications.

VIII. CONCLUSION

Bibliometric analysis of literature within online databases provide in-depth information on research output in any chosen area of study. The bibliometric analysis conducted on longitudinal data on FOSS literature has provided useful insight into various parameters of FOSS literature. A significant growth has occurred in the area of FOSS research output especially during the last decade.

IX. REFERENCES

- [1] "What is free software?," GNU Operating System, Feb. 17, 2015. <http://www.gnu.org/philosophy/free-sw.html>.
- [2] "The open source definition," Open Source Initiative. <http://www.opensource.org/docs/osd>.
- [3] David A. Pendlebury, White Paper "Using Bibliometrics in Evaluating Research," Research Department, Thomson Reuters, Philadelphia, PA USA. https://services.anu.edu.au/files/system/Pendlebury_White_Paper.pdf.
- [4] University of Waterloo Working Group on Bibliometrics, Winter 2016, White Paper on Bibliometrics, Measuring Research Outputs through Bibliometrics, Waterloo, Ontario: University of Waterloo, DOI: 10.13140/RG.2.1.3302.5680
- [5] António Bob Santos, "Open Innovation research: trends and influences – a bibliometric analysis," Journal of Innovation Management, JIM 3, no. 2, pp. 131-165, 2015.
- [6] Subhash Khode and Seema Thakkar, "Bibliometric Analysis of Literature Published in Emerald Journals on Open Source," Gyankosh: The Journal of Lib. & Info. Management, vol.3, no. 2, pp 44-49, Jul-Dec, 2012.

- [7] Gregory B. Newby, Jane Greenberg and Paul Jones, "Open source software development and Lotka's Law: Bibliometric patterns in programming," Journal of American Society for Information Science and Technology, vol. 54, Issue 2, pp. 169-178, Jan 2003.
- [8] Asish Maity, Soumen Teli, "A Bibliometric Analysis on the Directory of Open Access Journals (DOAJ) in the Subject Domain of LIS From the Year 2004-2014," International Journal of Innovative Research in Science, Engineering and Technology, vol. 4, Issue 4, pp. 1955-1962, April 2015.
- [9] Jayaprakash, "Bibliometric Analysis of Literature Published in Emerald Journals on Cloud Computing," International Journal of Computational Engineering & Management, vol. 18 Issue 1, pp. 21-26, January 2015.
- [10] Vahid Garousi, "A bibliometric analysis of the Turkish software engineering research community," Scientometrics, vol. 105, pp. 23-49, 2015. DOI 10.1007/s11192-015-1663-x
- [11] "Scopus," Elsevier. <https://www.elsevier.com/solutions/scopus>