Chapter 1.22
Electronic Usage Statistics

Patricia Hults
Rensselaer Polytechnic Institute, USA

ABSTRACT

This chapter provides an overview of electronic usage statistics, including methods of defining, collecting, and using the data. A survey of some of the systems of estimating journal usage in the print environment is followed by a description of the development of electronic usage practices. The important contributions of the COUNTER and SUSHI projects are reviewed, along with issues in the management and use of electronic statistics. Examples of ways these statistics can assist in decision making throughout a product's life cycle are included, as well as other ways usage statistics can prove useful. The chapter concludes with a brief look at the use of statistics in the bibliomining process.

INTRODUCTION

Unless you have a mathematical bent or are one of those individuals who find satisfaction memorizing the major league baseball stats, the topic of user statistics is not immediately intriguing. In fact, it can be mind numbing and tedious, but user statistics are extremely useful, particularly now that we are able to get real, meaningful information—they cannot be ignored. This chapter will start with an examination of early, pre-electronic usage statistics. It will then look at the development of electronic statistics, including both the COUNTER and SUSHI standards. Management issues in collecting and using these statistics will be explored. Some of the applications of these data will be discussed, in the context of an electronic product's life cycle. The value of usage statistics beyond just product evaluation will also be covered.

Generations of librarians have struggled to find ways to practically measure usage of the material they so carefully select. The information on just how many times a book or journal was used is critical in both selection and retention decisions, and in broader collection development strategies. Without a sense of how many times something is used, it becomes impossible to evaluate its worth.

BACKGROUND

Books have always presented less of a problem. You could count the number of times a book was checked out, whether you were counting circulation cards or looking at automatically generated circulation statistics. There was still the bugaboo of in-house use, but there were significant amounts of real circulation data available. Journals presented a much larger challenge. Not only were many libraries organized so that journals were never checked out and sat on open shelves; journals came both bound and in single issues, so that the unit of count was unclear. Indexes, while officially in book format, generally never circulated and therefore their use was as hard to quantify as it was for journals.

Because librarians are an ingenious group, all sorts of methods were devised to estimate in-house use of journals and books and in general to evaluate the worth of a particular title. These efforts ranged from using photocopying requests (Cooper & McGregor, 1994), making correlations between check-outs and in-house use (Walter, 1996), counting journals left on study carrels and near photocopy machines (Bader & Thompson, 1989; Chen, 1972), sticking voluntary usage log sheets on journal protective covers (Konopasek & O'Brien, 1982), and more. Some librarians sent their work-study students skulking in the stacks, trying to measure the ratio of actual vs. recorded use.

Other efforts included using external criteria such as journal impact factors and citation analysis (McCain & Bobick, 1981; Rice, 1979). The journal impact factor is a measure of the number of times a journal is cited in published articles. Interestingly, at least one recent study examining electronic usage and impact factor found no correlation between impact factor and local use of the journals. Duy and Vaughan (2006) examined use of electronic journals from three major vendors; the American Chemical Society, Elsevier, and Wiley, and they found there was not a correlation between impact factor of a particular journal and actual use of that journal on their campus. What they did find more predictive was a local citation figure, calculated by determining how many times a specific journal was cited in articles by campus faculty.

DEVELOPMENT OF ELECTRONIC STATISTICS

Librarians continued the tradition of ingenuity when journals, books, and databases began to be available in electronic format, and they quickly began trying to extract more reliable statistics from the new medium. Before publishers began supplying usage information, librarians explored other sources including institutional Website logs, statistics supplied by A-Z list providers, and those generated by link resolvers. While each of these offered interesting insight into patterns of use, they fell far short of accurately and fully capturing the information librarians sought. Unless an institution had loaded the product on its own server, it was clear that publishers would be the primary suppliers of usage statistics. Initially this data varied widely in what was being measured, and many times, what was being measured was fairly meaningless. An example is the number of pages called up from anywhere within the publisher's site, including help pages, menu pages, and so forth. This type of count served only to create an inaccurate impression of use.

In response to pressures from librarians and for their own internal management needs, publishers began attempting to measure journal usage. Some began to supply pages that captured the number of downloaded files from a particular site. This was progress, but it was still very messy. Article and chapters were often divided into multiple files to reduce download time and each component of a single article might be counted as an individual use, greatly inflating overall usage rates. One publisher's response to statistics requests was to