Impact of e-print Archives: A Case Study of Physics Archives

E R Prakasan Anil Sagar Anil Kumar V L Kalyane Vijai Kumar

Abstract

E-print archives (electronic preprints) have become the first choice of many researchers for current research, breaking scientific discoveries, and keeping up with colleagues (and competitors) at other institutions. The citations being received to e-prints are evidences of the above stated importance of the archives. Impact Factor and Immediacy Index devised by Thomson Scientific Inc. (formerly Institute for Scientific Information (ISI), Philadelphia) are two parameters based on citations to evaluate traditional journals. These parameters for nearly 6000 scientific journals from all over the world are published yearly as Journal Citation Reports (JCR). As a case study, the present work has made an endeavour to calculate impact factors and immediacy indices of archives uploaded in the Physics categories of arXiv.org of Cornell University, USA by treating them as individual journals. The Physics e-prints uploaded in arXiv.org are classified into twelve sub-categories: viz. AstroPhysics (astro-ph); Condensed Matter (cond-mat); General Relativity and Quantum Cosmology (gr-qc); High Energy Physics - Experiment (hep-ex); High Energy Physics - Lattice (hep-lat); High Energy Physics - Phenomenology (hep-ph); High Energy Physics - Theory (hep-th); Mathematical Physics (math-ph); Nuclear Experiment (nucl-ex); Nuclear Theory (nucl-th); General Physics (Physics); and Quantum Physics (quant-ph). Calculated impact factors and immediacy indices reveal that Nuclear Experiment has made good impact among the categories and followed by High-Energy Physics classes of arXiv.org (High Energy Physics – Theory; High Energy Physics – Lattice; High Energy Physics – Experiment; and High Energy Physics - Phenomenology). The citation trackers like Scopus and web-tools like 'Google-Scholar', 'Citebase', 'Citeseer' etc. will further strengthen the concluded impact of the archives.

Keywords: E-Print Archives, Electronic Preprints, Open Archives, Citation Impact, Impact Factor, Immediacy Index, Physics Open Archives, arXiv.org

1. Introduction

'Open access' (OA) means that a reader of a scientific publication can read it over the Internet, download and even further distribute it for non-commercial purposes without any payments or restrictions. The four most important OA channels are electronic-refereed-scientific periodicals, research-area-specific archive (e-print) servers (in this paper called subject-specific repositories), institutional repositories of individual universities, and self-posting on authors' home pages (Björk, 2004). R&D policy makers around the world have recommended mandating that researchers provide Open Access (OA) to their research articles by self-archiving them free for all on the Web (Harnad, 2001). OA is now firmly on the agenda for funding

7th International CALIBER-2009,
Pondicherry University, Puducherry, February 25-27, 2009

agencies, universities, libraries and publishers. What is needed now is objective, quantitative evidence of the benefits of OA to research authors, their institutions, their funders and to research itself. OA articles have significantly higher citation impact than non-OA articles (Harnad et al., 2004). Brody (2004) is also supported in a web-based analysis of usage and citation patterns. One universally important factor for all authors is impact made by their research papers, typically measured by the number of times a paper is cited.

Now the Open Archives (OA) era has revolutionized with new ideas about starting a global database for finding the number of citations received to the OA submissions. Citebase (Brody, 2003) and Citeseer are two such web tools, which serve this partially. Studies have begun to show that open access increases impact, although more studies and more substantial investigations are needed to confirm the effect. Hitchcock (2004) has given the progress in these directions in the form of a chronological bibliography with some explanation.

2. Citation Studies

The citation analysis in the fields of high-energy Physics and astroPhysics, indicates that the number of citations to traditional preprints has gradually declined over the past 10 years, and that citations to electronic preprints have nearly doubled every year since 1992 (Youngen, 1998a, 1998b). The electronic preprint servers are often the first choice of physicists and astronomers for finding information on current research, breaking scientific discoveries, and keeping up with colleagues (and competitors) at other institutions (Prakasan, 2004a; 2004b). In addition to these benefits, electronic preprints allow the free, unrestricted access to scientific information without concern for international, institutional, or political barriers.

Recently Lawrence (2001) and Brody, et al. (2004) have demonstrated that articles which are available on-line at no charge are cited at substantially higher rates than those which are not. Kurtz et al. (2004) has shown that restrictive access policies can cut article downloads to half the free access rate.

A new measure that becomes possible with online publication is the number of downloads or 'hits', opening a new line of investigation. Brody et al. (2004) have been prominent in showing there is a correlation between higher downloads and higher impact, particularly for high impact papers, holding out the promise not just for higher impact resulting from open access but for the ability to predict high impact papers much earlier, not waiting years for those citations to materialise (e.g. Brody and Harnad, 2004b). The effect can be verified with the Correlation Generator (Correlation Generator).

Citation analysis can be used to find emerging fields, to map the time-course and direction of research progress, and to identify synergies between different disciplines (Brody, 2004). Citation analysis is being mainly used for measuring the impact made by journal articles. But Rousseau (1997) has attempted to compare the impact made by the 'first and second international conferences on bibliometrics, scientometrics and informetrics' with some top journals in the field. Information scientists are already computing web impact factors (Bjorneborn and Ingwersen, 2001).

Garfield, probably the world's foremost proponent of citation analysis through two measures: impact factor and immediacy index, first mentioned the ideas in 1955. The analysis of citations is among the means by which policy-makers, scientists, and information

professionals seek to achieve a greater understanding of the qualitative forces that affect communications in science (Tomer, 1986). Like nuclear energy, the two measures have become a mixed blessing, expected that it would be used constructively while recognizing that in the wrong hands it might be abused (Garfield, 1999). As long as scientists publish articles containing lists of cited references, it will be possible to calculate impact factors (Garfield, 2001). Garfield (2004) has also stated that "it has been demonstrated that on line access improves both readership and citation impact". The same impact factor can indicate the 'influence' and 'performance' of e-print archives they make among scientists.

Sen et al. (1989) had calculated Impact Factors of journals which are not included in JCR. The calculation is based on three factors:

1. the number of citable items published in the journal during years (Y-1) and (Y-2), say y1 and y2 respectively;
2. the number of times those items are cited in year Y in SCI journals, say x1; and
3. the number of times those items are cited in year Y in the journal X itself, say x2;

3. E-print Service

arXiv is an e-print service in the fields of Physics, mathematics, non-linear science, computer science, quantitative biology and statistics. arXiv is owned, operated and funded by Cornell University, a private not-for-profit educational institution. arXiv is also partially funded by the National Science Foundation, USA (arXiv.org, 2008). It was the brainchild of Paul Ginsparg, a physicist. The contents of arXiv conform to Cornell University, USA academic standards. It receives about 10,000 downloads per hour on the main site alone (there are a dozen mirror sites), is an essential resource for research physicists. arXiv's high level of usage by both authors and readers makes it an excellent database for analysing research trends as well as an important test-case for the OA literature (Brody and Harnad, 2004a). The categorised services of the present arXiv.org have helped scientists to look in to items of his/her interest. The categories are divided into five main categories, viz. Physics, Mathematics, Nonlinear Sciences, Computer Science, and Quantitative Biology.

The Physics category is again categorised in to 12 sub-classes as follows:

1. AstroPhysics (astro-ph)
2. Condensed Matter (cond-mat)
3. General Relativity and Quantum Cosmology (gr-qc)
4. High Energy Physics - Experiment (hep-ex)
5. High Energy Physics - Lattice (hep-lat)
6. High Energy Physics - Phenomenology (hep-ph)
7. High Energy Physics - Theory (hep-th)
8. Mathematical Physics (math-ph)
9. Nuclear Experiment (nucl-ex)
10. Nuclear Theory (nucl-th)
11. Physics (Physics)
12. Quantum Physics (quant-ph)

4. Objectives of the study

The present paper is an outcome of a study which attempts to calculate the Impact Factors and immediacy indices for the e-print archives uploaded in the Physics sub-classes of arXiv.org without the first factor x_2 for the year 2007. JCR-2007 was made use of to elicit the latest Impact Factors of some journals.

5. Materials and methods

The citations received during 2002 – 2007 in Science Citation Index (SCI) Database are used as the base data for calculating the above parameters. There is no direct search mechanism for citations received for these categories. Search mechanism and analysis are somewhat different from the direct search in Web of Sciences or Web of Knowledge. For example, the search query used for retrieving the citations received to the 'Condensed Matter (cond-mat)' category of Physics for the year 1997 in the 'cited author/reference' field of SCI is as follows.

A*-COND-MAT97-* OR B*-COND-MAT97-* OR C*-COND-MAT97-* OR D*-COND-MAT97-* OR E*-COND-MAT97-* OR F*-COND-MAT97-* OR G*-COND-MAT97-* OR H*-COND-MAT97-* OR I*-COND-MAT97-* OR J*-COND-MAT97-* OR K*-COND-MAT97-* OR L*-COND-MAT97-* OR M*-COND-MAT97-* OR N*-COND-MAT97-* OR O*-COND-MAT97-* OR P*-COND-MAT97-* OR Q*-COND-MAT97-* OR R*-COND-MAT97-* OR S*-COND-MAT97-* OR T*-COND-MAT97-* OR U*-COND-MAT97-* OR V*-COND-MAT97-* OR W*-COND-MAT97-* OR X*-COND-MAT97-* OR Y*-COND-MAT97-* OR Z*-COND-MAT97-* OR <ANON>*-COND-MAT97-*

6. Inferences

The number of open archives is rising steadily, and new publishing models are rapidly evolving to test new ways to increase readership and access. Proponents of open access claim the increasing influence of e-prints as they achieve higher readership through free access. Scholars in diverse disciplines are adopting open-access practices and being rewarded for it. Table 1 presents last five years uploads to different sub-categories of the Physics archives at arxiv.org. A total number of freely available 1,81,075 e-prints!!! 'Condensed Matter'; and 'AstroPhysics' categories have a maximum number of submissions (45211 and 44300 respectively) and Nuclear Experiment category is at minimum with 2255 submissions.

Table 1 Number of submissions to the Physics sub-categories of Arxiv.org (2003-2007)

	No. of e-prints added in					
Archive category	**2003**	**2004**	**2005**	**2006**	**2007**	**2003-2007**
astroph	7899	8242	8746	9271	10142	**44300**
condmat	8367	8953	9233	9243	9415	**45211**
grgc	1420	1554	1662	1742	1920	**8298**
hepex	771	885	854	848	862	**4220**
heplat	575	586	663	504	607	**2935**
hepph	3963	4136	3915	3949	4012	**19975**
hepth	3276	3357	3239	3295	3438	**16605**
mathph	747	874	951	937	1170	**4679**
nuclex	323	474	460	441	557	**2255**
nuclth	1156	1142	1064	1049	1114	**5525**
Physics	1619	2086	2759	3157	3554	**13175**
quantph	2438	2605	2859	2959	3036	**13897**
Total	**32554**	**34894**	**36405**	**37395**	**39827**	**181075**

astro-ph = AstroPhysics; **cond-mat** = Condensed Matter; **gr-qc** = General Relativity and Quantum Cosmology; **hep-ex** = High Energy Physics - Experiment; **hep-lat** = High Energy Physics - Lattice; **hep-ph** = High Energy Physics - Phenomenology; **hep-th** = High Energy Physics - Theory; **math-ph** = Mathematical Physics; **nucl-ex** = Nuclear Experiment; **nucl-th** = Nuclear Theory; **Physics** = General Physics; and **quant-ph** = Quantum Physics

Figure 1 shows the comparison of the increasing trend of Physics archives uploaded in Arxiv.org and the citations received to them for the same period. The sudden growth in number of citations from 2002 to 2003 shows their impact among physicists all over the world.

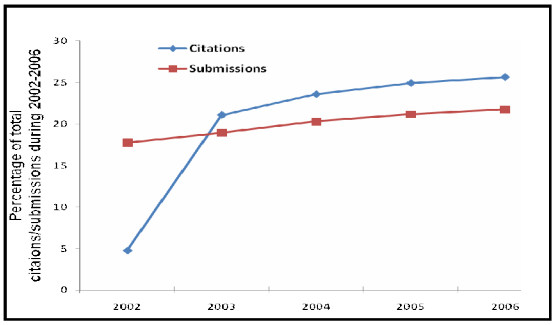

Figure 1 Year-wise percentage growth of total number of submissions to arXiv.org (2002-2006)

The Impact Factor for the year 2007 for the Physics classes of arXiv.org are computed and documented in Tables 2. The 'Nuclear Experiment' sub-class of Physics has high impact factor (0.63) followed by all high energy Physics categories and the 'General Physics' subfield makes less impact. (Current Science, high-impact factor journal of India has an Impact Factor 0.8 as per Journal Citation Reports-2007).

Table 2: Impact Factors calculated for the year 2007 for the Physics Classes of arXiv.org as per Science Citation Index

Archive category	Submissions in 2005 (a)	Submissions in 2006 (b)	Total submissions x1 = a+b	Citations in 2007 for 2005 submissions (y1)	Citations in 2007 for 2006 submissions (y2)	Total Citations Y = y1+y2	Impact Factor =x1/Y
astro-ph	8746	9271	18017	756	2542	3298	0.18
cond-mat	9233	9243	18476	745	2431	3176	0.17
gr-qc	1662	1742	3404	199	607	806	0.24
hep-ex	854	848	1702	314	583	897	0.53
hep-lat	663	504	1167	110	332	442	0.38
hep-ph	3915	3949	7864	1026	2483	3509	0.45
hep-th	3239	3295	6534	1011	2646	3657	0.56
math-ph	951	937	1888	91	111	202	0.11
nucl-ex	460	441	901	188	377	565	0.63
nucl-th	1064	1049	2113	257	477	734	0.35
physics	2759	3157	5916	122	218	340	0.06
quant-ph	2859	2959	5818	421	933	1354	0.23

astro-ph = Astrophysics; **cond-mat** = Condensed Matter; **gr-qc** = General Relativity and Quantum Cosmology; **hep-ex** = High Energy Physics - Experiment; **hep-lat** = High Energy Physics - Lattice; **hep-ph** = High Energy Physics - Phenomenology; **hep-th** = High Energy Physics - Theory; **math-ph** = Mathematical Physics; **nucl-ex** = Nuclear Experiment; **nucl-th** = Nuclear Theory; **physics** = General Physics; and **quant-ph** = Quantum Physics

Open archives are increasingly and immediately accessed through the Web and instantly get cited than the traditional journal articles. For researchers, the time of accessibility to the open archives has drastically reduced. The phenomena have caused the time for citing the open archives. As the impact factor for 2007 is calculated, the study has computed the immediacy index (a measure to evaluate how immediately the journal articles are cited) of the archives. The calculation of immediacy index is much easier than impact factor. In the context of journals, it is the ratio of the number of citations to the articles published in the calculating year with the number of articles published in the same year. So, in case of e-print archives, it can be defined as the ratio of the citations to the e-prints uploaded in the calculating year with the number of e-prints uploaded in the same year. Table-3 provides the calculated immediacy indices of the sub-classes of physics archives at arxiv.org for the year 2007. 'High Energy Physics –

Theory' and 'High Energy Physics – Phenomenology' are the two sub-categories with high immediacy index and 'Mathematical Physics' has comparatively low immediacy index.

Table-3: Immediacy Index calculated for the year 2007 for the Physics Classes of arXiv.org as per Science Citation Index

Archive category	No. of sub-missions in 2007(x)	Citations in 07 for 07 submissi-ons(y)	Immediacy index(x/y)
astroph	10142	565	0.06
condmat	9415	540	0.06
grgc	1920	205	0.11
hepex	862	87	0.10
heplat	607	50	0.08
hepph	4012	472	0.12
hepth	3438	626	0.18
mathph	1170	13	0.01
nuclex	557	53	0.10
nuclth	1114	83	0.07
physics	3554	47	0.01
quantph	3036	234	0.08

astro-ph = Astrophysics; **cond-mat** = Condensed Matter; **gr-qc** = General Relativity and Quantum Cosmology; **hep-ex** = High Energy Physics - Experiment; **hep-lat** = High Energy Physics - Lattice; **hep-ph** = High Energy Physics - Phenomenology; **hep-th** = High Energy Physics - Theory; **math-ph** = Mathematical Physics; **nucl-ex** = Nuclear Experiment; **nucl-th** = Nuclear Theory; **physics** = General Physics; and **quant-ph** = Quantum Physics

7. Conclusion

It can be believed that the physicists working in condensed matter physics and astrophysics etc. make use of the freedom to access the e-prints uploaded at arxiv.org. 'Nuclear Experiment' category could make great impact among the categories even though the number of submissions to the category is comparatively less among the categories. High-Energy Physics classes of e-print archives have made great impact among physicists and the immediacy factor is also high for the class compared to other classes. No wonder if the scientists and physicists turn up to freely available e-print archives and open archives for their R&D activities.

If the study incorporates the self citations to the e-print archives, the computed impact factors may go up and the categories may definitely compete with the physics journals with impact factors of even more than one. Further, the impact may go up if the study could have used new autonomous Open Archive web tools like 'Citebase', 'Citeseer', 'Google Scholar' etc. and of course the databases like 'Scopus' etc. The increase in number of uploads in e-print archive services are very high as against number of articles publishing in traditional journals. This factor definitely affects the impact factor of the categories studied.

Since the e-print archives are instant information feeding mechanism with an ephemeral effect, the e-print archive impact factor can be a divergent idea rather than e-print archive immediacy index. This connotation can be complemented if half-life of e-print archives has been calculated.

Although many authors believe that their work has a greater research impact if it is freely available, studies to demonstrate that impact are few [Antelman, 2004]. Once the impact and immediacy in citations of subject e-print archives are compared, scientists will submit their research documents in the open archive categories with high impact factors

and immediacy index. In that case, the continued emphasis on 'impact factors' will not misguid the readers as stated by Brunstein [2000]. Wherever the readers can make a comparison of sources they want to publish considering impact factors as the criteria, they may slant towards the high impact side.

References

1. **Antelman, Kristin.** Do Open-Access Articles Have a Greater Research Impact?, College & Research Libraries, 2004, pp. 372-382.
2. **Arxiv.org.** 2008. Available at http://arxiv.org
3. **Bjork, B.C**. Open access to scientific publications - an analysis of the barriers to change. Information Research, 2004, 9(2), paper 170, Available at http://informationr.net/ir/9-2/paper170.html.
3. **Bjorneborn, L., and Ingwersen, P.** Perspectives of webometrics. Scientometrics, 2001, 50(1), pp. 65-82.
5. **Brody, T.** Citebase Search: Autonomous Citation Database for e-Print Archives. In Proceedings of the SINN 03 conference on Worldwide Coherent Workforce, Satisfied Users - New Services For Scientific Information, Oldenburg, Germany, September 2003, Available at http://physnet.physik.uni-oldenburg.de/projects/SINN/sinn03/proceedings/brody/brody_sinn03_paper.pdf.
6. **Brody, T**. Citation Analysis in the Open Access World, 2004. http://eprints.ecs.soton.ac.uk/10000/01/tim_oa.pdf (Accessed on 10/11/2008)
7. **Brody, T., and Harnad, S. Earlier** Web Usage Statistics as Predictors of Later Citation Impact, 2004a, http://www.ecs.soton.ac.uk/~harnad/Temp/timcorr.doc. (Accessed on 10/11/2008)
8. **Brody, T., and Harnad, S.** Using Web Statistics as a Predictor of Citation Impact. 2004b. Author eprint available at http://www.ecs.soton.ac.uk/~harnad/Temp/timcorr.doc. (Accessed on 10/11/2008)
4. **Brody, T., Stamerjohanns, H., Harnad, S., Gingras, Y., Vallieres, F., And Oppenheim, C.** The effect of Open Access on Citation Impact. In Proceedings of the National Policies on Open Access (OA) Provision for University Research Output: an International meeting. Southampton University, Southampton UK. 19 February 2004. Available at http://opcit.eprints.org/feb19oa/brody-impact.pdf. (Conference proceedings)
9. **Brunstein, J**. End of Impact Factors. Nature, 2000, 403(6769), pp. 478-478.
9. **Correlation Generator** Available at http://citebase.eprints.org/analysis/correlation.php Generates a graph (or table) of the correlation between citation impact and usage impact from the Citebase database. (Accessed on 10/11/2008)
11. **Garfield, E.** Journal Impact Factor - A Brief Review. Canadian Medical Association Journal, 1999, 161(8), pp. 979-980.
13. **Garfield, E.** Impact Factors, and Why They Wont Go Away. Nature, 2001, 411(6837), pp. 522-522.
10. **Garfield, E.** American Scientist Open Access Forum listserv, 2004, 18 January 2004. http://users.ecs.soton.ac.uk/harnad/Hypermail/Amsci/3431.html (Accessed on 10/11/2008)
15. **Harnad, S.** Research access, impact and assessment. Times Higher Education Supplement, 2001.1487, p. 16.

16. **Harnad, S., Brody, T., Vallieres, F., Carr, L., Hitchcock, S., Gingras, Y, Oppenheim, C., Stamerjohanns, H., And Hilf, E.** The Access/ Impact Problem and the Green and Gold Roads to Open Access. Serials Review, 2004, 30. Available at http://www.ecs.soton.ac.uk/ ~harnad/Temp/impact.html.

11. **Hitchcock, S.** The effect of open access and downloads ('hits') on citation impact: a bibliography of studies. 2004. http:// opcit.eprints.org/oacitation-biblio.html. (Accessed on 10/11/2008)

17. **Kurtz, M.J**. Restrictive access policies cut readership of electronic research journal articles by a factor of two. In Proceedings of the National Policies on Open Access (OA) Provision for University Research Output: an International meeting. Southampton University, Southampton UK. 19 February 2004. Available at http://opcit.eprints.org/ feb19oa/kurtz.pdf. (Conference proceedings)

20. **Lawrence, S.** Online or Invisible?. Nature, 2001, 411, p. 521.

21. **Prakasan, E.r., Anil Sagar, Anil Kumar, Lalit Mohan, Gaderao, C. R., Surwase, Ganesh, Kadam, Sandeep, Kalyane, V. L., And Vijai Kumar.** Nuclear Theory and Nuclear Experiment E-print Archives: Science Citation Index - based Analysis. In Proceedings of the 22nd Annual Convention & Conference on Digital Information Exchange: Pathways to Build Global Information Society (SIS 2004), 2004a, pp. 172-183, IIT Madras, Chennai, 22-23 January 2004,. Available at http:// eprints.rclis.org/archive/00001244/ (Conference proceedings)

22. **Prakasan, E. R., and Kalyane, V. L.** Citation analysis of LANL High-Energy Physics E-Prints through Science Citation Index (1991-2002). 2004b. Author e-print, E-LIS, 26 August 2004. Available at http://eprints.rclis.org/ archive/00002200/

23. **Rousseau, Ronald.** The proceedings of the first and second international conferences on bibliometrics, scientometrics and informetrics: A data analysis. In Proceedings of the Sixth Conference of the International Society for Scientometrics and Informetrics, Jerusalem, Israel, 1997. (Conference proceedings)

24. **Sen, B.k., Karanjai, A., And Munshi, U.M.** A method for determining the impact factor of a non-SCI journal. Journal of Documentation, 1989, 45(2), pp.139-141.

26. **Tomer, C**. A statistical assessment of two measures of citation: the impact factor and the immediacy index. Information Processing and Management, 1986. 22 (3), pp. 251-258.

27. **Youngen, G.K**. Citation Patterns to Electronic Preprints in the Astronomy and Astrophysics Literature. Library and Information Services in Astronomy, 1998a, 53.

28. **Youngen, G.K.** Citation Patterns of the Physics Preprint Literature with Special Emphasis on the Preprints Available Electronically, Author e-print, UIUC Physics and Astronomy library, c. 5 November 1998, presented at ACRL/STS on 6/29/97, 1998b. http:// gateway.library.uiuc.edu/phx/preprint.html (Accessed on 10/11/2008)

About Authors

Mr. E R Prakasan, Scientific Assistant, International Nuclear Information System (INIS) Section, Scientific Information Resource Division, Knowledge Management Group, Bhabha Atomic Research Centre, Mumbai.
E-mail: prak@barc.gov.in or erpbarc@gmail.com.

Mr. Anil Sagar, Scientific Information Resource Division, Bhabha Atomic Research Centre, Mumbai.
E-mail: anilsagarbarc@yahoo.com

Mr. Anil Kumar, Scientific Officer, Scientific Information Resource Division, Bhabha Atomic Research Centre (BARC), Trombay, Mumbai.
E-mail: anilbarc72@rediffmail.com

Mr. V L Kalyane, Scientific Officer 'F',
Scientific Information Resource Division, Bhabha Atomic Research Centre (BARC), Trombay, Mumbai.
E-mail: vlk@barc.gov.in

Dr. Vijai Kumar, Associate Director,
Knowledge Management Group and Head, Scientific Information Resource Division at BARC, Mumbai.
E-mail: vijai@barc.gov.in