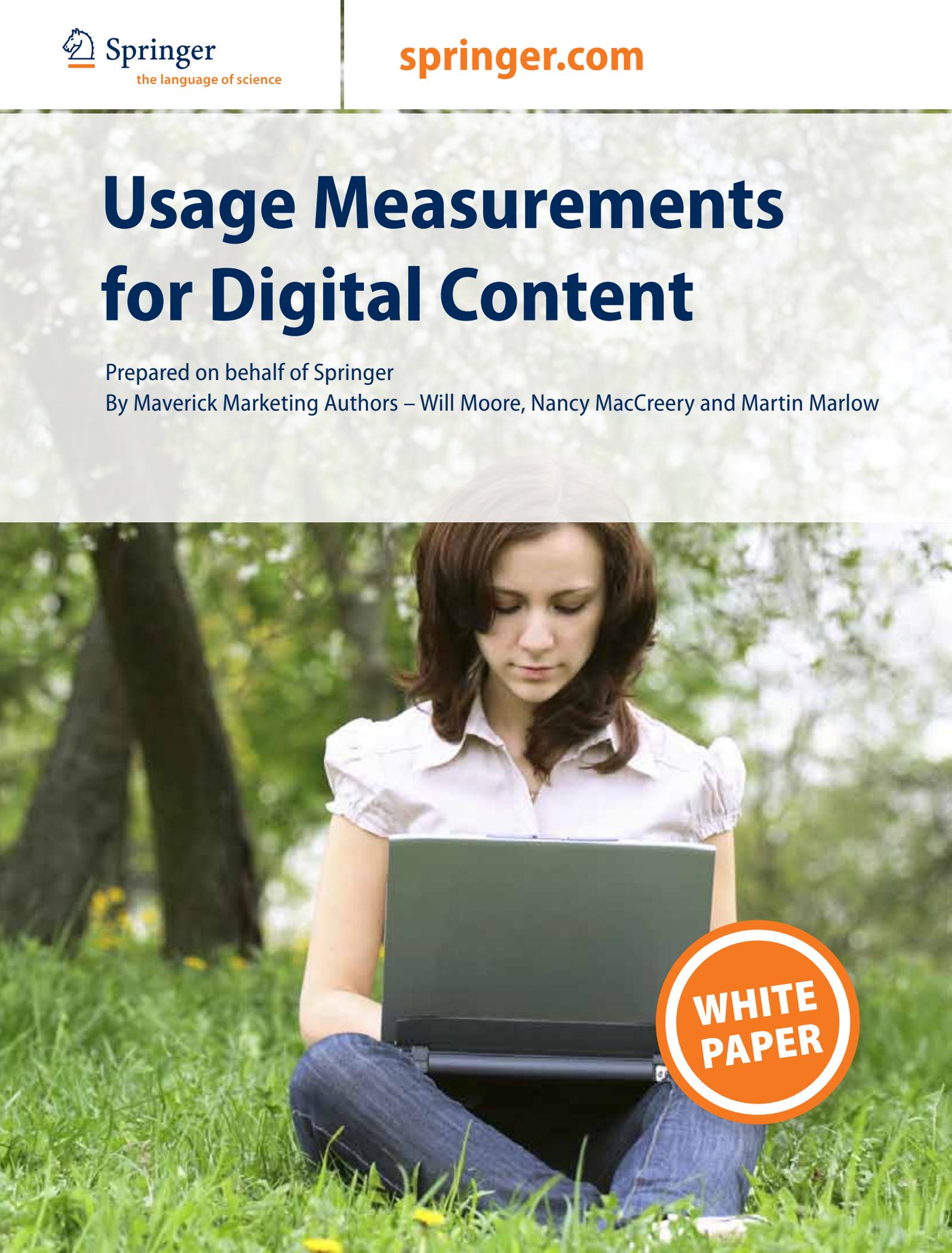


Usage Measurements for Digital Content

Prepared on behalf of Springer

By Maverick Marketing Authors – Will Moore, Nancy MacCreery and Martin Marlow



**WHITE
PAPER**

Executive Overview

Libraries face a dynamic and progressively complex environment. Due to the explosive growth of content, increased budget pressures, and changes in the way that content is delivered and used, there is a growing need to find and evaluate methods to judge the usefulness of library resources. This is a difficult challenge, but necessary to maintain the trustworthy resources required by patrons and the viability of libraries into the future.

- ▶ The growing amount of content is reflected in the growth of the number and availability of scholarly journals. Simultaneously, the number and availability of eBook resources has exploded.
- ▶ User expectations have changed. Patrons expect instant search, global reach and will utilize more sources in their research.
- ▶ With digital content, more usage statistics are available to support decision making, but there are multiple factors to consider. Subject matter differences and nuanced usage models should be considered as well as more traditional measurements, such as how often the content is accessed. Content measures include straight usage statistics, the Impact Factor, the Eigenfactor and newer metrics that begin to take into account the linkages and interdependencies that characterize modern, interdisciplinary research.
- ▶ As libraries strive to become more user-centric, usability and findability become increasingly important. Measures to ensure portals are very usable, open to referrals from search engines, and contain previously hard to find 'long tail' content should also be evaluated – to maximize their value, content resources should help to facilitate and interact with new research methods and user demands.

Introduction

Libraries are facing explosive growth in the amount of available content, significant changes in delivery formats and mechanisms, as well as changing expectations among library patrons about the availability of content and the manner in which they interact with it. In addition, libraries are under increasing pressure to defend their budgets as Universities are forced to cut more costs in a slowing economy. Amidst all of this, libraries continue to provide a valuable core service, perhaps more critical than ever in an era where the availability of information is growing exponentially; that is to sift through vast collections of information and make choices about the value and relevance of content while providing trustworthy and convenient access to their patrons.

Libraries must allocate limited resources to carefully select content and validate that their choices are meeting user needs. This situation presents both opportunities and challenges as libraries work to stay in step with the evolving needs of their user base.

Questions are plentiful in this changing landscape. Are your current resources meeting the needs of your user base? How is the user base changing? Which resources are actually being used, and how? What impact do new electronic delivery platforms have on the way that patrons use content? And the big question – how relevant is the content, the delivery technology, and are all of these variables working together to meet user needs for critical research information now and for the future?

There are many ways to interpret the increasing amount of data generated from the use of digital content. Basic usage data can appear to be a straightforward and expedient measure of value and cost effectiveness – i.e. the more a piece of content is used, the more apparent value it has – but there are other important variables and measures to consider when making renewal or purchase decisions. Not considering them can lead to critical content being cut/not renewed despite its relevance to library patrons which can be demonstrated when more qualitative considerations are taken into account.

In an electronic delivery platform, the relevance of a particular text can no longer be judged in isolation; it must be considered in context, measured not just on how often it is accessed or how much time a patron spends reading it, but also on the behaviors that it drives. Digital content usage measurement is complex, but needs to be understood if the correct buying decisions are to be made. This paper will attempt to shed some light on these areas and provide some useful information on how librarians might determine the answers for their institutions and patron base.

Trends affecting content and the user base

Content explosion and migration; how technology is changing the landscape of information

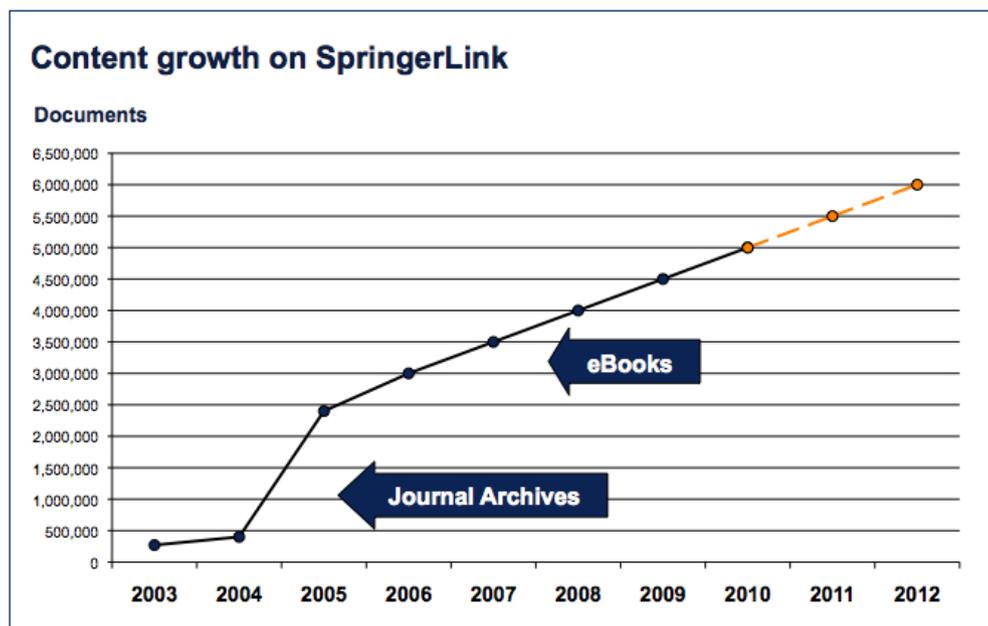
Each year, more and more content becomes available in various digital formats. This vast array of content, offered up through a range of both reliable and less reliable sources, is frequently available as text, audio, video, images or a combination of all of above. Much of this is produced through traditional publishing channels though an increasing amount is user generated. Scholarly publishing channels have also experienced unprecedented growth in the amount of content available, especially in the area of eBooks; where in certain disciplines specialized subject specific eBooks are flourishing.

Adding to this complexity, content is no longer static, but is increasingly dynamic, with users viewing it as a step on their path to discovery, rather than an end result. For example, an article or eBook might not have high usage measurements but may have value as a critical step in the path to other research.

Journals – Growing in number and availability

There were about 25,400 active scholarly peer-reviewed journals in early 2009, collectively publishing about 1.5 million articles a year. The number of articles published has grown at 3% per year while the number of journals has grown by 3.5% per year.¹ The largest growth among journals is in scholarly journals for research use. The average number of journals available in library collections more than doubled between 2000 and 2009.² This doubling represents a significant shift in library purchasing patterns and allocation of resources.

One factor in the rapidly expanding availability of journals is quick adoption of electronic delivery among journal publishers. According to the ALPSP's 2008 report on scholarly publishing practice, 96% of STM and 87% of arts, humanities and social sciences journals were accessible electronically in 2008 which represents a steady increase compared to comparable surveys conducted in 2003. During that same period, print books have also made the transition to eBooks, albeit at a slower growth rate as publishers have struggled to adapt their book publishing model to suit electronic delivery and availability.



¹ "STM Report: an Overview of Scientific and Scholarly Journal Publishing", Mark Ware and Michael Mabe, September 2009. p. 18

² "ALPSP Survey of Librarians", Ian Russell, October, 2009. p. 5

The latest CIBER research suggests that eBooks will be the next publishing success story, although demand here could be even more spectacular, simply as a result of the enormous size of the student population, hungry for highly digested content.

In the print world, only rudimentary usage statistics typically exist. Properly-designed eBook collections allow librarians to monitor the usage of these books in much greater detail, and at the same time support the decision-making about collection development and budgets.

EBooks – While journal information is growing, eBooks are now exploding

Although slower to adapt than journals, eBooks have begun to see explosive growth. With consensus emerging around a handful of digital formats and delivery platforms, publishers and authors are now leveraging digital publishing for mainstream as well as more specialized content. Authors and publishers, freed from print run volume requirements as well as the constraints and delays associated with print production are using digital publishing platforms to bring to market content that would not until now have been economically feasible using traditional printing and distribution systems. More specialized books that would not have met volume requirements to be published in print versions can now be published and distributed digitally in very short time frames – giving researchers new access to a variety of current, focused content not previously available, or affordable as part of a library collection.

The number of eBooks published has grown steadily since 2005, but the rate of growth has really accelerated over the past two years, with 2009 wholesale eBook sales almost 200% larger than vs. 2008.³ For a variety of reasons, detailed below, eBook collections are very well-suited as library resources. In recent years, publishers and commercial content delivery platforms have highlighted the collection model as a primary offering to academic libraries. The increased interest and growth in eBook collections has been driven by the several main factors: user convenience (24/7 on and off-campus access), part of a larger strategic move to electronic access, student demand for multiple copies (particularly for use around key deadlines such as exams); searchability, a simplified purchasing model and reduced pressure on shelf space.⁴

Where eBooks are readily available, they are increasingly being accepted and utilized by library patrons. Springer's own research data confirms this trend. While overall digital downloads were up 33% in 2008, eBook chapter downloads were up 70%. In fact, eBooks usage more than doubled between 2007 and 2009. Within eBooks, the top 20% titles account for 'just' 53% of usage. This spread is far less concentrated than it is for journals, where a 20/80 rule generally still applies. The usage spread is an important statistic to note – it indicates a larger trend in the usage model for eBooks where patrons are increasingly finding that the 'long tail' effect applies to academic research as it does to commercial content and will often yield more fruitful and specific results. These specialized eBooks may not appear to have high usage, but when accessed, they have significant value to the user. As suggested by the MESUR project research, for usage measurements to truly demonstrate the value of a piece of content, they must capture this qualitative information alongside the more traditional quantitative measurements.

Improved electronic search combined with growing eBook collections makes information readily accessible that was once difficult to reach. Books that sold only a few titles in the past are seeing increased access in academic research due to their increased availability and improved methods for locating the relevant information within them. That said; content still counts in eBooks as it does in journals. Among the top used titles are many of the 'usual suspects', including Springer Handbooks Series, Prokaryotes and bestselling textbooks. Textbooks often contain links to multimedia or other resources, alternative learning content that enhances learning but will not be captured in the usage statistics for the eBook.

Users and changing usage patterns

As technology has enabled more and more content to be accessed electronically, research behaviors have also evolved to reflect these changes. Top ranking activities carried out on library resources still include browsing and searching but now also include downloading. Access and navigation to articles is increasingly driven by search, rather than browsing.

Library users, and especially students, are accustomed to the experience provided by commercial search engines like Google and expect immediate results – this means instant search results and access to information with a global reach. It also means that they are used to sifting through results of questionable relevance, scanning brief snippets of information to quickly determine the value and sourcing quality of a particular bit of information, so a library resource platform needs to be deep linked and crawlable, available 24/7, with all resources assessable to its user base online.

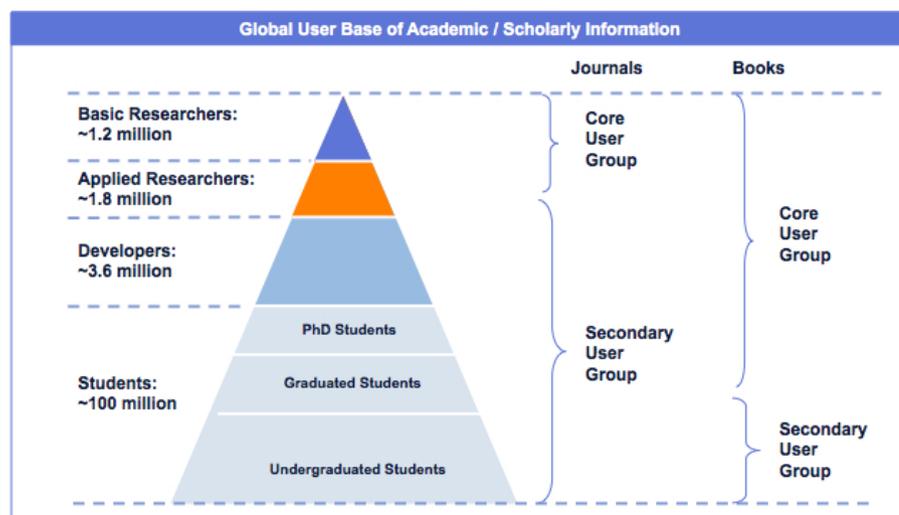
³ www.IDPF.org Annual Industry Statistics

⁴ STM report p. 17

Similarly, more efficient access to resources in a library setting leads to researchers willing to leverage more and more types of sources. As is the habit in the use of commercial search engines, scanning, rather than reading an entire piece of content, is increasingly prevalent as the usage model. In addition, the availability of active links to more and more electronic resources enables researchers to more quickly explore more materials on a topic and follow those linkages to increasingly focused or even supplementary materials. As mentioned above this 'long tail' effect is a significant and important change in the way that research is increasingly performed and will have a correspondingly significant impact on usage measurements. As more material is more easily found, the result is a change in reading patterns, with researchers reading more, averaging 270 articles per year, but spending less time per article, with reading times down from 45-50 minutes in the mid-1990s to just over 30 minutes per article.⁵

With an increasingly global outlook, patrons also expect libraries to provide access to content with global reach. There are also changes in the geographic distribution of content origination and consumption. Output from East Asia (China, Singapore, South Korea and Taiwan) has grown dramatically. Between 1995 and 2005, China's output grew at 17% and Taiwan's at 16% per year, compared to 0.6% for the USA and 1.8% for the EU.⁶ The growth and availability of eBooks has also expanded the library user base beyond researchers to include more of the general student population, who use different resources than the core researchers.

Global User Base of Academic / Scholarly Information



Measuring the value of content to users

Considering users and the changing environment is important when measuring value. There are many factors that may influence the value perception of library services and content. Making sure that the factors with the most valence to the users are measured is key to keeping the library in step with user needs while balancing faculty requests and budget pressures.

Compared to print, digital media increases the opportunities for measurement and tracking is less costly, but complexity is increased due to the number of variables, measurement types and volume of data.

Usage Statistics

Usage statistics can be recorded for all digital scholarly content, starting immediately after publication. Usage data can be a rapid indicator of scholarly trends, and according to the MESUR⁸ report, a better term for this data is "attention data". Utilizing Usage data comes with significant challenges – to begin with, defining what exactly is meant by usage: it could be downloading a .pdf file, downloading an abstract, emailing an abstract, clicking a link to view an article or perhaps even clicking within an article.

In general terms, this new form of information seeking behavior can be characterized as being horizontal, bouncing, checking and viewing in nature. Users are promiscuous, diverse and volatile and it is clear that these behaviors represent serious challenges for traditional information providers, nurtured in a hardcopy paradigm and, in many respects, still tied to it. Libraries must move away from bean counting dubious download statistics, and get much closer to monitoring the actual information seeking behavior of their users.⁷

⁵ STM report p. 27 - 28 / ⁶ STM report p. 21

⁷ Ciber report: Information Behavior of the Researcher of the Future. A Ciber Briefing Paper, January 2008. p. 9

⁸ Bollen, Sompel, and Rodriguez, Towards Usage-based Impact Metrics: First results from the MESUR Project, arXiv:0804.3791v1 [cs.DL] 23 Apr 2008

There also are several factors that should be noted when measuring and interpreting usage statistics to make sure comparisons between resources are relevant. For example, the age of the publication can impact its usage rate, and there are great differences in how publications are used over time. The peak age of journal articles varies by discipline. For humanities, it is about 20 years, but for chemistry, engineering and medicine, it is around 5 years.⁹

Subject area constituencies also influence usage statistics and care must be taken when comparing publications or publishers. For example, the 420 articles read annually by faculty members in medicine is nearly three times the 156 read by humanities faculty. Faculty members from scientific disciplines are in the mean, averaging 348 articles per year.

Springer's research indicates that inclusion of eBooks in a library's collection broadens the user base to include more students from the general population. These students are in addition to the core base of users which are more typically faculty and researchers. This broader population of users has a flattening effect on the usage of materials in the library collection. The result of this is that statistics on individual publications are diminished as the usage of materials is increased but spread out over more titles. Users are still deriving value from the content, only they are increasingly using more titles, and using them differently.

One new alternative measurement that utilizes usage data is the Usage Factor. This calculation uses statistics generated by COUNTER, which tracks monthly full-text article requests for many scholarly journals.

The Usage Factor is calculated as follows:

Total usage of a journal (Counter journal usage for a specified period) is divided by the total number of articles published online during the same period.¹⁰ The result is a number that illustrates the relative value of a particular journal in the context of other like publications. Viewing the usage in the context of all journals is helpful in understanding the value of a single publication, but again, it is only one measurement to consider.

Another factor that should also be considered in measuring usage is the impact of linking to an abstract versus linking to the full text of an article. In following a link to an article abstract, a user may very quickly determine the relevance of the piece without spending significant time reading the text. In this instance, what is often more important in this instance is not measuring how much time a user spent on a given piece of content but understanding what the user did next. Did they link to the whole article? Did they follow a hyperlink in this abstract, perhaps to a more relevant piece of content? In both cases, the time spent with the abstract would be brief, but its relevance to the user may be very high. As articles and texts increasingly become part of a network of connected texts, the linking trends of a particular article may be as important a factor in gauging its value as the more traditional access and usage measurements.

The MESUR (Metrics from Scholarly Usage of Resources) project is capturing and measuring these types of citation and usage connections. As noted above, it is becoming increasingly important to view content not as atomized texts, but as part of a larger network of connected information. MESUR is developing measurement statistics that look at a document across three key factors: document type, agent (author, publisher, user, etc.) and context (citations, co-authors). By examining a document in this manner, particularly in looking at the 'citation networks' that develop as documents are accessed and then used to locate other relevant documents, librarians will be increasingly able to make determinations about the intrinsic value of a piece of content as well as the importance of that content in the larger network of the library collection.

The Impact Factor, Findability and Usability

A goal of some of the more qualitative measurements is to explore not how often a resource is used, but rather the context and the structure of the reading patterns it is used in. For example, there might be a paper that is not read very often but is a key support for another highly used paper, by a high profile scientist.

An important form of qualitative feedback is responses from faculty and users, which can be gathered a number of ways and used to measure the specific needs of the library's unique user base.

⁹ STM report, pg 29-30

¹⁰ Final report on the investigation into the feasibility of developing and implementing journal usage factors, Shepard, P.T. (2007) p. 18

One of the best known measures that use quantitative information to approximate qualitative usage is the Impact Factor. The Impact Factor is the number one ranking value for scientific journals, which has become a substantial part of any journal development discussion since it was first published in the late 1950s. Impact factors are a benchmark of a journal's value and recognize how frequently peer-reviewed journals are cited by other researchers in a particular year. The impact factor helps to evaluate a journal's relative importance, especially when compared with others in the same field.

The Impact Factor is the ratio of two elements – the total number of citations counted in the Impact Factor year Y divided by the number of articles published in the previous two years. Citable articles are somewhat loosely defined. In general, they include original research articles and review articles. They may also include editorials, if the editorial contains a lengthy reference list. EBooks are not included in the Impact Factor, so it may not give an accurate picture when researchers link back and forth from eBooks and journals. A new development, the 5-year journal Impact Factor, was released in 2009 by Thomson Reuters that tracks citations over a longer period of time. The 5-year journal Impact Factor is the average number of times articles from the journal published in the past five years have been cited in the JCR (Journal Citation Report) year. It is calculated by dividing the number of citations in the JCR year by the total number of articles published in the five previous years. "Long tail" content in stable disciplines like the humanities, where valuable information becomes more visible and gains usage over time, would have a higher JCR.

Like usage data, the value of the Impact Factor cannot be compared among different scientific disciplines. For instance, Microbiology journals have, on average, much higher Impact Factors than Mathematics or Engineering journals. The citation and usage patterns in these disciplines are entirely different, which affects the numerical values of their Impact Factors, so when making purchasing decisions, care needs to be taken that journals are compared with others in the same field to yield appropriate results.

Findability measures, like page rank, capture how well content is optimized for search engine indexing and are an indicator of how easily it can be found using the commercial search engines. The number of denials – when content is found but not accessible by library resources – is somewhat opposite of findability and can indicate lack of resource breadth and contribute to user dissatisfaction. Measuring user satisfaction with the breadth of resources should be included in general user surveys, and this data should be considered when making content choices.

The "Y factor," a measure that combines traditional impact factor data with "page rank" data similar to what Google uses in its search algorithm, was introduced in 2006 by researchers at the Los Alamos National Laboratory. Google's ranking system keeps track of which Web sites link to other Web sites, much like article citation. However, Google also factors the prestige of a linking site into its calculations; if a major research institute's site links to a Web page, this means more than a link to the same page from a personal blog. The Y factor applies the page rank approach to journal citation networks. In studying the measure, they found significant differences between the highest impact factor journals and Y factor journals in physics, computer science, and medicine, and less difference in a subspecialty of medicine, dermatology.¹² In other words, it is important to consider not only if a work is being cited, but rather who is citing it, in determining the value of the content. Cutting content that shows lower usage or impact by traditional methods may backfire if the Y factor is high – researchers may miss out on quality content.

The Eigenfactor works like Google's PageRank, both of which are based on social network theory; where Google follows page links, Eigenfactor uses citations. They evaluate the importance of each journal (or Web page) based on the structure of the entire network. Developed by Bergstrom, the Eigenfactor provides an online suite of tools that "ranks journals and is available at no charge. The Eigenfactor attempts to account for the prestige of citing journals; incorporates many non-standard items such as newspapers and PhD dissertations into the citation network; and evaluates items over a 5 year, rather than 2 year period. Of particular interest to librarians, the "cost-effectiveness search" relates this data to the going subscription rates for journals as means of determining value-for-money. The "Article Influence" metric within the Eigenfactor is comparable to the impact factor, but that is just one aspect of the broader framework.¹³

In keeping up with social networks and the interest in blogs as an emerging medium, there is an even more novel evaluation metric, termed the "Blog Citation Index" or BCI. An increasing (but still small) amount of scholarly discourse occurs via blog posts; the comments on these posts; and the links to the posts from other sites (which are akin to citations). The BCI would formally track these connections and provide a measure to

Despite criticisms of the measure, journal impact factors tend to be stable over time, which validates their usefulness as an indicator of the most prestige of a journal.¹¹

¹¹ "The history and meaning of the journal impact factor", *Journal of the American Medical Association* Vol. 295 No 1, Garfield, E. 2006 pp. 90-93.

¹² "Refining dermatology impact factors using PageRank", *Journal of the American Academy of Dermatology* Vol. 57, No 1. Dellavalle, R.P., Schilling, L.M., Rodriguez, M.A., Van de Sompel, H., Bollen, J., pp. 116-119.

¹³ "Journal Status", *Scientometrics*, Vol. 69 No 3, Bollen, J., Rodriguez, M.A., Van de Sompel, H., pp. 669-687

compare among articles. The blogosphere is a radically open place that has blossomed in recent years. The BCI proposal is an extension of earlier calls for “recommender systems” to enrich and track new forms of peer review¹⁴ as a measure of the value of content.

Looking ahead, there are other factors which can be measured when building a user-centric environment. Springer research has identified key factors in an “ideal” online research tool. The first is usability: usability is a term used to denote the ease with an individual can employ a particular tool or other human-made object in order to achieve a particular goal, which is an important factor to measure when building a user-centric environment.

Other important factors identified are fast downloading times, and five core functionalities.

The top five functionalities include:

- ▶ the ability to find related articles
- ▶ citation tracking
- ▶ search within a journal
- ▶ ability to export bibliographic info, and
- ▶ ability to search other databases for articles by the same author

In other words, usage and perceived value is affected by the technology and design of the platform – how easily the researcher can navigate to the desired content as well as its value once accessed. These usability factors should be investigated before buying decisions are made – they not only impact usage, can reflect on the overall impression of a library’s usefulness to the research community it serves.

Wrap up and Moving Forward

As the amount of content grows and becomes more dynamic and libraries strive to become more user focused, the tools used to determine the value of content should evolve as well. Finding effective ways to measure the value of content is a complex challenge. It is not enough to measure how often a resource is used, it is also important to consider how it is used – the more qualitative factors, like the Impact Factor, page ranking, Eigenfactor and the linkages and interdependencies that characterizes modern, interdisciplinary science research.

- ▶ Multiple factors should be taken into account when evaluating to make sure measures reflect the value to the changing needs of the user base. Librarians should learn about the strengths and weaknesses of the alternatives; contribute to the discussion about their ongoing development; encourage and facilitate faculty awareness of these alternatives; and begin to use them in deciding how to develop and promote scholarly resources.
- ▶ Compare “apples to apples” when evaluating content for purchase decisions, Take into account subject matter and other differences to make sure comparisons are valid. For example, consider newer metrics, like the Blog Citation Index, for areas that are changing rapidly, such as biomedical research, or the Five Year Impact Factor for stable content that grows in influence over time.
- ▶ Look closely at alternative content providers. Ideally, they should be providing choice, depth and breadth in versatile platforms plus premium content. Their portals should be very usable, open to referrals from search engines, and contain previously hard to find ‘long tail’ content. Overall, they should help to facilitate and interact with new research methods and user demands.

About Springer:

Springer is a leading publisher developing a leading edge platform and actively contributing to industry debate to move the industry forward.

Through SpringerLink, springerlink.com, the Springer eBook Collection spans the world of science, technology and medicine, publishing more than 3,000 eBook, eReference Work and eBook Series titles each year. The collection offers high resolution illustrations, exceptional search capabilities and bookmarks. The eBook Collection speeds research by interconnecting all content on SpringerLink. Institutional licensing options are available.

To use the Eigenfactor, go to: <http://www.eigenfactor.org/>

¹⁴ “A blog citation index?” <http://ericsschnell.blogspot.com/2007/11/blog-citation-index-bci.html>. Schnell, E