

## WEB LOG ANALYSIS OF E-JOURNAL USAGE AND SCHOLARLY COMMUNICATION: A CASE STUDY OF E-JOURNAL (FULL-TEXT) DOWNLOAD PATTERNS OF NAL SCIENTISTS AND ENGINEERS

*R Guruprasad<sup>1</sup>; Khaiser Nikam<sup>2</sup>; M Gopinath Rao<sup>3</sup> and Vidyadhar Y Mudkavi<sup>4</sup>*

At present most scientists have access to full-text e-journals. In most cases, this facility is provided right at the desktop. In this paper, we present a case study of full-text e-journal use patterns among the scientists and engineers at the National Aerospace Laboratories (NAL), a constituent of the Council of Scientific and Industrial Research (CSIR). The facility at NAL is provided right at the desktop through the NAL-CSIR-NISCAIR e-conglomerate. National Institute of Science Communication and Information Research (NISCAIR) provides e-access to more than 4040 world-class e-journals to all science and technology personnel of the CSIR fraternity. This CSIR-NISCAIR initiative allows any scientist in any CSIR Laboratory to access this electronic information to keep abreast of the technological developments in his / her area of specialization. The analysis of data of the full-text e-journal use patterns covers the period 2005 to 2007. The major findings highlighted are: (a) the mean number (per-month) of full-text downloads for the above three years was found to be different through Kruskal Wallis test of 'One Way Analysis of Variance' at 1% level of significance and (b) when the Chi-Square test was applied to test whether there is independence between the years and the publishers, the calculated value of Chi-Square was 845, which is highly significant. Hence we conclude that for the full-text downloads data, the years and the publishers are not independent. This Chi-Square test was carried out for only those publishers (8 in number) for which the data was available for all the three years (2005-2007).

**KEYWORDS:** Scholarly communication, Electronic information usage patterns, Scientists, Scientific communication, Electronic journals, NAL, CSIR, NISCAIR, Full-text e-journal download statistics.

- 1 Scientist, Knowledge and Technology Management Division, National Aerospace Laboratories (CSIR), PB No.1779, Airport Road, Bangalore. & Ph.D. Research Scholar, Dept. of Studies in Library and Information Science, University of Mysore. E-mail: gprasad@nal.res.in.
- 2 Chairperson, Professor and Research Guide, Department of Studies in Library and Information Science, University of Mysore, Manasagangotri, Mysore 570 006.
- 3 Professor, College of Agriculture, Dept. of Statistics, University of Agricultural Sciences, GKVK, Bangalore.
- 4 Head, Computational and Theoretical Fluid Dynamics Division, National Aerospace Laboratories (CSIR), NWTC, Belur Campus, NAL, Bangalore. E-mail: vm@ctfd.cmmacs.ernet.in.

## 1 OBJECTIVES AND SCOPE OF THIS STUDY

The main objective of this study is to examine the online usage patterns of electronic journals or full-text download patterns of e-journals by aerospace engineers and scientists through a technologically innovative method, such as, web log analysis. In fact, these statistically aided tools are proving very useful in studying the online e-user's information seeking behaviour. These methods also aid in understanding the searching and browsing behaviour of e-journal users. In fact, some researchers have adopted log analysis to examine the use of electronic journals in terms of both volume and patterns of use. After 2005, 2006 another method to study the information seeking behaviour of users called the Deep Log Analysis is becoming popular. This method overcame some of the major pitfalls of the Web Log Analysis method.

We present a study of the full-text e-journal download patterns of aerospace scientists and engineers (at NAL) during the period 2005-2007. Even though 14 e-journals are available for e-access to every NAL scientist right at his/her desktop, only 8 publishers data for the years 2005-2007 have been considered for this study as the other data were not complete. A chi-square test was applied to see whether there is independence between the years and the publishers. The calculated value of Chi-Square was 845, which is highly significant. Hence, our conclusion is that for the full-text downloads, the data, years and publishers are not independent. The mean number (per-month) of full-text downloads for the three years (2005-2007) was found to be different applying Kruskal Wallis test of 'One Way Analysis of Variance' at 1% level of significance. The trend appears to be that more and more scientists would download e-journals right on their desktop computers through their e-conglomerate. This could lead to innovative information seeking approaches and to get more authentic information about the use of online electronic journals in terms of both volume and patterns of use.

A review of the literature in this area indicates that there is hardly any Indian study for ascertaining the e-journal usage patterns among aerospace scientists and engineers. Hence, the value of this study which used techniques like Web Log Analysis where the data will always remain 'unfiltered', 'pure' and without any 'human intervention'.

## 2 REVIEW OF LITERATURE

Jansen, Spink, and Saracevic (2000) [1], analyzed transaction logs containing 51,473 queries posed by 18,113 users of Excite, a major Internet search service.

## Web Log Analysis of E-Journal Usage and Scholarly Communication

The data provided were on: (i) sessions - changes in queries during a session, number of pages viewed, and use of relevance feedback, (ii) queries - the number of search terms, and the use of logic and modifiers, and (iii) terms - their rank/frequency distribution and the most highly used search terms. The authors then shift the focus of analysis from the query to the user to gain insight to the characteristic of the Web user. With these characteristics as a basis, the authors then conducted a failure analysis, identifying trends among user mistakes. The conclusion gives a summary of findings and a discussion of the implications of these findings.

Earlier the above-cited authors (1999) [2] conducted a transaction log analysis of the same data. Approximately 2,500 (about 5%) of these queries were from the use of relevance feedback. Given the high level of research activity and historical success of relevance feedback in assisting users in locating relevant information, this is a surprisingly small percentage of usage. In order to investigate this phenomenon on the Web, the authors analyzed users sessions that contained relevance feedback queries. Next they identified states and patterns in these sessions. They also attempted to classify the sessions as successfully or not. This analysis provided insight on the use of relevance feedback on the Web, its success or failure, and why it is so seldom used. The authors relate their conclusions to system design of information retrieval systems on the Web.

Spink and Saracevic (1997) [3] investigated the sources and effectiveness of search terms used in mediated on-line searching under real-life (as opposed to laboratory) circumstances. A stratified model of information retrieval (IR) interaction served as a framework for the analysis. For the analysis, the on-line transaction logs, videotapes, and transcribed dialogue of the presearch and on-line inter- action between 0 users and 4 professional intermediaries were used.

Five sources of search terms were identified: (1) the users' written question statements, (2) terms derived from users' domain knowledge during the interaction, (3) terms extracted from retrieved items as relevance feedback, (4) database thesaurus, and (5) terms derived by intermediaries during the interaction. Distribution, retrieval effectiveness, transition sequences, and correlation of search terms from different sources were investigated. Search terms from users' written question statements and term relevance feedback were the most productive sources of terms contributing to the retrieval of items judged relevant by users. Implications of the findings are discussed.

Kantor and Saracevic [4], in anticipation of the explosive growth of digital libraries, undertook a complex study to evaluate 21 diverse services at 5 major academic research libraries. This work stands as a model for evaluation of digital libraries, through its focus on both the costs of operations and the impacts of the services that those operations provide. The data have been analyzed using both statistical methods and methods of Data Envelopment Analysis. The results of the study, which are presented in detail, demonstrate that a cross-functional approach to library services is feasible. They also highlight a new measure of impact, which is a weighted logarithmic combination of the amount of time that users spend interacting with the service, combined with a Likert-scale indication of the value of that service, in relation to the time spent. The measure derived, incorporating information obtainable from the user, together with information which is readily available in server/client logs, provides a means for transferring these measurement principles to the Digital Library environment

Rusch-Feja and Siebeky (1999) [5], of the Max Planck Society, a German organization for basic research, conducted a survey of researchers' use and acceptance of e-journals from April 15-May 15, 1999. The results show a significantly high acceptance of e-journals and an unwillingness to return to print versions only. The frequency of use of e-journals from four scholarly publishers was evaluated. The researchers also rated the advantages and disadvantages of e-journals related to various aspects, such as currency, ease of access, timeliness, upto dateness of information, additional searching mode, etc. The questionnaire also allowed capture of additional information, such as which additional journals were desired in electronic form, which services could be done without in case of budget restrictions, what information might be felt necessary for additional assistance in using e-journals, etc. Data from publisher-provided usage and transaction statistics shed more light on the distribution of use among the inter- and cross-disciplinary fields of research within the 84 Max Planck Institutes and additional Working Groups and Research Centers. Subject-oriented comparisons between the researchers' use in the biomedical section, the chemical-physical-technical section, and the humanities section of the Max Planck Society are drawn. A review of the recommendations resulting from the survey, as well as a suggestion for expansion of the information provision structures through establishment of a new Center for Information Management within the Max Planck Society, are given.

Nicholas, Huttington, Dobrowolski, Rowlands, Jamali and Polydoratou (2005) [6], say that the publication age or date of documents used (or not used) has long fascinated researchers and practitioners alike. Much of this fascination can be attributed to the weeding opportunities the data is thought to provide for libraries in their never-ending battle to find the space to accommodate their expanding

collections. In general journal article age studies have shown an initial increase in use/citation, then a gradual or sharp decline, depending on the discipline concerned. This characteristic has been termed obsolescence or decay and was largely measured, in the absence of accurate journal usage/borrowing data, by citations. In the sciences the decay rate was shown to be the greatest. This was largely attributed to the rapid obsolescence of much of the scientific content. New research findings, methods or ensuing events rendered the material obsolescent. Of course, when reviewing the data we need bear in mind that citation studies reveal “use” by authors, whereas library loans or downloads represent actual use by readers, and it is readers that libraries and digital libraries principally target.

Morris (1999) [7] noted that every Website has a different set of goals, but there's one thing we all have in common: We want more traffic! Although a sure-fire way to build Web site traffic quickly remains as elusive as a sure-fire way to predict stock prices, there are some tried-and-true methods that can help you build your Web site traffic slowly but surely.

The ambitious site owner will use various promotional tactics on an ongoing basis, but this article is not about any one traffic-building technique. It is about using the Web server log files to direct efforts and measure success. One may not have time to do everything, so the focus should be what works, Careful analysis of the information in the log files gives more promising traffic-building ideas, and also assist in measuring which ones live up to their promise.

Jamali, Nicholas and Huntington (2005) [8], provide a review of the log analysis studies of use and users of scholarly electronic journals. The advantages and limitations of log analysis are described and then past studies of e-journals' use and users that applied this methodology are critiqued. Those aspects of online journals' use and user studies that log analysis can investigate well and those aspects that log analysis cannot disclose enough information about are highlighted. The review indicates that although there is a debate about reliability of the results of log analysis, this methodology has great potential for studying online journals' use and their users' information seeking behaviour.

Nicholas, Huntington, Jamali, Tenopir (2006) [9] describe the early findings of an exploratory deep log analysis of journal usage on OhioLINK, conducted as part of the MaxData project funded by the US Institute of Museum and Library Services. OhioLINK, the original “big deal”, provides a single digital platform of nearly 6,000 full-text journal for more than 600,000 people in the state of Ohio. The purpose of the work was also to test a new method of analysing online information user behaviour – deep log analysis. The raw server logs were obtained for the period June 2004 to December 2004. For this exploratory study one month (October) of the on-campus usage logs and seven months of the off-

campus transaction logs were analysed. The paper presents a number of usage analyses including: number of journals used, titles of journals used, use over time, a returnee analysis and a special analysis of subject, date and method of access.

Connaway and Snyder, (2005) [10] say that other studies analyse e-book usage reports and user surveys to better understand what e-books are being accessed and how users perceive their e-book experience. Another possible methodology for identifying the e-books that users are accessing and how they are accessing them is transaction log analysis, which can be integrated with other data collection methods, including those mentioned above. This type of analysis allows the researcher to unobtrusively identify user search and retrieval patterns and to evaluate systems. Transaction log analysis provides both macro analysis, an analysis of aggregate use data and patterns, and microanalysis, an analysis of individual search patterns. The data can be used to develop systems and services based on user behavior. There are limitations to transaction log analysis. The users may not be identifiable; therefore, it is usually impossible to associate user demographics with usage patterns or to determine where they access the resource or how and why they use the resource. The transaction logs provide massive amounts of data to manipulate. The types of possible analyses are dependent upon the data collected and stored in the system and when systems change, the data collected may change or no longer be available for analysis.

Taha (2004) [11] write that e-journals have revolutionised the scholarly communication process and changed the way that the researchers seek and exchange scientific information. The UAEU library, like many other libraries in the West, is going through a paradigm shift; switching from paper-based to digital formats. This work aims at better understanding the pathways taken when the researchers connect to favourite e-journals by analysing the generated transaction logs. Examples on electronic uses were derived from ScienceDirect™ database transaction logs of the articles that appeared in 2003. The data analysis revealed that the sheer predominance of electronic access to e-journals was in the domains of environmental, materials, medical, and business studies since these areas are gaining a significant priority to UAEU research funding and graduate programmes. The implication of the choice and heavy use of favourite e-journals is discussed with particular reference to the future of e-information services, and subscription to core e-journals. This study argues that the emergence and fit of e-journals into library services has significantly improved efficiency of information practices as well supported a range of research activities and trends.

Mariner (2002) [12] collected raw data for measuring the use of web sites<sup>1</sup>. The raw web log files were extracted and read to learn what pages and resources the patrons were taking advantage of on the web site.

---

1 See Thomas Dowling's "Lies, Damned Lies, and Web Logs", netConnect, Spring 2001, for an overview of what the web logs can and can't tell you).



## Web Log Analysis of E-Journal Usage and Scholarly Communication

There are several ways the content of web log files can be extracted. In an in-house model, it is reasonably easy to dump the tab-delimited text into a database or spreadsheet and study the numbers. However, an informal poll of 150 attendees at the 2001 Library and Information Technology Association (LITA) National Forum found that only two were using an in-house model. More commonly, either a free or a commercial product crunches the raw log files into friendly and easy-to-read graphs and tables that show the use of the library web site.

Rozic-Hristovski, Hristovski, Todorovski (2002) [13], The Central Medical Library (CMK) at the Faculty of Medicine, University of Ljubljana, Slovenia, started to build a library Website that included a guide to library services and resources in 1997. The evaluation of Website usage plays an important role in its maintenance and development. Analyzing and exploring regularities in the visitors' behavior can be used to enhance the quality and facilitate delivery of information services, identify visitors' interests, and improve the server's performance. The analysis of the CMK Website users' navigational behavior was carried out by analyzing the Web server log files. These files contained information on all user accesses to the Website and provided a great opportunity to learn more about the behavior of visitors to the Website. The majority of the available tools for Web log file analysis provide a predefined set of reports showing the access count and the transferred bytes grouped along several dimensions. In addition to the reports mentioned above, the authors wanted to be able to perform interactive exploration and ad hoc analysis and discover trends in a user-friendly way. Because of that, they developed their own solution for exploring and analysing the Web logs based on data warehousing and online analytical processing technologies. The analytical solution developed proved successful, so it may find further application in the field of Web log file analysis.

Philip (2004) [14] study reports an analysis of referral URL data by the Cornell University IP address from the American Chemical Society servers. The goal of the work was to better understand the tools used and pathways taken when scientists connect to electronic journals. While various methods of referral were identified in this study, most individuals were referred infrequently and followed few and consistent pathways each time they connected. The relationship between the number and types of referrals followed an inverse-square law. Whereas the majority of referrals came from established finding tools (library catalog, library e-journal list, and bibliographic databases), a substantial number also originated from generic Web searches. Scientists are also relying on local alternatives or substitutes such as departmental or personal Web pages with lists of linked publications. The use of e-mail as a method to refer scientists directly to online articles may be greatly underestimated. Implications for the development of redundant library services such as e-journal lists and the practice of publishers to allow linking from other resources are discussed.

Nicholas, Huttington, Jamali, and Whatkinson (2006) [15] used deep log analysis (DLA) techniques, a more sophisticated form of transaction log analysis, to demonstrate what usage data can disclose about information seeking behaviour of virtual scholars – academics, and researchers. DLA works with the raw server log data, not the processed, pre-defined and selective data provided by journal publishers. It can generate types of analysis that are not generally available via proprietary web logging software because the software filters out relevant data and makes unhelpful assumptions about the meaning of the data. DLA also enables usage data to be associated with search/navigational and/or user demographic data, hence the name ‘deep’. In this connection the usage of two digital journal libraries, those of Emerald Insight and Blackwell Synergy are investigated. The information seeking behaviour of nearly three million users is analyzed in respect of the extent to which they penetrate the site, the number of visits made, as well as the type of items and content they view. The users are categorized by occupation, place of work, type of subscriber (“Big Deal”, non-subscriber, etc.), geographical location, type of university (old and new), referrer link used, and number of items viewed in a session.

Griffiths, Hartley, and Willson (2002) [16] write that investigation of how people use electronic information resources is important as a means of understanding how to make systems more usable and as a means of understanding information-seeking behaviour of end users. Important developments have been the use of transaction logging data and of protocol analysis. However a trend in understanding information-seeking behaviour and in information system evaluation has been the move from controlled experiments to the study of information seeking in a natural setting. The collection of transaction logging data of end users searching full text electronic information sources in a natural setting poses numerous challenges and raises methodological issues. The paper explores the challenges and describes a novel approach that was developed to address the limitations of traditional quantitative methodologies as applied to CD-ROM databases. The transaction logs for the research project were required to provide more qualitative data, capturing the search from start to finish as experienced and seen by end users.

The paper discusses transaction log analysis and protocol analysis as methods of data collection before outlining the novel methodology in detail, illustrating its operation with some examples of the data collected and concluding with a brief discussion on its use in a number of projects. The intention of the paper is to concentrate on the methodology itself rather than report results of specific research. Appendices are included which provide system requirements/technical detail and contact information of suppliers.



## Web Log Analysis of E-Journal Usage and Scholarly Communication

Harley and Henke (2007) [17] point out that almost every American research university and library has made significant investments in digitizing its intellectual and cultural resources and making them publicly available. There is, however, little empirical data about how these resources are actually used or who is using them (Harley, 2007, [18]). Those who fund and develop digital resources have identified the general lack of knowledge about the level and quality of their use in educational settings as pressing concerns. As part of a larger investigation into use and users of digital resources (Harley et al., 2006, [19]),<sup>1</sup> an experimental analysis of two commonly-used methods for exploring the use of university-based web-based resources: transaction log analysis (TLA) and online site surveys. The article first provides an overview of these two methods, including their key challenges and limitations. Then follow a description of an implementation of TLA and online surveys in combination on two local sites and the results of that test, including an exploration of the surveys' response rates and bias. From that test, conclusions are drawn about the utility of these two methods and the particular analytic methods that may provide the most valuable and efficient results.

Gardis (2001) [20] emphasizes that measuring the usage of electronic resources is vital both for the collection development strategy and to plan an efficient training programme for users. This work aims to provide both a case study and a framework for future developments in measuring the performance of a digital library. The survey starting point has been the server log of the website of the Pediatrics and Adolescent Medicine Library of the Turin University. The log has been analyzed from September December 1999, March to May 2000 and July 2000 to January 2001. Parameters taken into account include the number of hits, users and single IP addresses in the time period, most requested pages, errors, percentage of users accessing the site from inside or outside the institution's building. Besides, using both Altavista and Fast Search Gardis has tried to determine which web sites point to the library. The library also delivers access to 160 e-journals (2000-10-30). For 94 of them usage reports are available. Where possible, the researcher has tried to assess the number of total users, the hits on table of contents and abstracts the number of full text downloads both in HTML and PDF format. Users of the website also access the following databases: Medline, CINAHL, Cochrane Database of Systematic Reviews, Best Evidence (all delivered by OVID). The usage of these resources has been analyzed for the period October 2000 to April 2001 with particular attention to: number of users for each resource, structure of the queries performed by users, search by subject, author, title, journal, and boolean operators.

Hao-Ren Ke, Kwakkelaar, Yu-Min Tai, Li-Chun, Chen (2002) [21] analysed usage of the Taiwan-based ScienceDirect OnSite E-journal system, one of the

largest and most heavily used full-text Science, Technology, and Medicine (STM) databases worldwide. The Internet and World Wide Web have introduced new and powerful ways for finding and sharing information. Web based electronic resources facilitate the exploration of user behavior far beyond what is possible in a print environment, as Web services can be configured to record (log) all relevant user transactions. This article analyzes transaction logs of the Taiwan ScienceDirect OnSite (SDOS) E-journal system, states the findings and interprets the results. The Taiwan SDOS hosts the bibliographic information and full-text articles of more than 1,300 journals published by Elsevier Science. There are an estimated 625,000 users for the Taiwan SDOS system. The system is hosted locally by Academia Sinica, a leading academic institution of Taiwan, and can be accessed by SDOS subscribers, including all major Taiwanese universities and research institutes.

Davis and Price (2005) [22] write that the design of a publisher's electronic interface can have a measurable effect on el- journal usage statistics. A study of journal usage from six COUNTER-compliant publishers at 32 research institutions in the United States, the United Kingdom and Sweden indicates that the ratio of PDF to HTML views is not consistent across publisher interfaces, even after controlling for differences in publisher content. The number of full text downloads may be artificially inflated when publishers require users to view HTML versions before accessing PDF versions or when linking mechanisms, such as CrossRef, direct users to the full text, rather than the abstract, of each article. These results suggest that usage reports from COUNTER-compliant publishers are not directly comparable in their current form. One solution may be to modify publisher numbers with 'adjustment factors' deemed to be representative of the benefit or disadvantage due to its interface. Standardization of some interface and linking protocols may solve these differences and allow for more accurate cross-publisher comparisons.

Huttington, Nicholas, and Jamali (2007) [23] in their paper argue that metrics can be generated from search transactional web logs that can help evaluate search engine effectiveness. Search logs from the BBC website were analysed and metrics extracted. Two search metrics — the time lapse between searches and the number of searches in a session — were developed to see whether they could measure search success or satisfaction. In all, 4 million search statements by 900,000 users were evaluated. The BBC search engine possessed a number of functional attributes which sought to improve retrieval and these were subjected to the two metrics to help determine how successful they were in practice. There was some evidence to support the proposition that the search outcome metrics did indeed indicate the effectiveness of engine functionality. The authors argue that this result is significant in that the identification of search outcome metrics will pave the way for assessing the effectiveness of site specific search engines and a greater understanding of the effectiveness of search engine functionality.

Moen and McClure (1997) [24] state that the assessment and evaluation of electronic networks and network-based resources is increasing in scope and application (Bertot & McClure, 1996a, 1996b; McClure and Lopata, 1996). Web server transaction log file analysis is a network-based assessment technique that is particularly useful when performed in conjunction with other ongoing evaluation activities. The investigators designed an experimental data collection technique to analyze usage of GILS records on a Web server at the Environmental Protection Agency (EPA). The analysis had four purposes: (1) to determine the overall Web site traffic including the location of users, the portions of the site accessed, and the number of document downloads; (2) to determine the use of the Web site GILS directory traffic including the location of users, portions of the site accessed, and number of document downloads (both hits and accesses); (3) to experiment with new log analysis techniques that go beyond domain, hit, and browser counts; and (4) to assist federal agencies that operate Web-based GILS servers to develop, implement, and maintain ongoing log file analysis.

Warren (2002) [25] states that in the past libraries had only a couple of service points for providing information; patrons could personally consult librarians at the reference desk or use the phone to contact the library. Now, libraries rely on their websites to provide access to e-journals, online databases, and other services. Although it is standard procedure for most librarians to count the number of questions they receive at the reference desk or via the phone, collecting statistics about library website usage is a more recent practice. Even if libraries collect web statistics, raw web logs must be transformed into a more “readable” format by appropriate software. Librarians need to know how to interpret the reports generated from such software in order to understand what web statistics can and cannot measure. Further, web statistics must be interpreted so that the data can impact decision-making practices within libraries. The Library & Information Services Branch at the National Institute of Environmental Health Sciences (NIEHS) has taken the initiative to integrate web log statistics into its overall management practices. This study shows that web log analysis produces information about the overall use of the NIEHS Library website, in addition to providing insight for collection development, marketing, and management issues within the library. Data for the paper was collected from two utilities: web log analysis software called Analog and a ClickCount script that counted the number of clicks on certain web links. The data was then summarized to highlight trends and provide a template for future analysis of the library’s website statistics.

### **3 SCIENCE AND SCIENTIFIC COMMUNICATION.**

Science is undergoing some fundamental changes [26]. Much of science is experiencing greater specialization, while, on the other hand, some parts of big science is getting even bigger. Also, a great deal of research is becoming more

multi-disciplinary. This has led to collaboration among universities, government, and industry all of which extend across national borders. Science education is becoming not only multi-disciplinary, but also collaborative, as more and more faculties teach across disciplines, departments and universities.

Learning is fundamental to science and communication is the heart of learning. Garvey [27], a psychologist at the Johns Hopkins University, summed up nearly two decades of scientific communication research by saying that “communication is the essence of science”. Fourteen independent studies conducted from 1958 to 1998 observed that scientists spend a large, and perhaps increasing, proportion of their time communicating. Recent studies place this proportion in the range of 50 to 60 percent on an average of scientist’s time as spent in communicating.

Traditional scientific patterns have evolved into a multitude of channels, including data and image transmissions, informal discussions, e-mails/messages, laboratory notes and technical reports, conference presentations and proceedings, journal articles, patents and books, to name a few. Each channel provides several distribution means and can involve a variety of media. All these channels of distribution result in a complex pattern of information flow.

Many studies conducted by the American Psychological Association (APA) for the National Science Foundation (NSF), during 1963-1968 and later in the 70’s led many to believe that electronic technologies could minimize redundancies and produce more efficient communication capabilities.

#### **4 SCHOLARLY SCIENTIFIC JOURNALS**

The first two scientific scholarly journals were believed to have started at about the same time, in the mid 1600s [26]. One was the *Le Journal des Scavans*, founded by M. de Sallo and the other was the *Philosophical Transactions*, a monthly journal of articles by the Royal Society of London to its members. In fact, Henry Oldenburg produced the first issue of a scientific journal, the *Philosophical Transactions of the Royal Society of London*, in 1665 [28]. By the end of the 17th century there were about 30 to 90 scientific and medical journals published worldwide and this number rose to about 750 by the end of the 18th century. Currently, there are about 80,000 to 100,000 scholarly journals published worldwide.

In another study Okerson [29] reviewed the history of e-journals and discussed a few examples from the early 1990s. She also counted the number of e-journals listed in two directories from 1991 to 1999. The number of titles grew from 27 in 1991, to 3,634 in 1997, and to 8,000 in 1999.

Finally, questions about a journal’s accessibility and readership can also hinge on the extent to which it allows readers free access to electronic versions.

## 5 ELECTRONIC JOURNALS

A good proportion of scientific scholarly journals are now available in electronic media. Most are digital replicas of traditional journals, with the majority being provided in both paper-based and electronic media and a few are exclusively electronic journals. Much of the debate on the future of electronic publishing [28] focuses on opportunities for readers, writers, and publishers. The accessibility of scholarly e-journals, their potentially lower production costs, the possibility of multimedia publication, and reference linking are treated as compelling features of the electronic medium that will enable them to thrive.

## 6 INFORMATION-SEEKING AND READING PATTERNS: *TRENDS*

Evidence suggests that the amount of reading and the time spent in reading have been relatively stable over the past 20 years, there have been some changes in the ways in which scientists identify the articles they read and there are appreciable differences in the sources of these articles. Surveys [26] from 1993 to 1998 show that scientists identify articles they read by browsing through journal issues or bound volumes. 62% of readings are identified in this way, by automated searches accounts for 12 %, by having other person tell them about the articles amounts to 11%, by using citations found in other articles, books etc. adds up to 9%, or by other means such as current awareness services, printed indexes, and so on fills the remaining 6%. The same study indicates that during the period 1993 to 1998, scientists surveyed averaged about 120 readings of scholarly articles per year. In general, reading has shifted from personal subscriptions to library-provided journals, due in large part to a decline in the number of personal subscriptions and to better library services.

There are a number of factors that influence information-seeking and reading patterns [26]. Variation among scientists' communication patterns is partially attributable to personal characteristics such as level of education and experience, and general communication capabilities, in addition to the nature of the discipline or area of specialization.. There are also situational factors as well, such as size of the organization, level of research funding, amount of funds available for information services, and availability and access to library services.

The authors noted that scientists read at least one article from an average of 18 scholarly journals. However, they tend to read only a few of these journals extensively and most of them sparsely. For example, across all journals read by scientists only five percent of them are read more than 25 times by a scientist (on an average) and about 80 percent are read less than 10 times. The amount of reading of a journal has a major bearing on whether it should be purchased, depending, of course, on the price compared with the cost of using alternative sources of the article. In the past, libraries have been the principal alternative to purchasing journals.

So, what do the various trends reflect? Since their birth in the 17<sup>th</sup> century, scientific scholarly journals have become the typical genre of publication and, for most fields of science, ‘the most inevitable, and the single most important channel of communication’ [26]. Over the last 40 years, numerous studies indicate that journals are extensively read; the information they contain is very useful for research, teaching and lifelong learning; and the information is valuable in terms of the favourable outcomes from its use.

## **7 USE AND USERS OF SCHOLARLY E-JOURNALS: TRANSACTION LOG ANALYSIS (TLA) APPROACH**

Jamali, *et al*, [9], highlight the advantages and limitations of the log analysis approach. Even though there is a debate about the reliability of the results of the log analysis, this methodology has immense potential for studying the use of online journals and their users’ information seeking behaviour. It is a well understood fact that finding about the usage patterns of scholarly journals has been important for both the librarians and the publishers for a very long time. Interest of libraries in the use of journals is two-fold. First, research and academic libraries spend the biggest portion of their acquisition budget on serials. Secondly, virtually all academic and research libraries are moving towards electronic access to journals. In such an environment, the users who have the world of knowledge at their fingertips are physically disappearing from the librarian’s view. Therefore, understanding the usage of e-journals, and the information seeking behaviour of users is of great importance both for libraries and publishers.

Before the advent of online journals [9], most of the studies on journal usage were based on (a) citation analysis, (b) re-shelving data or (c) questionnaires. All the three have their own limitations. Citation analysis does not represent all of journal usage as authors do not cite all the articles they read, and moreover, not every journal reader is an “author”. Re-shelving data are not accurate. In this case, it is not possible to distinguish between the use of individual articles or the whole journal. It also does not indicate use of personal subscriptions and the type of use. Questionnaire based studies rely heavily on what people think they do or might do – not what they actually do, and this could result in misinterpretations.

With the widespread use of computer and network technologies for facilitating access to scholarly journals, a new methodology has emerged for studying journal usage and scholarly information seeking behaviour. Computers record or log all user transactions in a plain text file known as a “transaction log”. Log files contain data about many of the details of the users’ interaction with the system. Hence, some researchers have adopted log analysis to find out about the use of e-journals in terms of both the volume and patterns of use.



### 7.1 Transaction-Log Analysis Methodology: Usefulness

Let us first try to understand as to why the Web Server Transaction Log Analysis Methodology plays an important role in understanding the e-journal full-text download patterns. Web server transaction log file analysis is a network-based assessment technique that is particularly useful when performed in conjunction with other ongoing activities [9]. Generally, the intent of the Web server log analysis is multi-purpose: First, one can determine the overall Web site traffic including the location of users, the portions of the site accessed, and the number of document downloads; Second, one can determine the Web site directory traffic including the location of users, portions of the site accessed, and the number of document downloads (both hits and accesses); third, one can experiment with developing new log analysis techniques that go beyond domain, hit, and browser counts; and finally, one can assist government agencies to develop, implement, and maintain ongoing log file analysis.

### 7.2 Transaction Log Analysis Technique

Web server log analysis technique generally involves a three-fold process that includes determining the types of information server administrator and decision makers need; developing a program that can parse through, manipulate, and present value-added information from the log files; and analyzing the information generated from the program. Web servers automatically generate four different log files: access logs (e.g. hits), agent log (e.g., browser, operating system), error log (e.g. download aborts), and referrer logs (e.g. referring links). These files are text files that can range in size from 1 KB to 100 MB, depending upon the traffic at a particular site. Distinction between a hit and an access is critical to understanding the type of data contained in these files. A hit is any file from a web site that a user downloads. Download of a Web page with 6 images on it accounts for 7 hits (6 images + 1 text page). An access (or a page hit) is an entire page download regardless of the number of images, sounds, or movies on the page. Download of a web page with 6 images accounts for only one access.

## 8 DEEP LOG ANALYSIS METHOD

Deep Log Analysis (DLA) overcomes pitfalls of TLA. Nicholas [30, 31, 32] and his colleagues in CIBER conducted a series of studies on Emerald and Blackwell e-journals in order to evaluate the impact of the Big Deal on users' behaviour and generally to find out digital journal's users' information seeking behaviour. Based on the experience gained from investigating consumer health logs, they developed a more sophisticated methodology called Deep Log Analysis (DLA). Some of the salient features of DLA are:

- More attention is paid to the users in their analysis and the authors highlighted the importance of returnees and bouncers (calculating repeat visits of the users to the same site).

The strength of DLA is due to the following features:

- Use of SPSS (statistical analysis package) to analyse raw logs instead of proprietary log analysis software. SPSS provides more flexibility and enables researchers to define their own variables and breakdowns;
- Enriching log data with demographic data, such as user data gathered from the subscription of publishers;
- Classifying users based through a combination of their demographic attributes and their usage data; and
- Paying special attention to returnees – users who come back to the use the service.

Three Deep Log Micro Analysis Techniques are [30]:

- The construction of a subgroup of users for which researchers can feel confident in regard to their geographical origin;
- The analysis of a subgroup of users for which users whose IP addresses are more likely to reflect the use of the same individuals; and
- The tracking and reporting of the use made by individuals rather than groups.

DLA methodology thus provides a better, more accurate, and fuller picture than what is possible with standard survey techniques and provides some very powerful types of analyzes not obtainable from the standard log analyzing software [32].

### **8.1 Some Findings of Log Based Studies:**

Log analysis has been applied for different purposes [9] such as assessing system performance, studying user's searching and browsing behaviours, investigating the effectiveness of Big Deal subscriptions, studying literature decay, etc. Digital journal platforms or libraries also have different features. These factors make it difficult to compare the results of the different studies and make generalizations. Nevertheless, there appears to be a considerable degree of concentration of the studies in the use of journals. According to Morse and Clintworth [33], just 20 per cent of titles accounted for nearly 60 per cent of usage. Another study [34] showed that a small number of heavy users can have an extremely large effect on the number of total downloads. Another study [35], showed that 4.9 per cent of a journal collection satisfied 44 per cent of downloads and, on the other hand, 59 per cent of the collection represented only 10 per cent

## Web Log Analysis of E-Journal Usage and Scholarly Communication

of the use of the collection. The effect of log analysis limitations, particularly the problems with caching and proxy servers, on this asymmetric pattern of use is yet to be investigated. Log studies also indicate a relative preference for PDF versions of articles to HTML versions among users [36, 34, 30]. Questionnaire studies confirm this preference and highlight the fact that most users do not like reading on the screen [37, 38, 39]. This indicates that users of e-journals probably choose a PDF version because it is more printer friendly and better for archiving.

Log studies have been particularly helpful in understanding the searching and browsing behaviour of users of e-journals. The findings of eJUST project showed that there were two major starting points for journal web visits, i.e. through journal home pages and through PubMed. Entering journal web sites through homepages usually led to either browsing contents or searching for an article. More users read full text right away instead of reading abstracts first to see if articles were of interest; however, users of certain journals requested abstracts before reading full text. Users tend to read full text after browsing contents. Either abstracts or full text views in HTML preceded requests for full text in PDF format. However, three very common seeking patterns were noted:

- Journal homepage – TOC – HTML full text – PDF full text;
- PubMed – HTML full text – PDF full text; and
- Journal homepage – search – HTML full text – PDF full text.

The findings indicated that most requests were for full text in HTML, which were then followed by request for the full text in PDF, as if the final goal of most visits was to take away a PDF version of an article. Also, the study revealed that library catalogues and bibliographic databases, which are both searching mechanisms, were the top two sources that led users to journals [40].

Although, there has been an ongoing debate on the pitfalls of web log analysis, some of the studies indicate that there are possibilities for improving the methodology. DLA methods developed by Nicholas and his colleagues at CIBER have opened a new horizon in studying e-journal use and users. Several steps can be taken to enrich the log data and obtain more robust data:

- for instance, who is a user, what is a hit, what represents success, etc.) re-align as necessary, and assess statistical significance;

- The raw data should be re-engineered to provide more powerful metrics and to ensure that data gathering is better aligned to organizational goals;
- Enrich the usage data by adding user demographic data (e.g. occupation, subject, specialism);
- Categorizing the users into smaller groups rather than looking at a broad picture of the usage and tracing the usage by some individual users as case studies could help achieve a deeper knowledge of usage patterns and user behaviour; and
- Finally, to strengthen the results of log analysis and test the findings, some questionnaire, interview or observational studies should be conducted to explain the information seeking behaviour of the users discovered in the logs.

To sum up, log analysis is clearly useful for certain kinds of studies, like shedding light on the format of the articles scientists read (PDF or HTML), the age of the articles (obsolescence), and the way scientists navigate to the required material (searching and browsing behaviour). However, it is not at all helpful at discovering the value and use of the articles retrieved, or about what lies behind expressed information seeking behaviour. So far log analysis has not been a very efficient technique for finding out about the differences of information seeking behaviour among users from different subjects, or about the effects of the status of users on their information seeking behaviour. These are the areas in which log analysis methods must be improved. The results of log analysis should be enhanced by a triangulation of the findings of studies with other methodologies (e.g. employing a combination of log analysis, questionnaire surveys and observation studies).

## **9 NATIONAL AEROSPACE LABORATORIES, BANGALORE AND ITS SCIENTISTS**

National Aerospace Laboratories (NAL) [41], a constituent of the Council of Scientific and Industrial Research (CSIR), is India's pre-eminent civil R&D establishment in aeronautics and allied disciplines. NAL's primary objective, as articulated in its new Vision Statement, is the "development of aerospace technologies with a strong science content and with a view to their practical application to the design and construction of flight vehicles". NAL is also required "to use its aerospace technology base" for general industrial applications.

NAL has a staff strength of about 1250 with about 400 full-fledged R&D professionals (over 100 Ph.D.'s). It is thus in a unique position to offer R&D support, expertise and services to both aerospace and non-aerospace sectors of the industry. Scientists at NAL have been provided facility of accessing 3316 international journals from about 11 international journal publishers by being part of the NAL-CSIR-NISCAIR E-Conglomerate, right at their desktops.

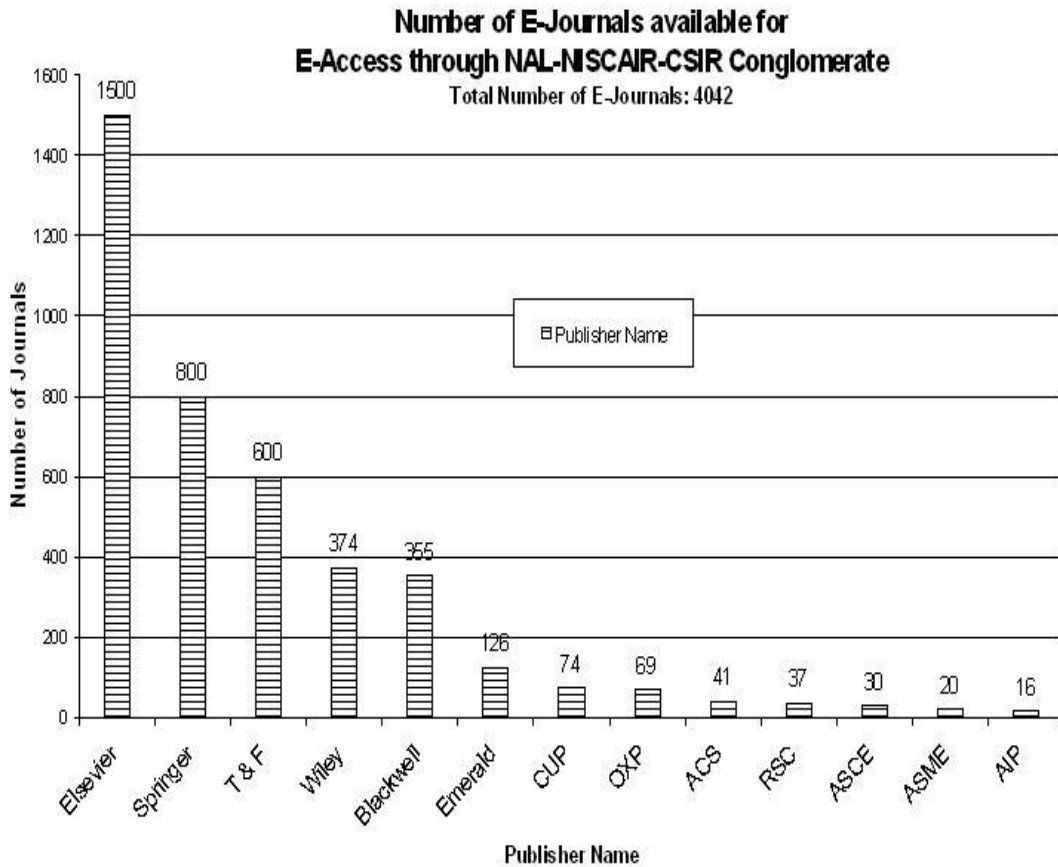
### 9.1 CSIR, NISCAIR and the E-Journals Conglomerate

CSIR) [42]—the premier industrial R&D organization in India was constituted in 1942 by a resolution of the then Central Legislative Assembly. Its Mission is to provide scientific industrial R&D that maximizes the economic, environmental and societal benefits for the people of India. Today it is one of the world's largest publicly funded R&D organizations having linkages to academia, R&D organizations and industry. CSIR's R&D portfolio embraces areas as diverse as Aerospace, Biotechnology, Materials, Energy, Chemicals and so on. Indeed, it spans A to Z of Indian Science.

The National Institute of Science Communication and Information Resources (NISCAIR) is one of the constituent units of CSIR in the area of information science [43]. Its mission is to become the prime custodian of all information resources on current and traditional knowledge systems in science and technology in the country, and to promote communication in science to diverse constituents at all levels, using appropriate technologies. One of the main mandates of NISCAIR is to provide formal linkages of communication among the scientific community in the form of research journals in different areas of science and technology (S&T). NISCAIR aims to provide access to more than 4042 world-class e-journals to all S&T personnel of the CSIR fraternity from their desktops, through pooling and sharing of resources. NISCAIR deals with 13 popular international publishers. The objectives of E-journals Consortia are: (a) to strengthen the pooling, sharing and electronically accessing the CSIR library resources, (b) to provide access to world S&T literature to CSIR labs, (c) to nucleate the culture of electronic access resulting into evolution of digital libraries.

Till date CSIR has entered into agreement with the 13 publishers to access about 4040 international journals across the labs. Details of the e-access is given below:

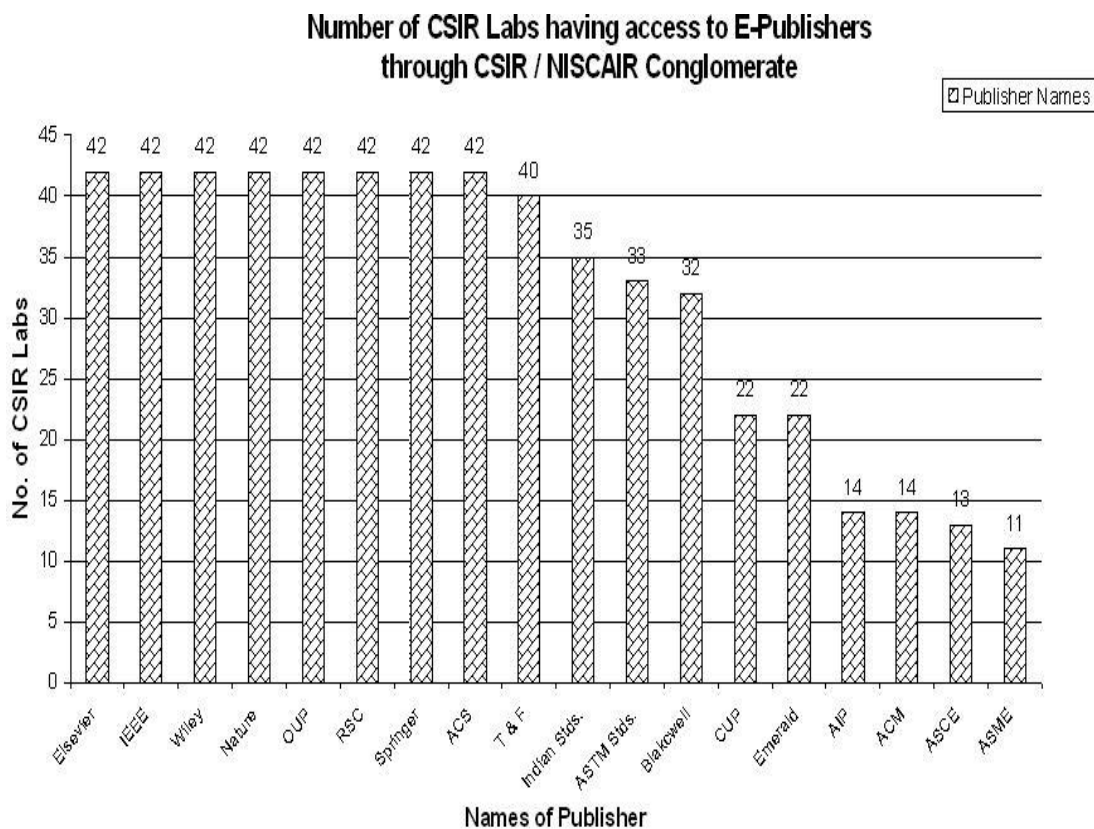
## 9.2 Full-text e-journals download patterns of NAL Scientists



**Fig. 1: Highlights the number of Scientific Journals available for E-Access through the CSIR-NISCAIR E-Conglomerate Data obtained from NISCAIR Server:**

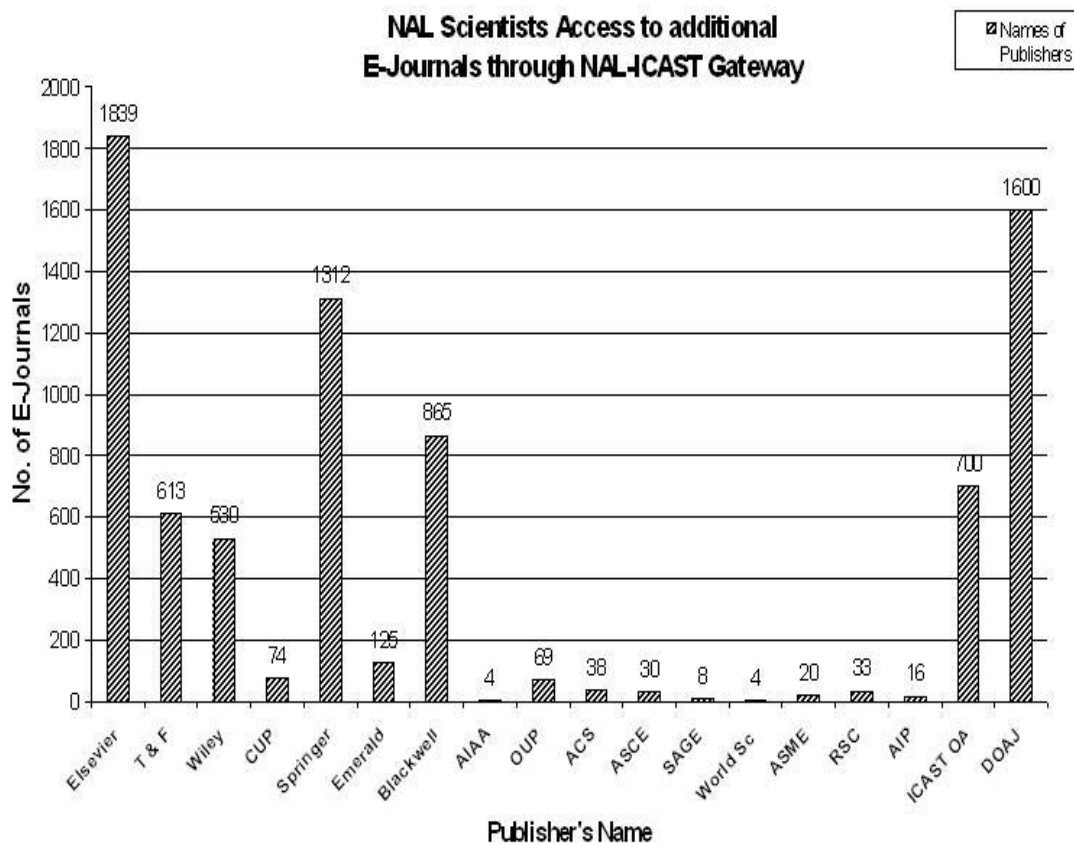
- The maximum number of e-journals for the conglomerate is from Elsevier, followed by Springer and T&F. Interestingly, the trend nearly follows exponential curve;
- Wiley and Blackwell e-journals are also available in good numbers;
- Journals from ASME and AIP are the lowest; and
- There are 12 publishers whose e-journals are available for e-access for the conglomerate





**Fig. 2: List of CSIR Labs Having Access to the following International Scientific Journal Publishers through the CSIR/NISCAIR E-Conglomerate**

- There are 42 CSIR labs which have access to 8 e-publishers from this Conglomerate;
- 40 CSIR Labs have E-Access to T&F and 35 Labs have E-Access to Indian Standards;
- 33 CSIR Labs have E-Access to ASTM Standards and 33 labs have E-Access to Blackwell;
- 22 CSIR labs have E-Access to Emerald and CUP; and
- Only 11 CSIR labs have E-Access to ASME.



**Fig. 3: NAL Scientists access to additional E-Journals through NAL-ICAST Gateway**

- NAL scientists have e-access to 1839 e-journals from Elsevier and 1600 e-journals from DOAJ and 1312 e-journals from Springer;
- A moderate number of e-journals for e-access belong to Blackwell, Taylor and Francis and Wiley;
- NAL scientists have open access to 700 e-journals through ICAST Gateway; and
- The minimum of e-journals for which e-access is available is for publishers AIAA and World Science. Low score on AIAA is surprising given the specialization of the scientists.

Table 1, 2, 3: Highlights the full-text usage statistics of E-Journals by NAL Scientists for the Years 2005, 2006, 2007.

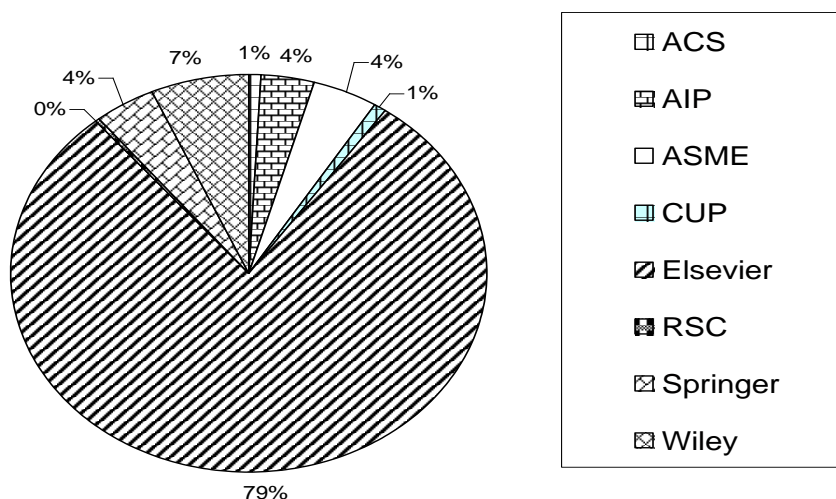
## Web Log Analysis of E-Journal Usage and Scholarly Communication

**Table-1: Year 2005**

Sl. No	Publ.	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total-Publ. Wise
1	ACS	0	3	47	4	18	22	21	51	23	10	2	3	204
2	AIP	0	0	0	0	0	47	153	172	417	46	19	26	880
3	ASME	115	54	115	56	83	98	54	53	30	43	11	325	1037
4	CUP	12	15	31	47	16	12	22	2	8	0	6	13	184
5	Elsevier	1384	1026	2221	1056	1903	2000	1026	1914	1503	1120	1814	2100	19067
6	RSC	3	2	8	0	12	9	8	9	7	7	0	5	70
7	Springer	19	172	183	128	63	70	69	31	60	61	51	36	943
8	Wiley	28	67	147	158	123	129	144	348	117	133	84	153	1631
<b>Total: 23067 (Month Wise All Publishers)</b>		<b>1561</b>	<b>1339</b>	<b>2752</b>	<b>1449</b>	<b>2218</b>	<b>2387</b>	<b>1497</b>	<b>2580</b>	<b>2165</b>	<b>1420</b>	<b>1987</b>	<b>2661</b>	<b>23067</b>

ACS=American Chemical Society, AIP=American Institute of Physics, ASME= American Society of Mechanical Engineers, CUP=Cambridge University Press, RSC=Royal Society of Chemistry

**Year 2005: NAL Full-Text Download Usage Statistics: All Publishers**



**Fig. 4: Year 2005: NAL Full-Text Download Statistics: All Publishers**

- 79% of full-text downloads for the Year 2005 is from journals published by Elsevier;
- Only 7% of full-text downloads for the same year is from journals published by Wiley;

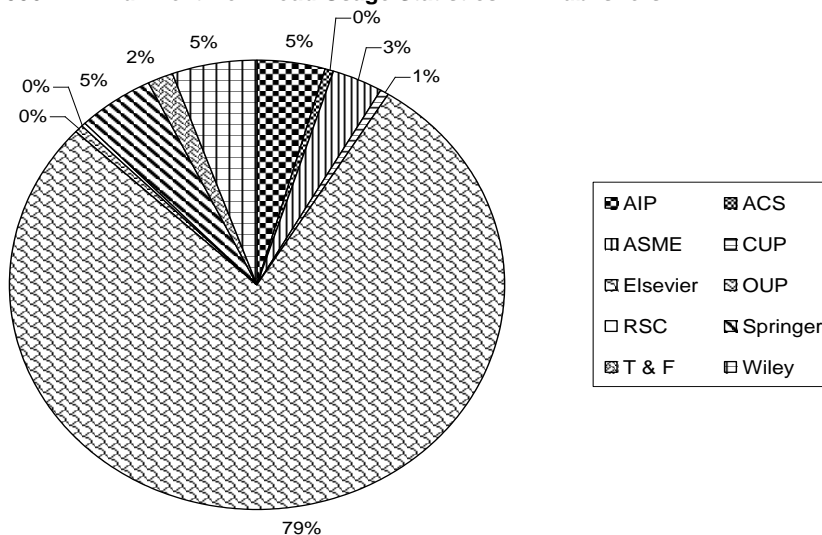
- 4% each of full-text downloads are from publishers Springer, AIP and ASME; and
- Only 1% each of full-text downloads are from publishers ACS and CUP.

**Table 2: Year 2006**

Sl. No	Publisher	Jan	Feb	Mar	Apr	May	Jun	July	Aug	Sep	Oct	Nov	Dec	Total-Publ. Wise
1	AIP	15	20	36	89	104	69	246	268	241	136	181	215	1620
2	ACS	1	8	7	3	5	12	2	10	8	19	13	40	128
3	ASME	0	39	67	40	63	50	103	66	177	83	397	112	1197
4	CUP	47	6	5	2	8	6	1	26	37	21	19	36	214
5	Elsevier	3202	3482	1888	2249	2837	1856	1698	2192	1512	2198	1939	2738	27791
6	OUP	0	0	0	0	5	6	12	6	38	53	29	0	149
7	RSC	1	1	3	10	7	4	3	5	6	15	16	21	92
8	Springer	137	100	90	98	98	81	63	175	314	93	313	283	1845
8	T & F	6	194	28	19	34	25	13	34	44	24	114	56	591
9	Wiley	216	91	124	120	210	201	119	161	172	131	237	174	1956
<b>Total: 34595 (Month Wise All Publishers)</b>		<b>3625</b>	<b>3941</b>	<b>2248</b>	<b>2630</b>	<b>3371</b>	<b>2310</b>	<b>2260</b>	<b>2943</b>	<b>2549</b>	<b>2773</b>	<b>3258</b>	<b>3675</b>	<b>34595</b>

ACS=American Chemical Society, AIP=American Institute of Physics, ASME= American Society of Mechanical Engineers, CUP=Cambridge University Press, RSC=Royal Society of Chemistry, T&F= Taylor and Francis, OUP=Oxford University Press

**Year 2006: NAL Full-Text Download Usage Statistics: All Publishers**



**Fig. 5: Year 2006: NAL Full-Text Download Usage Statistics: All Publishers**

## Web Log Analysis of E-Journal Usage and Scholarly Communication

- 79% of full-text downloads for the Year 2006 is from journals published by Elsevier;
- 5% each of full-text downloads are from publishers Springer, Wiley and AIP;
- 3% of full-text downloads are from publisher ASME;
- 2% of full-text downloads are from publisher Taylor and Francis; and
- The least percentage of full-text downloads are from the publisher CUP.

**Table 3: Year 2007**

Sl. No	Publisher	Jan	Feb	March	Apr	May	Jun	July	Aug	Sep	Oct	Nov	Dec	Total-Publ. Wise
1	AIP	174	52	184	374	254	248	350	234	332	334	448	136	3120
2	ACS	20	4	12	42	84	76	48	40	40	78	76	46	566
3	ASME	124	118	208	278	158	142	128	188	122	240	0	0	1706
4	CUP	52	0	0	0	84	72	62	0	70	50	22	36	448
5	T & F	-	-	-	-	-	-	-	-	-	-	-	-	--
6	Elsevier	4006	3453	5759	5105	4396	4302	4333	5653	4103	4645	3689	3335	52779
7	RSC	44	24	34	56	62	42	48	24	44	94	26	18	516
8	Springer	349	323	268	426	346	409	465	442	410	465	317	395	4615
9	Wiley	322	322	452	406	786	444	456	418	322	362	406	232	4928
10	ASCE	8	6	16	20	18	52	20	6	16	20	0	0	182
<b>Total: 68860 (Month Wise All Publishers)</b>		<b>5099</b>	<b>4302</b>	<b>6933</b>	<b>6707</b>	<b>6188</b>	<b>5787</b>	<b>5910</b>	<b>7005</b>	<b>5459</b>	<b>6288</b>	<b>4984</b>	<b>4198</b>	<b>68860</b>

ACS=American Chemical Society, AIP=American Institute of Physics, ASME= American Society of Mechanical Engineers, CUP=Cambridge University Press, RSC=Royal Society of Chemistry, T&F= Taylor and Francis, ASCE=American Society of Civil Engineers. Download Statistics for Taylor and Francis have not been tabulated for the Year 2007 in view of non-availability of data.

- 76% of full-text downloads for the Year 2007 are from the journals published by Elsevier;
- 7% of full-text downloads are from the publishers Springer and Wiley;
- 5% of full-text downloads are from the publisher AIP;
- 2% of full-text downloads are from the publisher ASME; and
- 1% of full-text downloads are from the publishers RSC, ACS, CUP.

Year 2007: NAL Usage Full - Text Download Statistics: All Publishers

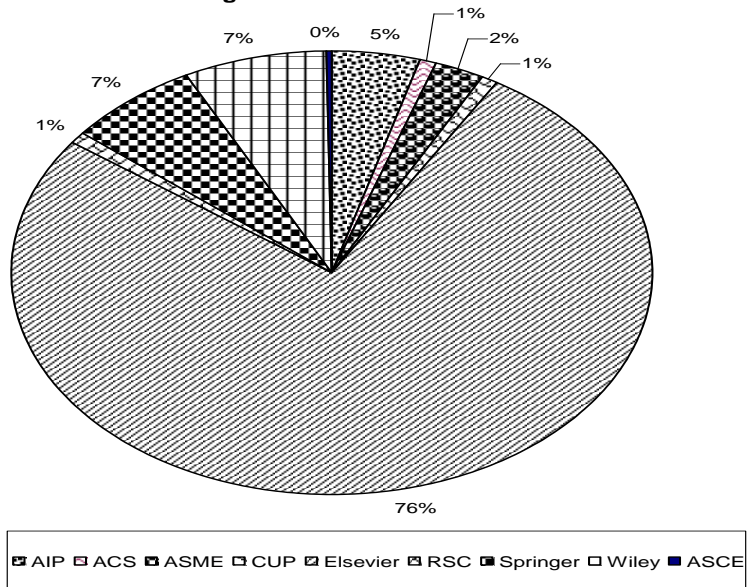


Fig. 6: Year 2007: NAL Usage Full-Text Download Statistics: All Publishers

Table 4: Highlights the consolidated monthly total downloads, Publisher Wise for the Years 2005, 2006, 2007

Sl. No.	Name of the Publisher	2005	2006	2007
1.	ACS	204	128	566
2.	AIP	880	1620	3120
3.	ASME	1037	1197	1706
4.	CUP	184	214	448
5.	Elsevier	19067	27791	52779
6.	RSC	70	92	516
7.	Springer	943	1845	4615
8.	Wiley	1631	1956	4928
9.	ASCE	-	-	182
10.	OUP	-	149	-
11.	Taylor and Francis	-	591	-

- Chi-Square test was applied to test whether there is independence between the years and the publishers;
- The calculated value of Chi-Square was found to be 845, which is highly significant;



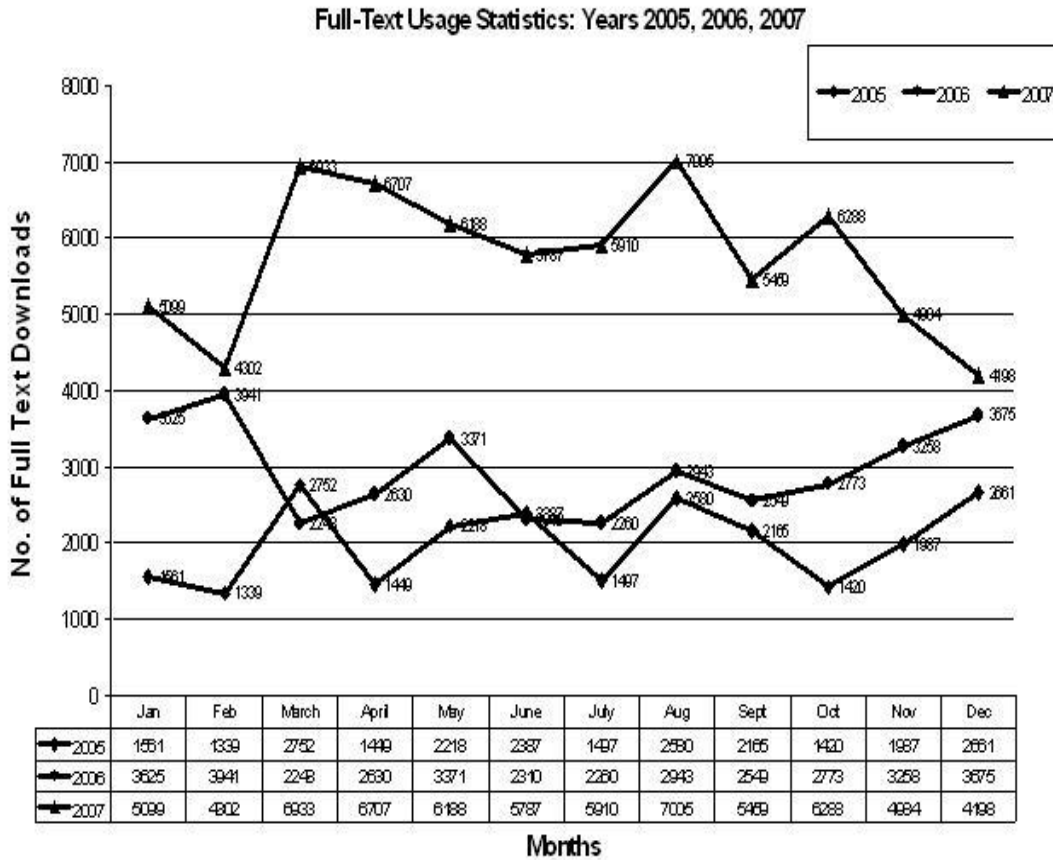
## Web Log Analysis of E-Journal Usage and Scholarly Communication

- Hence we conclude that for the full-text downloads data the years and the publishers are not independent; and
- This Chi-Square test was carried out for only those publishers (8 in number) for which the data was available for all the three years (2005-2007).

**Table 5: Highlights the total number of downloads (Month Wise, All Publishers) for the Years 2005, 2006, 2007**

Sl. No.	Name of the Month	2005	2006	2007
1.	January	1561	3625	5099
2.	February	1339	3941	4302
3.	March	2752	2248	6933
4.	April	1449	2630	6707
5.	May	2218	3371	6188
6.	June	2387	2310	5787
7.	July	1497	2260	5910
8.	August	1631	1956	7005
9.	September	2165	2549	5459
10.	October	1420	2773	6288
11.	November	1987	3258	4984
12.	December	2661	3675	4198
<b>Grand Total:</b>		<b>23067</b>	<b>34595</b>	<b>61405</b>

- From this table it is observed that the mean number (per-month) of full-text downloads for the above three years was found to be different through Kruskal Wallis test of 'One Way Analysis of Variance' at 1% level of significance



**Fig. 7: Line Graph: Full-Text Usage Statistics: Years, 2005, 2006, 2007**

- In 2005, full-text usage varied between 1561 in the month of January to 2661 in the month of December with a peak of 2752 in the month of March 2005;
- In 2006, the number of full-text usage varied little with 3625 in the month of January to 3675 in the month of December with a peak of 3941 in the month of February, 2006; and
- In 2007, the full-text download increased with 5099 in the month of January to a maximum of 6933 in the month of March and 7005 in the month of August and declined to a value of 4198 in the month of December 2007.

## 10 CONCLUSION

The coming of age of e-journals has altered the way scholarly information is disseminated throughout the world [44]. They have not only affected the way information is spread, but the way information is acquired and how scientific researchers seek that needed information.

We discuss a ‘new methodology’ [9] that has emerged for studying journal usage and scholarly information seeking behaviour, popularly called the “transaction log analysis”. Other methodologies including ‘Deep Log Analysis Method’ [30, 31, 32] are also discussed.

We presented analytical data on use patterns of full-text e-journal for the period 2005 to 2007 by scientists of NAL. The major findings we wish to highlight are:

- (a) The mean number (per-month) of full-text downloads for the above three years was found to be different through Kruskal Wallis test of ‘One Way Analysis of Variance’ at 1% level of significance; and
- (b) Chi-Square test was applied to test whether there is independence between the years and the publishers. The calculated value of Chi-Square was found to be 845, which is highly significant. Hence we conclude that for the full-text downloads data, the years and the publishers are not independent. This Chi-Square test was carried out for only those publishers (8 in number) for which the data was available for all the three years (2005-2007).

A review of literature in this area indicated that there is hardly any Indian study on ascertaining the e-journal full-text download and usage patterns amongst the aerospace scientists and engineers in India. Hence, this study appears to be first of its kind in India. In the context of exploding aviation activity in the Asian region in general and India in particular, aerospace is seen as a key technology. The value of this study is further enhanced in applying state-of-the-art online e-journal usage tools like the Web Log Analysis, which ensures that the data remains ‘unfiltered’, ‘pure’, and most importantly without ‘human intervention’. Applying such tools further ensures authenticity of the data analyzed.

## 11 ACKNOWLEDGEMENTS

The authors sincerely thank Dr. A. R. Upadhya, Director, NAL for his support, and for permitting to publish this article. Thanks to Dr. Ranjan Moodithaya, Head, KTMD and Dr. M. N. Satyanarayana, Jt. Head, KTMD for their encouragement. We also thank Mr. Prakash Chand, Scientist-in-charge NISCAIR/CSIR e-journal conglomerate and his colleagues for providing us relevant full-text e-journal usage statistics of NAL scientists.

## REFERENCES

1. Jansen, Bernard J; Amanda Spink; Tefko Saracevic (2000). "Real life, Real Users, Real Needs: A Study and Analysis of User Queries on the Web". *Information Processing and Management*, 36(2); 207-227; Pergamon Press.
2. Jansen, Bernard J; Amanda Spink; Tefko Saracevic (1999). The Use of Relevance Feedback on the Web: Implications of Web IR System Design, World Conference on the WWW and Internet, Honolulu, Hawaii.
3. Amanda Spink; Tefko Saracevic (1997). "Interaction in Information Retrieval: Selection and Effectiveness of Search Terms". *JASIS*. 48(8); p.741-767.
4. Kantor, Paul B; Tefko Saracevic (2008). "Quantitative Study of the Value of Research Libraries: A Foundation for the Evaluation of Digital Libraries." *ASIS'99*.
5. Diann Rusch-Feja; Uta Siebeky (1999). "Evolution and Usage and Acceptance of Electronic Journals". *D-Lib Magazine*. 5(10).
6. David Nicholas; Paul Huttington; Tom Dobrowolski; Ian Rowlands, Hamid, R. Jamali; Panayiota Polydoratou (2005). "Revisiting 'obsolescence' and Journal Article 'Decay' through Usage Data: an analysis of digital journal use by year of publication, Elsevier, 41(2005), pp. 1441-1461.
7. Charlie Morris (1999). "There's gold in them there log files", Web Developer's Virtual Library (WDVL).
8. Hamid, R Jamali, David Nicholas; Paul Hutington (2005), "The use and users of scholarly e-journals: a review of log analysis studies", *Aslib Proceedings, New Information Perspectives*. 57(6); p. 554-571, [www.emeraldinsight.com/0001-253X.htm](http://www.emeraldinsight.com/0001-253X.htm).
9. David Nicholas; Paul Huntingon; Hamid, R. Jamali; Carol Tenopir (2006). "What deep log tells us about the impact of big deals: case study OhioLINK". *Journal of Documentation (Emerald)*. 62(4); p.482-508.
10. Lynn Silipigni Connaway; Clifton Snyder (2005). "Transaction Log Analysis of E-Book Usage". *Against the Grain*. p. 85-89.
11. Ahmed Taha (2004). "Wired Research: Transaction Log analysis of E-Journal Databases to Assess the Research Activities and Trends in UAE University". Corporate Information Issues, Nord I&D, Knowledge and Change.
12. Vince Mariner (2002). "Logging Usability", Reed Business Information, a division of Reed Elsevier Inc.

## Web Log Analysis of E-Journal Usage and Scholarly Communication

13. Anamarija Rozic-Hristovski; Dimitar Hristovski; Ljupco Todorovski (2002). "User's Information Seeking Behaviour on a Medical Library Website. *J Med Libr Assoc.* 90(2).
14. Davis Philip, M. (2004). "Information seeking behaviour of Chemists: A Transaction Log Analysis of Referral URLs. *JASIST.* 55(4); p. 326–332.
15. David Nicholas; Paul Huttington; Hamid, R Jamali; Anthony Whatkinson (2006). "The information seeking behaviour of the users of digital scholarly journals". *Information Processing and Management.* 42(2006); p. 1345-1365.
16. Jillian, R Griffiths; Hartley, J R; Jonathan, P. Willson (2002). "An improved method of studying user-system interaction by combining transaction log analysis and protocol analysis". *Information Research.* 7(4).
17. Diane Harley; Jonathan Henke (2007). "Towards an effective understanding of Web Site Users: Advantages and Pitfalls of linking Transaction Log Analyses and Online Surveys". *D-Lib Magazine.* 13(3/4).
18. Harley, Diane (2007). "Why Study Users? An Environmental Scan of Use and Users of Digital Resources in Humanities and Social Sciences Undergraduate Education". First Monday, volume 12, number 1 (January 2007). Available at [http://firstmonday.org/issues/issue12\\_1/harley/index.html](http://firstmonday.org/issues/issue12_1/harley/index.html).
19. Harley, Diane; Jonathan Henke; Shannon Lawrence; Ian Miller; Irene Perciali; David Nasatir. (2006). "Use and Users of Digital Resources: A Focus on Undergraduate Education in the Humanities and Social Sciences. University of California, Berkeley". Available at: <http://cshe.berkeley.edu/research/digitalresourcestudy/report/>.
20. Paolo Gardis. "Measuring the performance of a biomedical digital library: web-site, e-journals and databases" (2001). University of Torino, EAHIL Workshop, 2001 Cyberspace Odyssey, Alghero, Sardinia.
21. Hao-Ren Ke; Rolf Kwakkelaar; Yu-Min Tai; Li-Chun; Chen (2002). "Exploring behaviour of E-journal users in science and technology: Transaction Log Analysis of Elsevier's Science Direct OnSite in Taiwan". *Library and Information Research.* Pergamon Press. 24(2002); p. 265 – 291.
22. Philip, M. Davis; Jason, S. Price (2006). "eJournal interface can influence usage statistics: implications for libraries, publishers, and Project COUNTER". *Journal of American Society for Information Science and Technology (JASIST).* 57(9); p. 1243-1248.
23. Paul Huttington; David Nicholas; Hamid, R Jamali (2007). "Employing Log Metrics to Evaluate Search Behaviour and Success: Case Study BBC Search Engine. *Journal of Information Science.* 33(5); p. 584-597.
24. Moen, McClure (1997). "Web Server Transaction Log Analysis Methodology". *An Evaluation of U.S. GILS Implementation.*
25. Nikki Warren (2002). "Website Log Analysis: Approaches for the Library of the National Institute of Environmental Health Sciences".
26. Tenopir, C; King, D W (2000). "Towards Electronic Journals: Realities for Scientists, Librarians, and Publishers, *Psychology: 11 (084) electronic journals (1)* [Special Libraries Association] 2000; p. xxii+488.
27. Garvey; William, D (1979). "Communication: The Essence of Science." Oxford Pergamon Press.
28. Rob Kling; Ewa Callahan (2005). "Electronic Journals, the Internet, and Scholarly Communication". *Indiana University, Bloomington, ARIST.* 37(1); p.127-177.

29. Okerson, A. (2000). Are we there yet? Online e-resources ten years after. *Library, Dends.* 48; p. 671-694.
30. Nicholas, D.; Huttington, P.; Watkinson, A. (2003), "Digital journals, Big Deals and Online searching behaviour: a pilot study". *ASLIB Proceedings.* 55(1/2); p. 84-109.
31. Nicholas, D.; Huntington, P.; Watkinson, A (2005). "Scholarly journal usage: the results of deep log analysis". *Journal of Documentation.* 61(2); p. 246-80.
32. Nicholas, D.; Huttington, P.; Watkinson, A.; Jamali, H. R. (2005). "The use of digital scholarly journals and their information seeking behaviour: what deep log analysis and usage data can disclose". *Journal of the American Society for Information Science and Technology.* 56(12).
33. Morse, D. H.; Clintworth, W. A. (2000). "Comparing patterns of print and electronic journal use in an academic health science library". *Issues in Science and Technology Librariananship.* 28. Available at: [www.istl.org/00-fall/refereed.html](http://www.istl.org/00-fall/refereed.html).
34. Davis, P.; Solla, L. (2003). "An IP-level analysis of usage statistics for electronic journals in chemistry: making inferences about user behaviour". *Journal of the American Society for Information Science and Technology.* 54(11); p. 1062-8.
35. Davis, P.M. (2002). "Patterns in electronic journal usage: challenging the composition of geographic consortia". *College and Research Libraries.* 63(6); p. 484-97.
36. Ke, H.; Kwakkelaar, R.; Tai, Y.; Chen, L. (2002). "Exploring behaviour of e-journal users in science and technology: transaction log analysis of Elsevier's ScienceDirect OnSite in Taiwan". *Library and Information Science Research.* 24(3); p. 265-91.
37. Tenopir, C. (2003). "Use and users of electronic library resources: an overview and analysis of recent research studies". *Report for the Council on Library and Information Resources,* August 2003, available at: [www.clir.org/pubs/reports/pub120/pub120.pdf](http://www.clir.org/pubs/reports/pub120/pub120.pdf) (accessed 20 March 2005).
38. Tenner, E.; Zheng Ye, Y. (1999). "End-user acceptance of electronic journals: a case study from a major academic research library". *Technical Services Quarterly.* 17(2); p. 1-14.
39. Worlock, K. (2002). "Electronic journals: user realities – the truth about content usage among the STM community". *Learned Publishing.* 15(3); p. 223-6.
40. Davis, P.M. (2004). "For electronic journals, total download can predict number of users". *Portal: Libraries and the Academy.* 4(3); p. 379-92.
41. National Aerospace Laboratories, [www.nal.res.in](http://www.nal.res.in)
42. [www.csir.res.in](http://www.csir.res.in)
43. [www.niscair.res.in](http://www.niscair.res.in)
44. Amy, C Gleeson (2001). "Information seeking behaviour of scientists and their adaptation to electronic journals". *Masters paper for the M.S. in Library Science degree, School of Information and Library Science, University of North Carolina, Chapel Hill.*