Usage data analysis for supporting decisions in LIS management

Giovanna MIRANDA*, Francesca GUALTIERI**
A GIDIF, Rbm group study

*Scientific Information & Library Services, sanofi aventis spa, Research Centre, Milan, Italy
**Scientific Information & Library Services, Rottapharm|Madaus, Monza, Italy

Introduction

Library and Information Services (LIS) budget is almost totally absorbed by the cost of journals and books, in print or electronic version, and databases access, but users have the unreasonable expectation that almost every thing is online free of charge. On the contrary, often, to meet the budget, new journals, books or databases purchase requires the cancellation of other currently received titles or decrease/cancel accesses to databases.

The central library model of content acquisition, the easiness of access for users - more and more scattered worldwide - push librarians to purchase journals and books in electronic versions only, but without having significant savings. Moreover, spreading access to chemical, biomedical and competitive intelligence databases to users working on different sites imply huge costs, and ask selecting carefully who access to what. LIS have to choose between providing access to a broad array of online journals, books and databases, subject to the license terms of the publishers, or keeping a small but highly focused print/electronic journals and books collection and selected databases.

To face this economic trend and make a better decision, LIS need to know very well their users: i.e. needs, habits, preferred titles and databases and ways to interact with information.
Access to journals or books (online or traditional paper) and databases is a key factor that shapes frequency of usage, and its evaluation is an essential part of the library planning. All the methods described in the literature to anlayse usage have advantages and disadvantages. LIS have to choose the appropriate method that better fits their information sources and typology of user behaviors and could help them to make the best decisions.

To identify if usage analysis is widely used in Italy, beginning 2010, we have invited the members of GIDIF, RBM (the Italian librarians and documentalists of pharmaceutical industries and biochemical research institutes association), to participate in a survey asking what methods they use to obtain statistic and to analyse the use of information resources. The survey is focused on qualitative and quantitative methods, and tools described in literature to obtain information on source use, namely journals and books (in paper or electronic format) and databases.

Methods/Process

The survey was carried out among 70 of the LIS members of Gidif, Rbm belonging to 30 pharmaceutical industries, 23 Universities and Biomedical Institutes and 17

hospitals. The survey refers to the most well-known qualitative/quantitative methods available.

Questions asked members which methods they used to analyse the usage of print and electronic sources (reshelving, survey, publisher reports, etc); which parameters in their opinion was the most representative, i.e. number of access, downloaded TOCs, abstracts and how they analysed the usage, i.e. the most used journals, cost-benefit, return of investment analysis. The last question was related to whether the library gave their users access to non-subscribed jounals through pay per view agreements.

Following a brief description of the methods cited in the literature, or commercialized tools, to obtain information on the use and cost-benefit of information sources.

Reshelving Statistics on print use. It is a locally recorded count of journals and books placed back in the shelves in the libraries, circulated to patrons, or used for the library photocopy service program. Each data registered is equals one use of a journal or books. Print use, simply indicating that a user was interested in accessing a particular journal for some reason. The general opinion on methods measuring print use is that they are time consuming and inaccurate.

Descriptive Statistics. It derives from a survey to analyse preferences of users (ex. traditional print versus online format, preferred titles, etc), or interview to the top management and/or researcher's opinions. This method is too subjective and sometime the interest showed for a journal does not correspond to the actual usage.

Local Citation Analysis. The journals are ranked by citation counts and those with high citation counts are considered to be of high use. It should be noted that the online availability of a journal was found to significantly increase local citations.

Impact Factor. Besides to be considered a valuable tool to evaluate journal quality, to choose relevant journals in a subject area and to evaluate research productivity, Impact Factor is considered also a tool to evaluate journal use. Nevertheless in evaluating Impact Factor as journal use, LIS have to bear in mind that journals published only in print formats had no significant changes in impact factor through the years, and that those published in print and online formats had a significant increase in impact factor. Moreover the free availability of Open Access Journals significantly affects journal use and citations.

Access Use. It is a measure of local online journal use derived from transaction log counts of the main routes that users accessed online journals (ex. IP address). This parameter is useful when you have to track the usage of a consortium or a network of libraries, to see the different use of the libraries.

Publisher reports. All institutional subscription administrators have access to their usage statistics: how often users use e-journals to retrieve, read, and/or download. Example: numbers of successful requests by month, journal and page type.

Depending on the publisher, the reports includes a variety of counts, such as: use by IP address, table of contents, downloaded abstracts or full text articles (use by file type PDF or html), and page downloads. Sometimes in giving the number of rejected sessions defined as unsuccessful log (turnaway). In case of the e-books the reports gives the number of downloaded chapters, whereas for databases reports gives

users sessions or searches run, viewed pages, alerts and other data (abstract, full text, etc).

Uniformity of data is essential to compare statistics and obtain reliable information.

The COUNTER (Counting Online Usage of Networked Electronic Resources) Code of Practice for Journals and Databases specifies: the data elements to be measured; definitions of these data elements; usage report content, format, frequency and methods of delivery; protocols for combining usage reports from direct use and from use via intermediaries. Not all vendors are COUNTER Compliant. To be considered COUNTER Compliant, the vendors must be compliant with the release 3 of the code.

The advent of the SUSHI protocol (Standardized Usage Statistics Harvesting Initiative) - developed by NISO (National Information Standards Organization) in cooperation with COUNTER - has greatly facilitated the handling of large volumes of usage data. Its implementation by vendors will allow the automated retrieval of the COUNTER usage reports into local systems, making this process much less time consuming for the librarian or library consortium administrator.

Reading Factor. Defined as the ratio between the number of electronic consultations of an individual journal and the mean number of electronic consultations of all journals considered. Value of 1 represents an average consultation rate, while a value >1 represents a higher than average consultation rate.

There is no way to determine whether this factor corresponds to a partial or a complete reading, an effect of curiosity or even an accidental mouse click. A revised version of Reading Factors counts "clicks" only if a visitor spends an appropriate amount of time looking the documents (e.g. 30 seconds). Reading Factor measures the interest in a journal within the limits of a given readership and cannot be taken as an indicator of the journal quality. It could be helpful in revealing differences on the use of journals of equivalent Impact Factor.

The cost-benefit analysis. There are studies suggesting that journals need to be evaluated by usage related to cost, and core journals are required to maximize local usage and minimize cost. The cost-benefit analysis assess journal values, considering several benefit factors including coverage, impact factor, use, location, and inter-library loan requests, as well as subscription cost. Some studies questioned whether the use of cost-benefit analysis was appropriate in evaluating journals, since 'use' actually does not equal 'benefits'.

Return of Investment (ROI) - The definition for an ROI varies depending on the context. Quantitative measurements of ROI include time saved by library users; the money users save by using the library instead of alternative sources; and revenue generated with the assistance of the library. Qualitative measures include the reliance of users on library-supplied content and services for decision-making; the level of decisions that the library supports; and the relative value of having the support of a professional information function that knows. Another key qualitative value metric revolves around the importance of information provided by the library that the user would not have found or had access to without the library's intermediation.

ERM (Electronic Resource Management). It is a service, powered by SwetsWise, designed to aid librarians with the management, evaluation and access of electronic resources. It gives an overview on the library collection, showing the content, end users access, licensing terms and conditions options and the possibility to analyse

the cost per use of subscriptions, pay-per-view and document delivery analysis. ERM gives COUNTER compliant usage statistics.

EBSCONET Usage Consolidation. It is an EBSCONET tool that allows combining COUNTER compliant usage from content providers with other statistics, creating an overall picture of how the library collection is used. It is possible to leverage cost and other data to further analyse collection.

JUR (Journal Use Reports). It is a Thomson Reuters product. JUR gets a picture of journal performance capturing different types of use data into a single interface and providing a systematic quantitative approach to gather and review statistics. It analyses usage by citations, usage or both. It defends ROI by integrating user activity with researcher output to see how their library collection is contributing to academic output. JUR is COUNTER-compliant.

JURO (Journal Usage Report Online) is an open source software that enables libraries to capture journal usage from different sources: COUNTER-compliant usage reports of vendors and in-house journal statistics, or imported from any other system. It generates different usage reports based on the user preferences.

360 Counter. It is a Serials Solutions product. 360 Counter analyses usage data received from COUNTER-compliant reports. It gives automated cost-per-use calculation and support SUSHI.

Scholarly IQ provides access to compliant COUNTER 3 reports through the SUSHI protocol. Supported reports include COUNTER Journal, Database and Book Reports.

Results

We have received 37 completed questionnaires (53% of answer). The percentages of answer per typology of members were 50% Pharmaceutical Industries, 35% Universities and Biomedical Institutes and 82% Hospitals.
89% of respondents verify statistics of use whereas 11% does not do any control. The usage statistic are surveyed especially for electronic resources, 84% for e-journals vs. 41% print journals, 22% e-books vs. 11% print books, and 57% databases.
The use of print sources is verified through qualitative methods: reshelving, registration of loans and circulation lists (69%) and/or survey, interviews, opinions (42%).
The use of electronic resources are verified most of all through quantitative methods: log counts, IP address or publisher's reports (76%). A minority verifies use through qualitative methods: interviews, surveys, opinion (7) or Impact Factor (12%). No one use citation analysis or commercial tools (SwetsWise, Ebsconet, JUR, etc).
In LIS opinion the most representative parameter to monitor the use of e-journals and e-books are downloads of full texts (52%) and the number of on-line access (28%); to monitor databases the number of searches run (95%).
Decisions to renew or cancel subscription to e-journals, e-books and databases are made on usage statistics (73%), cost-benefit analysis (48%) or other parameters such as users' opinion or internal policy (21%). Only 6% use ROI (Return of Investment) analysis.

To give access to e-journals not subscribed, the libraries stipulate pay per view contracts with the editors (22%) and/or document delivery services (92%).

Discussion and Conclusions

Information can be considered only a cost until it is not used. Usage analysis is an emerging area of bibliometric research that helps focus the need to standardize the collections and select the information sources. It could help not only in making decision and choosing a more tailored journal collection and database selection, but also to identifying problems and weakness, to justify expenses and budget allocations, to monitor the library's expansion, to prioritize research areas and to predict prospective. Last but not least, gaining support of top management in deciding budget allocations.

The decisions should be based considering many issues such as subscription costs, impact factor, user request, journal coverage, language of publication, availability elsewhere, local usage and local citation. Staff-time devoted to services and library salary budget should be considered to analyse cost-benefit value and the return of investment (ROI) of the sources and library services. LIS have to choose and adapt the best method to apply to their libraries.

The majority of members of Gidif, Rbm in Italy monitor the use of information resources, both in print or electronic format. The methods used for print resources are qualitative, e.g. descriptive methods, surveys or interviews of users, and quantitative for electronic resources, mainly the publisher's report. The tools developed and commercialized by some agents or information producers are not yet in use. Decisions to renew or cancel subscription to e-journals, e-books and databases are made mainly on usage statistics, cost-benefit analyses and ROI are not yet used to a great extent.

Suggested Reading

H. Kyung Chung.
An analysis model of creating a core journal collection for academic libraries.
Library Collections, Acquisitions & Technical Services, 2009, 33, 17-24

J.D. McDonald
Understanding online journal usage: A statistical analysis of citation and use
Journal of the American Society for Information Science and Technology, 2006, 57, no.13

The Counter Code of practice
http://www.projectcounter.org/r3/Release3D9.pdf

H.R. Enssle, M.L. Wilde
So you have to cancel journals? statistics that help
Library Collections, Acquisitions, and Technical Services, 2002, 26, 259-281

J. Luther
University investment in the library: What's the return?
A case study at the University of Illinois at Urbana–Champaign

http://libraryconnect.elsevier.com/whitepapers/0108/lcwp010801.html