Scan it and they will comewill they cite it?

Michael Fosmire
Purdue University, fosmire@purdue.edu

Scan it and they will come…but will they cite it?

Michael Fosmire[1]

SUMMARY.

As the number of retrospective digitization projects of journal content increases, there is a need to assess the impact of these projects on the productivity of researchers. Librarians making collection development decisions about acquiring these backfiles need to know how useful they are to researchers. This study provides data on usage of a range of years of the *Physical Review*, and citation information from *Physical Review Letters* to other *Physical Review* articles. The usage of the online archive of *Physical Review* articles indicates that articles are accessed all the way back to the first issue, with an average number of downloads on the order of ten per article per year. Both usage and citation rates show exponential decay rates, however, with different intrinsic time scales. The citation half-life is consistent with previous studies of the physics literature, while the usage half-life computed here is in conflict with older analyses of print usage of the physics literature, although in line with some recent online usage studies in medicine. An analysis of the citation data indicates a potential order of 10% enhancement in citations to articles available in the online archive, but the statistical error is of the same magnitude, so no firm conclusions can be drawn from that data. A few more years of citation data may be able to resolve the question of impact of the online archive on citation rates.

KEYWORDS scientific journals; citation analysis; online usage;

[1] Michael Fosmire, MS Physics, MLIS, is Assistant Professor and Science Librarian, Purdue University, West Lafayette, IN. email: fosmire@purdue.edu. The author is indebted to Martin Blume, Mark Doyle, Claire O'Neill, and Gerald Young for providing generous amounts of usage and citation data for the APS journals.

INTRODUCTION

It is hard to believe that only five years or so ago electronic journals were still a novelty and not a way of life for researchers. Now that almost all STM publishers publish online versions of their journals, the next decision that the industry is facing is what to do about material that predates their online publishing initiatives, i.e., whether they should digitize their backruns. And, for librarians, the relevant question is, should librarians purchase this retrospective content, and what would be a fair price? What added value does the online content for older material provide, in terms of additional productivity for researchers? Despite initial skepticism on the part of many publishers, who didn't believe anyone would use the old stuff, the results of the American Physical Society's conversion of its entire journal run show that, if you scan it, they will come (see Figure 1).

However, the downloading of articles is only one measure of the usefulness of a journal backfile. Another is the rate at which scientists integrate older materials into their research, compared to the rate they did when the articles were only available in print. If a goal of digitization is to increase productivity by allowing the older material to more easily be integrated into current research, one way that may show up is by an increased level of citations to the older literature.

A quick look at Thomson ISI's *Journal Citation Reports* shows that the *Physical Review*'s article half-life (an indication of how long an article is relevant to research) has increased, fairly monotonically, from 1998 to 2002. *Physical Review D* hasn't seen a change in the article half-life, but the other sections have seen a 5% to 40% increase in their half-life. This may indicate that the older material is being used increasingly over time, coincident with the arrival of the online backfiles. However, data prior to 1998 was not available to the author, so it is unclear whether this increase was already taking place prior to the existence of the online backfile.

Several studies have looked at the obsolescence of the physics literature. Gupta's (1990) citation study found an exponential decrease over time in the density of citations to the *Physical Review* from fifteen major journals in physics, and found a half-life in the citation density (Gupta's density corrects for the effect of growth in the literature over time) of about five years. Chen's (1972) shelving study found an obsolescence rate of about 14.5 years, and Sandison (1974) reinterpreted the results to show that there was no obsolescence rate at all, that is, the physics literature remains relevant and its utility doesn't decay over time. However, the Chen and Sandison studies were limited by low statistics and, potentially, systematic sources of error, and did not rigorously correct for changes in the gross research output published in physics at that time.

Brown (2001a) has investigated a situation (e-prints from arXiv.org) where convenient online access to physics research is available, albeit unrefereed. She sees large citation rates for these preprints, where, for example, about 2/3 of all articles submitted to the high energy theory preprint server were ultimately cited somewhere in the time interval investigated. Electronic preprints may be cited at a high rate because they are the newest research available, or they may be cited preferentially because they are easily and freely available. This study may help to disentangle those two drivers of research citation patterns. If the former is a main driver of citing behavior, there should be little effect on the overall citation rate from access to the *Physical Review* backfile. If the latter is more valid, there should be some noticeable result.

Also, if the information system of physics has been efficient in communicating relevant results through time (through complete and accurate referencing, for example), then perhaps having an easily searchable and accessible backfile is less important, since relevant research has already been identified. However, if transmission of the relevant results is imperfect, or if research in new fields, or new discoveries, requires use of previously unused data, then an increase in citation levels might appear.

METHODS

The American Physical Society's (APS) journals were chosen as subjects to look for potential citation behavior changes as a function of access to the online version of the journal. Since the APS was the first major physics publisher to provide online access, and the first to make their entire backfile available online, the statistics on their journals are the most complete. As premier journals in physics, the *Physical Review* sections also provide a good cross-section of high-quality research in all areas of physics. The APS, unlike some of the latecomers to the archival market, has also made their backfile available at a reasonable rate to the library and research community, so access to the backfile of their articles is as widespread as for any publisher.

However, compiling the citing data manually appeared to be a daunting task. In 2002 alone, over 17,000 articles were published in the *Physical Review*. As tracking a statistically significant subset of these articles, over several years, did not appear feasible, the publishers at the APS were asked whether they already had some of that data, and if they were willing to share it. Editor-in-Chief, Martin Blume, and Manager of Product Development, Mark Doyle explained that, since they had created Digital Object Identifiers (DOI's) for all their articles, and linked all their references to other articles in the APS online journal platform, they had all the citation information in their publication database and could run a query to extract the information needed for this study. The ability to automatically process the over 3 million such links, made the execution of this project a reality and, certainly of a much higher quality than could have been achieved manually.

To get the largest possible statistics of citations to the *Physical Review* database, one logical place to go would be the *Web of Science*, since it is the largest citation database. The *Web of Science* indexes several thousand journals, while this paper only analyzes the citations of one journal. However, that would have required manually tracking citations for tens of thousands of articles, which was a prohibitive endeavor. Furthermore, since the annual output of scientific papers in constantly increasing, an extra variable is introduced if only bulk citation counts are tabulated. One would not be able to disentangle the possibility that citation numbers had increased due to the fact that more papers were being published, or if they were just being cited more often than previously. Since APS had all the data in-house, both citations and total publishing output, and it was manipulable in an automated fashion, analyzing citations from the APS journal articles to APS journal articles constituted a much more controlled environment. This decreased the overall statistics in the sample, but it did reduce the number of uncontrolled variables in the analysis. The next question then became, which APS articles to use in this analysis.

SELECTION OF JOURNALS FOR STUDY

The APS publications contain three general types of journal articles. Lengthy review articles are published in *Reviews of Modern Physics*, full length research articles are published in *Physical Review* sections *A-E*, and shorter communications are published in *Physical Review Letters* (and some short communications are published in the other *Physical Review* sections as well).

In analyzing citation patterns, attempts were made to limit potential systematic sources of error. Admittedly, that is a difficult task, and can be only imperfectly done. For example, hot topics change from year to year, and, consequently, the citation pattern may change due to the different research histories in those areas. Also, the relative number of long or short articles would likely have different average numbers of citations, and would change the rate of citations per article from year to year.

Reviews of Modern Physics (*RMP*) was eliminated from the analysis, since only a few articles, relatively speaking, are published in that journal. With only on the order of fifty or sixty articles published a year, an analysis of the citation patterns would likely be dominated by the differences in subject matter published from year to year. Additionally, with the number of citations per paper significantly greater than in the rest of the APS publications, differences in the relative number of papers published in *RMP* versus the *Physical Revie*w sections, for example, would skew the citation rates of articles if they were lumped in with the rest of the *Physical Review* sections, so *RMP* was taken out of the mix entirely.

The *Physical Review A-E* sections have short communications and longer, full, research articles, so if the relative concentration of the two changes over time, that also would yield different results in the citation rates by year. Also, in 2002, coincidentally, the APS synchronized its online and print publishing cycle, so that an article's publication date corresponded to its online publication, rather than its publication in a print issue (which is typically on the order of a month or so later). The result is that the print version now comes out later than it did (and, for example, articles that might have appeared in the January 2003 print issue were placed in the December 2002 issue). This led to a 13% increase in articles published in 2002 over 2001, compared to a 3% increase in articles published in the previous year. Since the citation rate decays rapidly with time, the effect of the publication date shifting alone would lead to lower citation rates in 2002 relative to previous years.

However, the *Physical Review Letters* (*PRL*) did not see the same increase, and in fact had 3% fewer articles published in 2002 than in 2001. Since the *PRL* did not seem to be as affected by the change in publication cycle (and in general *PRL* has a shorter time from acceptance to publication than the other sections of the *Physical Review*, so it shouldn't have been as affected), the citation behavior of just *PRL* was analyzed . Also, the research in *PRL* is generally composed of exciting new findings, while the other sections of the *Physical Review* are more exhaustive treatments of an experiment or theory. Thus, the gestation period of the experiment is likely shorter for *PRL* than for other articles, and one would be able to see the effect of online access sooner in *PRL* than in the other sections of the *Physical Review*. In general, the curves of citation data were fairly similar between *PRL* and the other sections of *Physical Review*, so the qualitative conclusions would probably not have changed if the entire *Physical Review* had been analyzed. The *PRL* also has a tightly constrained article length of four pages, so the articles are very similar in size, and the variation due to article types mentioned above should be minimized by just analyzing those articles. The total number of *PRL* articles published per year is around 3,000, which contain about 20,000 citations to APS articles (see Table I).

PROCESSING THE DATA

The APS, from its publication database, provided the total number of articles published in each year from 1893-2002, which established a normalization baseline for this study. Also provided was a report of the number of citations from *Physical Review Letters* articles published in 1995-2002 to *Physical Review* articles in previous years (a composite of all sections), and, for comparison, the total number of citations from both *Physical Review A-E* and *Letters* to all articles in previous years.

In order to increase the statistics for relevant time periods, the data was combined from different years into three time periods, corresponding to the availability of portions of the backfiles. The chronology of events in the digitization process for APS is as follows. In 1998, the APS announced it had digitized and made available its backfile to 1985. At the end of 2000, they had digitized materials back to 1975, and in May of 2001, had completed the digitization back to Volume 1 in 1893. Data were combined from 1995-1997 to use as the control group, corresponding to no access to the electronic backfile. 1999-2000 constituted a second portion, where access back to 1985 was available. Finally, 2001-2002 provided a sample of data with access back to 1975, and, for 2002, back to 1893. Since the statistics before 1975 were rather small and didn't warrant a separate analysis, the 2001 and 2002 data were combined. The 1998 data were not analyzed, since that was a transition year, with online access being made available in the middle of the year.

In order to compare apples to apples over time, two normalization procedures for the raw citation information were carried out. First, since different numbers of articles were published each year, the raw citation information was divided by the number of articles published in the year being analyzed. This quantity was then divided by the number of articles published in the year of the cited article. This provides a double-normalized citation rate, c_{dn},as follows, for example, for citations from 2002 articles to those in 1977:

$c_{dn\,2002} = \dfrac{c_{raw\,2002}}{N_{2002} \cdot N_{1977}}$, where N_X = number of articles published in year X, and c_{raw} is the gross number of citations from 2002 articles to those published in 1977. In the graphs of citation data that follow, all data are doubly normalized in this way.

Also, since citation rates decay (rather steeply) with time (see Figure 2, with a factor of three decrease in citation rates after seven years), the appropriate comparison of data between years was to graph the rates as a function of Years Since Publication.

For regimes of the usage and citation data where exponential behavior was observed, a least-squares fit was done to compute a half-life τ, where the half-life is the time it takes for a citation rate to become half of what it was originally. Using a functional form of

$$c_{dn}(t) = c_{dn}(0)e^{-ln(2)t/\tau},$$

a least squares fit of the natural logarithm of the doubly normalized citation rate c_{dn} to the year (t), yields a slope of $ln(2)/\tau$.

Before discussing the results, a few caveats and details about the data are given. The APS data treats any object with a DOI as an article. Thus, in the early years of the *Physical Review*, for example, the book review section was counted as an article in the statistics, and some 'Minor Contributions' were indexed as one article, even though several small articles were lumped under that heading. Also, until 1897, the *Physical Review* had an 'academic' publishing year. So, 1893, for example, runs from July of 1893 to June of 1984. 1897 contains only a half-year's worth of articles, as the APS synchronized their volumes to the calendar year. All of these discrepancies are fairly minor, in the gross analysis, but they do lead to small changes in the 'actual' citation rates of the data.

RESULTS

The usage statistics, provided by Claire O'Neill, Data Analyst at APS, and detailed in Figure 1, show that users download articles from every year, with even the early years showing about three downloads per article per year. The first few volumes of the *Physical Review* show much larger download rates.

When looking at the download data, there appear to be two exponential trend lines. In the first four or five years, the exponential has a large decay rate, which is overtaken by a slower decay process that extends to the end of the statistically significant portion of the data. The half-lives computed for these two processes are 2 years and 13 years, respectively.

The associated citation information, as visualized in Figure 2, of 2002 citation data, shows similar trends to the downloads, although, notably, no citations to articles before 1913. The citation curves for all years (1995-2002) look qualitatively the same, as can be seen in Figure 3. The peaks and valleys of the download and citation data in the area before around 1950 are fairly randomly distributed relative to one another, and are likely just random fluctuations. Indeed, plotting an average of all years of this study smoothes out the peaks in the citation data. The citation data also show two exponential dependencies, with a knee in the exponential curves around seven years from the publication date. The half-lives of these processes are 5 years and 10 years, respectively. Before 1960, the total number of citations is at most of order 30, so random fluctuations wash out any signal in that time interval.

In addition to year by year results of citation rates, the total number of citations were computed for each year from 1995-2002. The third column of Table 1 shows an apparent dramatic increase in citations per article published between the pre-electronic access regime, and the era since the complete backfile was made available, with a sharp transition during the years when the backfile was implemented. However, when one factors in that each year several thousand more articles have been published and thus are available to be cited, (see the last column of Table II), the citation rate per article published that year per article in the *Physical Review* from 1893 to that year is very consistent. Thus, one needs to continuously cite more and more articles in the reference section of a paper in order to maintain the same overall citation rate over time.

Finally, the original purpose of this paper was to analyze what difference in citation behavior online access to the journal may provide. Figure 3 shows data plotted for the time intervals when no online access was available to the backfile (1995-1997), access was available back to 1985 (1999-2000), and access was available at least back to 1975 (2001-2002). Although the data look almost the same, there is a small enhancement of citation rate in the more recent data, as Figure 4 shows, when the 'background' citation rate from 1995-1997 is subtracted from the other years' data. The average of the enhancement shown in Figure 4 is computed and tabulated in Table 2.

DISCUSSION

It is very interesting that both the download rates and the citation rates show two exponential trends. It is certainly reasonable to assume that there are many different ways the literature is consulted and cited, and that each one may have its own decay curve. Burton and Kebler (1960) created a robust analogy between radioactive half-life and the half-life of literature. Fitting citation patterns for several disciplines, Burton and Kebler actually determined that their data did not fit one exponential curve, and hypothesized a citation format with two exponentials, one with twice the decay rate of the other. Although there was no intrinsic motivation given for using that particular functional form to fit the data, the concept that literature is made up of many kinds of articles, each with its own half-life, is compelling.

Burton and Kebler posit that 'classic' literature has a long half-life, while 'ephemeral' literature has a much shorter half-life. Thus, with enough statistics, one could easily expect different half-lives for citation and usage data. Potentially, there are many ephemeral papers with short half lives, that dominate the statistics for short time periods, but after several years, their usage disappears, and the usage of the 'classic' papers with long half-lives becomes the dominant process. Just as Gupta's (1990) paper proclaims the first conclusive evidence of an exponential decay rate in the citation of the physics literature, the present study is the first one the author is aware of that shows evidence of two distinct exponential decay rates. as first predicted by Burton and Kebler, and these decay rates show up in both usage and citation data.

To provide motivation for the existence of multiple usage rates, one may consider the following. In the first year or so of an article's life, it may be browsed in a current awareness capacity. One would expect a very short half-life for this activity, maybe as short as a few months (and, indeed, the download rate for the first year after publication is much higher than that of the preceding years, even steeper than a two-year half-life would indicate. Monthly, or even weekly, data would need to be analyzed to determine this very brief usage process). This current awareness browsing behavior would not be restricted to a scientist's specific research area, so one would expect a lot of amplitude in this channel.

Another use of the literature is as background for a specific research project, perhaps the result of a literature search. This would have a smaller total amplitude than the current awareness use, so it would not be visible in the data from the early years. But, over time, classic results stay relevant, so the half-life of usefulness would be much longer, perhaps the ten years computed here. There are also probably classes of articles, ranging from the seminal article which can be cited for hundreds of years, to the more incremental and applied, Burton and Kebler's ephemeral article, which has a short shelf life until a seminal article comes along and makes that article irrelevant for future research. Perhaps the five-year window of the two year half-life reflects the usage of the ephemeral articles, while the ten-year half-life refers to use of the seminal articles, whose usage will dominate the statistics when the ephemeral articles have ceased to be useful. Or, perhaps the five-year window still reflects a general browsing behavior of users, and the second decay rate reflects usage related to actual research purposes. More detailed study of specific articles and correlation with citations would shed more light on the actual processes yielding these characteristic lifetimes.

Other usage studies of the physics literature, notably Chen (1972) and Sandison (1974), show usage obsolescence rates of 14.3 years and infinity (and even negative half-lives), respectively. Thus, this study's initial download half-life of 2 years is much different than those results, and is more in line with other reports of online usage (for example, Anderson *et al*. 2001, for the journal *Pediatrics*). However, the longer half life shown in Figure 1, compares with Chen's findings for obsolescence.

Print-based usage surveys have significant systematic sources of error, however. For example, the most highly used issues are only shelved a finite number of times per day, and several people may have consulted the issue between shelvings. Also, a browser would perhaps consult many articles within a recent issue, but only one use is credited in the study. However, for earlier years, a reader is likely to only consult one article per use of a volume, in answer to a specific information need. Finally, researchers with personal subscriptions will show no usage in the library. Researchers are less likely to have older issues of journals in their possession than newer ones, so, again, the bias points to more library usage of older journals, increasing the apparent usage half-life. Thus, it is conceivable that Chen's study is not sensitive to current awareness browsing, but is sensitive to slower processes. However, for example, Chen's data does not take into account the increase in the total volume of literature published, which is corrected for in this study, and when Sandison takes that into account, in a very crude fashion, Chen's half-lives become infinite, or, indeed, the usage density (per article) in fact becomes larger with time. The findings here definitely do not support Sandison's conclusion.

The online usage data from this paper does eliminate many of the systematic problems of earlier usage studies. First, it measured all online usage of the *Physical Review*. This eliminates the variation in local journal usage studies caused by an institution having strength in one particular field and not another. It also compiles usage for everyone, everywhere in the world, and thus gives a good description of the behavior of all users of the physics literature. It is irrelevant whether a researcher has a personal subscription or an institutional subscription, since all uses are counted. The only data which is neglected is for people who browse the print collection. This is a non-trivial loss of data, but, as has been reported elsewhere (for example, Morse and Clintworth 2000), online usage swamps whatever numbers are gained in a journal sweep.

At Purdue University, I have noticed that *Physical Review* is not being reshelved nearly as often as in previous years, as users have switched to reading it online. Indeed, since the library ran out of space to house the entire print run locally, pre-1990 issues were moved into local storage without a peep from users. This is from a library in which, in the early 1990's, a considerable uproar was made when the subscription went from two print copies of the *Physical Review* (in one physical location) to one. An especially robust usage study of the medical literature by Tsay (1999, 1998) also shows an exponential usage rate for print materials, with a mean use half-life of 3.43 years, much closer to the findings of this paper than those of Chen or Sandison.

The citation rates contain similar exponential decay rates, again with two characteristic half-lives. The initial decay rate of this study, at 5 years, is fairly consistent with the results of Gupta's (1990) 4.9 years, and Burton and Kebler's (1960) 4.6 years for the physics literature. That these half-lives are so consistent is, in some ways, amazing, since they sample citation patterns of authors from the 1950's, the 1980's, and, here, the 21^{st} century. Over fifty years, the half-life of the literature has remained constant, despite the many changes in the way physics is done and the kinds of problems that are being tackled.

Gupta points out that, comparing results to the work of Chen (1972) and Sandison (1974), perhaps, the citation and usage data are not measuring the same thing. Tsay (1998) qualitatively supports this supposition of a difference, for the field of medicine, albeit with much different qualitative results than Chen or Sandison, and the current study confirms a difference in the data for physics. With such different characteristic time scales, the activities have some fundamental differences. For example, current awareness searching has no direct effect on citation habits—one reads many papers that don't end up being cited. And subtracting out the body of articles with half-life zero (i.e., are never cited) would tend to increase the effective half-life of citation data over direct usage data.

However, the fairly close correspondence between the longer half-lives for usage and citations (13 years and 10 years, respectively) is interesting. Perhaps this does show that usage and citation rates are measuring the same fundamental process for use of seminal articles of the literature. More investigation of that connection is necessary to draw any firm conclusions, however.

The tails of the distributions in Figures 1 and 2, the peaks and valleys from 1893-1950, are relatively randomly distributed. Indeed, when 1995-2002 citation data is averaged, the peaks and valleys pretty much disappear entirely. Also, the high download rates (up to 100 downloads per article per year) for early issues of the *Physical Review* are likely due to curiosity and/or historical research, rather than to any relevant results for current research, as evidenced by a lack of citations in *PRL* in that time interval to any articles before 1911.

Another difference between the Gupta results and those from this study, is that the peak citation rate here comes at one year after publication. Gupta sees the second year having more citations than the first after publication. This is a reasonable difference, because *PRL*, as a Letters journal, has a quick turnaround, and is meant for rapid communication of new developments, while the physics literature as a whole (which was sampled by Gupta) contains predominantly articles that are exhaustive examinations of a topic, and thus take longer to carry out, analyze, review, and get published. Thus, the second year enhancement in the Gupta data is likely due to this gestation period of carrying out research and getting it published. Indeed, when I looked at citations from all sections of the *Physical Review* to the *Physical Review*, the citation rates were about equal in year one and year two after publication, generally consistent with the Gupta results, and different from Tsay (1999) who found a maximum citation rate in the third year after publication in medicine.

The next step is to see whether detailed analysis of the citation levels in recent years shows any enhancement over the control years. The relative citation rates between the pre-electronic access period, the initial backfile, and the complete backfile look very similar (see Figure 3). In order to identify potentially small differences in the citation patterns, the 'background' of the average data from 1995-1997 was subtracted, and the difference was plotted for the two sets of data corresponding to the availability of the online backfile (see Figure 4). This shows that there is a net positive effect in both the 1999-2000 and 2001-2002 time intervals for citing materials that were available online. The mean values of this enhancement are given in Table 2. This is about a 10% effect on the total citation rate (the subtracted background). As can be seen, there is a weak positive correlation between online access and increased citation level, significant to two standard deviations for the combined data.

Checking before 1985 (the last two columns of Table 2), as a control to see if there is any systematic increase in citation rates irrespective of online access between those years, one finds that there is still a positive enhancement. For the 1999-2000 data, which is the control, it is a slightly lower level of significance than for the 1985-present data, while the 2001-2002 data, which should provide whatever signal is available, has a lower absolute enhancement, but the error is much smaller, yielding results of three standard deviations significance. The results from 1975-1985 show that there could likely be another factor at work in the enhancement of citation rate, since there should be no enhancement for the 1999-2000 data, and yet one shows up. The 2001-2002 data for this time interval has an anomalously low error level compared to the error levels of the other data, so it is perhaps just the result of a fluke correspondence in the data and is without real statistical significance.

Before 1975, the potential enhancement computed is a factor of ten smaller than the error, so no signal is apparent. As mentioned before, the total number of citations in a year up through the 1960 is at most of order 30, so attempting to find a signal is extremely difficult.

One thing that should be noted is that the 2002 data (like the 1996) data has a lower absolute citation rate than the other years (see Table 1). I believe this leads to some of the reduced enhancement in citation rates for the 2001-2002 data compared to the 1999-2000 data. Whether this is a fluke for 2002, or a real trend, it will take more years of data to determine (and again, the effect of the changed publishing cycle for 2002 is unknown). Perhaps a study with a few more years of data would produce enough statistics to provide more solid answers, and it should be easier to determine whether the decrease in the citation/article/article rate is real or not. The fact that each four-page *Physical Review Letter* article averages over six references to *Physical Review* articles, leads one to wonder whether the increase can continue indefinitely, or whether, as some pundits have noted, papers will eventually just become a collection of references to other work.

One unresolved question is the effect of e-print citing on the overall citation rate for *Physical Review* articles. The APS suggests that authors submit their papers to arXiv.org before submitting them to the *Physical Review*. This allows for an informal period of peer review before proceeding through the official certification process, and major errors can be corrected outside of the formal process of peer review. It leads to faster editorial reviews for the *Physical Review*, but it also means papers might be citing the e-print version of an article that ends up in the *Physical Review*, rather than the final, published version, and this might erode the overall citation rate for APS articles. As Brown (2001b) has discovered, the citation rate for arXiv papers is substantial and (Brown 2001a) still increasing with time. The effect of e-print citing habits is another topic worthy of investigation.

CONCLUSION

The usage data in Figure 1 do show that, ‘if you scan it, they will come.’ The APS backfiles are being used frequently, and, if nothing else, at least make it convenient for authors to actually consult the older articles they cite. The citation data further shows that articles are still cited in *Physical Review Letters* back to the early 1910’s. Double exponential behavior was seen in both download and citation data, which give a flavor of how long articles are useful for reading and citing, and show that there are likely multiple processes involved in the usage of the journal information. For example, with the download data, the large exponential decay rate is likely due to general browsing, while the smaller rate process is more likely related to focused research activities. The overall half-life for citation rates are consistent with previous studies (Gupta 1990; Burton and Kebler 1960), while the usage half-life is similar to recent online usage studies (Anderson *et al*. 2001; Tsay 1998), and markedly different from previous print usage studies in physics (Chen 1972; Sandison 1974).

The citation data is tantalizing but is not conclusive in determining whether online access to an electronic backfile has affected citation rates. One would not expect a huge effect on citation rate, since scientists have always been citing prior research, and, one assumes, already know most of the relevant results, from their own prior research and papers where they've cited past work. So, the 10% enhancement that is observed is perhaps of the right order of magnitude, or even higher than might be expected. However, as seen by the enhancement of the data in pre-1985 years for the 1999-2000 citation data, there might be some other systematic enhancement going on that this research was unable to discriminate. Perhaps future study of this type of data will be able to determine more definitively the nature of the observed enhancement.

REFERENCES

Anderson, Kent, John Sack, Lisa Krauss, and Lori O'Keefe (2001) "Publishing Online-Only Peer-Reviewed Biomedical Literature: Three Years of Citation, Author Perception, and Usage Experience," *Journal of Electronic Publishing* 6(3). http://www.press.umich.edu/jep/06-03/anderson.html

Brown, Cecelia (2001a) "The E-volution of Preprints in the Scholarly Communication of Physicists and Astronomers," *Journal of the American Society for Information Science and Technology* 52(3):187-200.

Brown, Cecelia (2001b) "The Coming of Age of E-Prints in the Literature of Physics," *Issues in Science and Technology Librarianship* 31. http://www.istl.org/01-summer/refereed.html.

Burton, RE, and Kebler, RW (1960) "The 'Half-life' of some scientific and technical literatures," *American Documentation* 11: 18-22.

Chen, C.C. (1972) "The Use Patterns of Physics Journals in a Large Academic Research Library," *Journal of the American Society for Information Science* 23:254-265.

Gupta, Usha. (1990) "Obsolescence of Physics Literature: Exponential Decrease of the Density of Citations to Physical Review Articles with Age," *Journal of the American Society for Information Science* 41(4): 282-287.

Morse, David H., and David A. Clintworth (2000) "Comparing Patterns of Electronic and Print Journal Use in an Academic Health Science Library," *Issues in Science and Technology Librarianship* 28. http://www.istl.org/00-fall/refereed.html

Sandison, A. (1974) "Densities of Use, and Absence of Obsolescence, in Physics Journals at MIT," *Journal of the American Society for Information Science* 25:172-182.

Tsay, Ming-Yueh (1998) "Library Journal Use and Citation Half-life in Medical Science," *Journal of the American Society for Information Science* 49(14):1283-1292.

Tsay, Ming-Yueh (1999) "Library Journal Use and Citation Age in Medical Science," *Journal of Documentation* 55(5): 543-555.

TABLE 1. Citation rates for *Physical Review Letters* to Other Articles in the *Physical Review*, by Year

Year	Articles	Citations	Citations per Article	Citations/article/ Article
1995	2574	12707	4.9	2.2E-5
1996	2636	12957	4.9	2.1E-5
1997	2632	14519	5.5	2.2E-5
1998	3045	18422	6.0	2.3E-5
1999	2822	18263	6.5	2.3E-5
2000	3001	19426	6.5	2.2E-5
2001	3055	20378	6.7	2.2E-5
2002	2986	20136	6.7	2.1E-5

TABLE 2. Increased Citation Rates for Years with Online Access Available, Compared to the 1995-1997 'Background' with no Online Access. The 15-year window corresponds to data since 1985 for 1999 data, and the 25-year window, to 1975 data. These correspond to the two years for which backfiles were posted, in 1998 and 2000, respectively.

Year	Mean Enhancement (less than 15 years from publication)	St Dev (σ) (less than 15 years from publication)	Mean enhancement (15-25 years)	St Dev (σ) (15-25 years)
1999-2000	2.7E-6	1.4E-6	1.6E-6	.99E-6
2001-2002	1.9E-6	1.3E-6	1.2E-6	.45E-6
Combined	2.3E-6	1.1E-6		

FIGURE 1. Downloads from *Physical Review* journals per article published that year, per year, averaged over 2001 and 2002. Guidelines are provided for the two characteristic half-lives. (Data courtesy of the American Physical Society.)

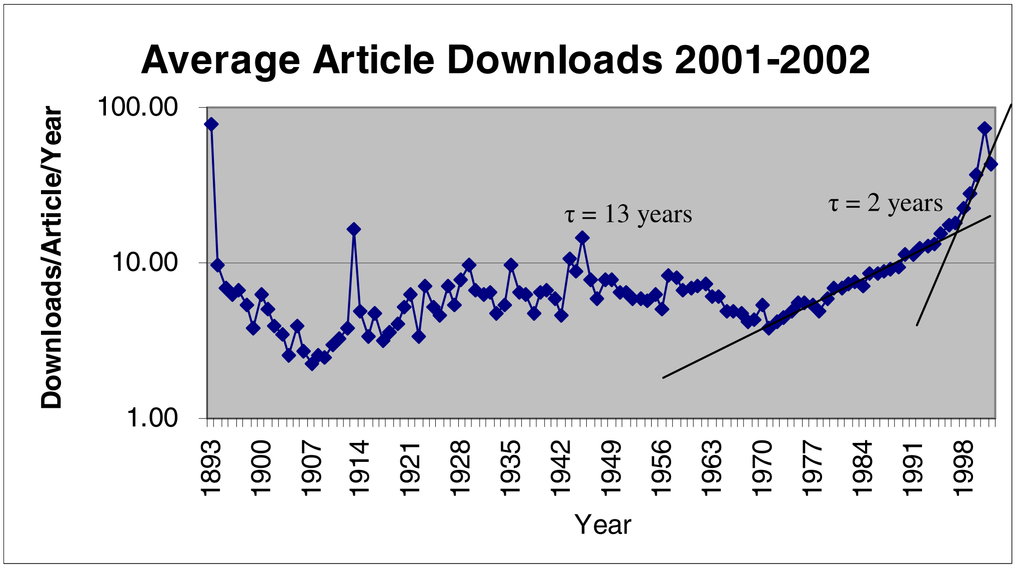

FIGURE 2. Citations from *Physical Review Letters* articles in 2002 to all sections of the *Physical Review*. Data is doubly normalized by dividing by the number of articles published in 2002 and the number of articles published in the year plotted. Guidelines are provided for the two characteristic half-lives. (Data courtesy of the American Physical Society)

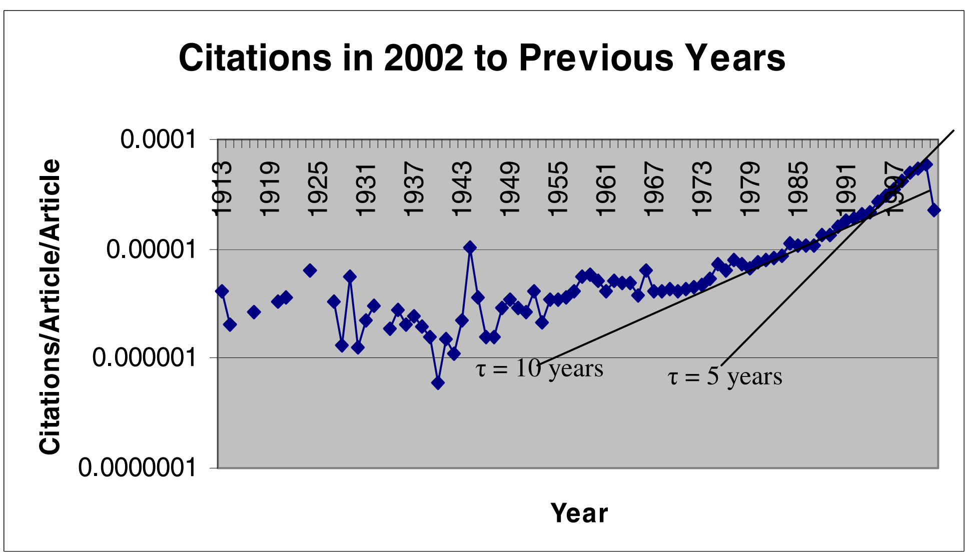

FIGURE 3. Doubly normalized citation rates from *Physical Review Letters* articles to all sections of the *Physical Review*, for three time intervals. The years 1995-1997 correspond to no access to an online backfile, 1999-2000 correspond to access to an online backfile to 1985, and 2001-2002 correspond to access to a backfile at least back to 1975. (Data courtesy of the American Physical Society)

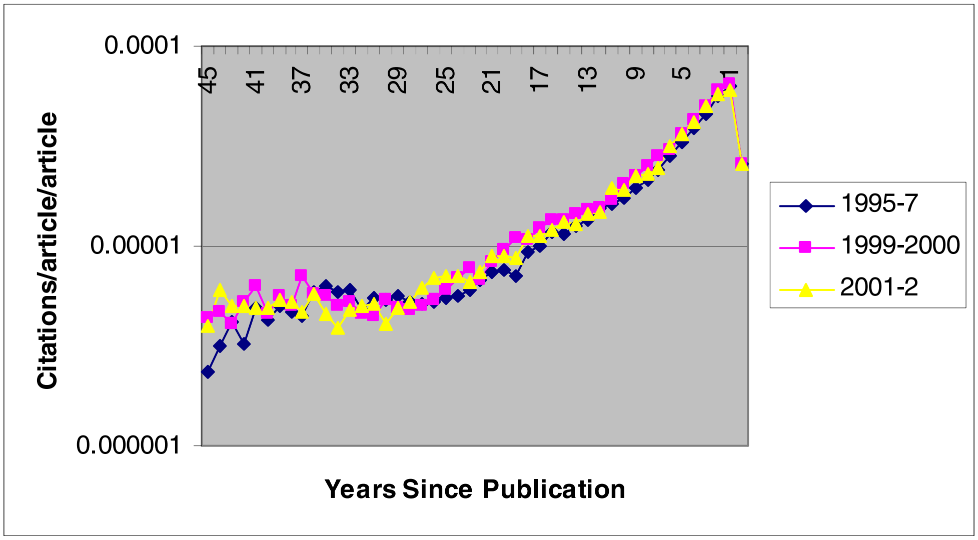

FIGURE 4. Differential doubly normalized citation rates from *Physical Review Letters* articles to all sections of the *Physical Review*, for the intervals 1999-2000 and 2001-2002 (i.e., subtracting the rate for 1995-7). (Data courtesy of the American Physical Society)