

***Revitalizing Older Published Literature: Preliminary Lessons from the Use of JSTOR***  
**Kevin M. Guthrie**  
**March 23, 2000**

Perhaps it would be best to begin this paper by stating explicitly what it is not. This paper does not present a scientific study. It does not purport to present evidence that will lead the reader to a carefully argued conclusion. Rather, it is an attempt to highlight some of the questions that usage of the JSTOR database is enabling us to ask and to begin to assess whether there are answers that will prove interesting or valuable to the scholarly community. At this stage, and with the relatively small amount of data and minimal degree of analysis that has been conducted, this report should be regarded as highly preliminary.<sup>1</sup>

JSTOR began as a research project sponsored by The Andrew W. Mellon Foundation at the University of Michigan. Its original objective was to test whether the digitized versions of older research journals might serve as a substitute for the paper versions, thereby offering libraries the possibility of long-term savings in shelving and archiving costs while simultaneously improving their usability. A pilot database was created that included the back runs of ten journals – five in history and five in economics – and access was made available at five liberal arts colleges and the University of Michigan.<sup>2</sup> By the summer of 1995, it was apparent that the concept held great promise, and JSTOR was established as an independent not-for-profit organization. JSTOR was founded to carry on the original objective stated above, but with the added charge that it develop an economic model that would allow it to become self-sustaining.

The JSTOR Phase I database now includes the backfiles<sup>3</sup> of 117 journal titles (see Appendix A) from 15 academic disciplines, a collection numbering nearly 5,000,000 pages. More than 650 academic institutions from 30 countries are now participants in this collaborative enterprise, with approximately 100 colleges and universities having had access to the database since early 1997. The amount of usage of the resource and

---

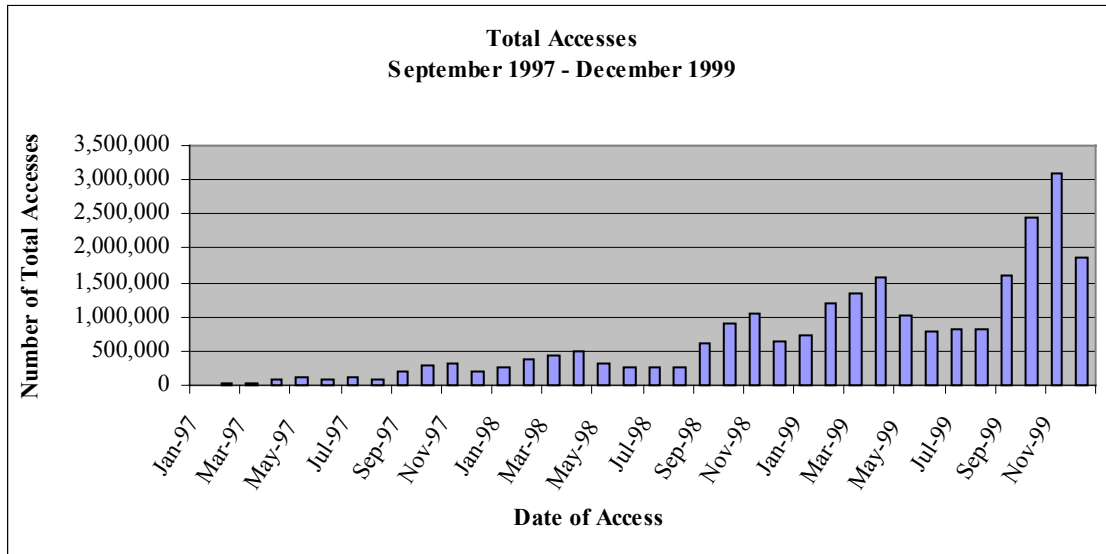
<sup>1</sup> Participating JSTOR libraries and publishers have requested that JSTOR exercise care when presenting and distributing JSTOR usage data. We therefore aggregate these data whenever possible and do not identify the usage at individual sites, for individual publishers, or for individual articles by title.

<sup>2</sup> The original test site libraries were: Bryn Mawr College, Swarthmore College, Haverford College, Denison University, Williams College in addition to the University of Michigan.

<sup>3</sup> JSTOR digitizes and makes accessible participating journals starting with the first volume published. In order to protect publishers' subscription revenue stream, JSTOR does not include current issues, but offers access up to a "moving wall" negotiated with each publisher.

its growth rate has been surprising. In 1999, over 1.4 million articles were printed from the JSTOR database, over 4 million searches were performed, and users accessed the database more than 17 million times.<sup>4</sup> Figure 1 illustrates the growth in the total number of accesses since the database was first made available.

**Figure 1**



When JSTOR was established, many people questioned the wisdom of converting journal backfiles. With comparatively little use of these materials in paper form, one could not help but wonder whether there would be sufficient interest in gaining access to the resource to warrant the substantial investments that would have to be made to create it. It is clear that it would not have been possible even to conceive of pursuing a project like JSTOR without the interest of the Mellon Foundation. Through its grant-making, the Foundation provided the financial resources necessary to establish the technological infrastructure required to create the database. Perhaps more importantly, however, the Mellon Foundation contributed staff time, most notably that of its President, William G. Bowen, to launch the enterprise.

<sup>4</sup> In assembling usage statistics, JSTOR counts significant accesses, not server hits. Accesses include actions such as viewing electronic table of contents or citation data, viewing an article, printing an article and executing a search.

The investments of the Mellon Foundation have made it possible for JSTOR to pursue and begin to fulfill its important not-for-profit mission, one component of which is to enhance the accessibility of little-used and inconvenient-to-retrieve journal literature. Another primary component of JSTOR's mission is to act as a trusted archive for the material under its care. This part of JSTOR's mission is reflected in the number of articles in the database that are not being heavily used today, but which may someday be a critical component of a new line of argument for an important paper or research article.

Early analysis of JSTOR's usage data allows us to begin to ask questions about how scholars and students use older literature in electronic form. Do scholars and students make use of the older articles? Are the materials being used more now than they were in paper format only? Can these data provide guidance about what material should be digitized? Does the usefulness of the older literature vary by academic discipline? These are some of the questions that we hope JSTOR will answer over the long run.

#### *Comparing JSTOR Use to the Usage of the Journals in Paper Format*

As part of the original JSTOR pilot project, an effort was made to collect circulation and usage information for the ten pilot journals. The hope was that the data would serve as a benchmark for comparison purposes. Unfortunately, it was not easy to collect reliable data. Since many of the journals were available in open stacks, it was not possible to obtain accurate circulation figures (although some circulation data were obtained from the University Reserves office at the University of Michigan Library). Instead of regular circulation data, two counting methods were employed to obtain information about use of these journals. First, slips of paper were placed in each journal volume with a request that a user mark when they had used the volume. Signs were also placed in the area of the journals to instruct users of the survey being conducted. Second, staff at the library pilot sites was instructed to check the shelves each business day for several months and make note of which volumes were not on the shelves. The volumes not on the shelves were counted as having been used.

Also, only the journal volumes housed on the main library shelves at the participating pilot libraries were included in this work. Usage of the paper volumes in faculty offices or in departmental libraries was not captured. Because of the lack of a controlled environment and the relatively narrow scope of this study, one must be careful about conclusions drawn when comparing these data to site license access to JSTOR at the institutions.

It does appear, however, that the electronic articles in JSTOR are being used much more frequently than they were used in the paper form. The paper usage data was collected over varying lengths of times at the five institutions that returned data, but a minimum of three months of information was collected. There were a total of 692 uses of the ten journals at the five test sites over the course of the entire survey period. Usage of the same journals in JSTOR at the same five sites for the months of September, October and November of 1999 yields a total of more than 7,696 article views. In addition, although there is presumably substantial overlap in articles viewed and those printed, 4,885 articles were printed -- a total of 12,581 views and prints during the three month time period. When compared to the 692 uses in the benchmarking survey, it would seem that the convenience of having electronic access is facilitating greatly increased use of the material.

Another way to assess whether usage of the older journals in electronic form is greater than in paper is by evaluating the growth in usage. It is rather unlikely that the usage of older articles in paper form was growing. That contrasts markedly with usage of JSTOR. Growth in the aggregate use of the JSTOR database has increased dramatically in the period since 1997 when it first became available. The table below shows the total accesses to the database by institution type.<sup>5</sup> Total accesses to all content in the database increased 4.4 times from 1997 to 1998 and 3 times from 1998 to 1999.

---

<sup>5</sup> JSTOR uses the Carnegie Classes of U.S. Institutions of Higher Education to place colleges and universities into one of five classes ranging from Very Large to Very Small. For a description of our methodology, see <http://www.jstor.org/about/class.html>.

**Table 1**

JSTOR Class	Accesses 1997	Accesses 1998	1997 – 1998 Growth Factor	Accesses 1999	1998 – 1999 Growth Factor
Very Large	817,893	3,291,648	4.0	8,550,945	2.6
Large	160,700	785,224	4.9	2,766,100	3.5
Medium	110,254	637,950	5.8	2,468,666	3.9
Small	110,312	490,854	4.4	1,323,894	2.7
Very Small	43,754	207,170	4.7	73,823	3.4
Totals	1,242,913	5,412,846	4.4	15,814,475	2.9

Because some of the growth in aggregate usage of JSTOR is a result of new institutions signing up for the database during this time period, we have compiled usage figures at institutions that had JSTOR installed prior to April 1, 1997. Aggregate accesses at these institutions increased by a factor of 3.4 times from 1997 to 1998 and by a factor of 2.5 times from 1998 to 1999. The cumulative growth of usage over the three-year time period at existing sites is 740%!

As one contemplates this impressive growth in JSTOR usage, it is perhaps valuable to note that JSTOR is available “for free” to end users. Libraries have paid participation site license fees that allow authorized users (faculty, staff, and students) to make unlimited use of the resource. For the most part, authentication is handled by IP address, thereby making the authentication process virtually invisible. This unfettered access contributes to the rapid growth in use of the resource; it is consistent with the kind of growth one is seeing in other resources available on the World Wide Web. This picture might be very different indeed if JSTOR were charging either users or libraries based on usage.

*The Interdisciplinary Appeal of JSTOR*

An additional variable that is likely to be a contributing factor to the increasing use of JSTOR is the addition of new content. During the past three years JSTOR has been digitizing new journals and making them available to participating institutions. Content in new academic disciplines introduces new scholars and students to the resource. Additional content in existing fields broadens the appeal of the resource within that discipline.

As the resource has grown, it is evident that the cross-title and interdisciplinary appeal of the resource has grown as well. Pulling from the search logs of a recent week of JSTOR use reveals that approximately 68,000 searches were conducted. Of these, just under 62,000 (90%) specified more than one title.<sup>6</sup> Because JSTOR offers the option to search by cluster (pre-defined discipline-specific collections), it is convenient for users to search across journals in a single discipline. Approximately 58,000 searches specified clusters. Of those cluster searches, 69% specified more than one cluster. This is quite significant because the JSTOR interface does not offer an option to select all clusters. The ability to search across disciplines is important to users.

#### *Nature and Distribution of Use*

There are a total of 831,087 articles in the JSTOR database. Our use of the term “article” may be a bit misleading in that it refers to all items that are indexed as an item for retrieval. Full-length articles are a sub-set of this total, of which there are presently 356,978. Other “articles” are items like book reviews, letters to the editor, membership lists, and the like.

The distribution of the use of JSTOR is interesting because it speaks to the extent to which JSTOR functions as an archive. Many libraries, particularly research and academic libraries, have a mission to collect not only that material that is likely to be used today, but also to collect and care for that information which may be valuable in the future. JSTOR has surprised us in the extent and degree that it has been used, but there is something to be learned also from what has not been used.

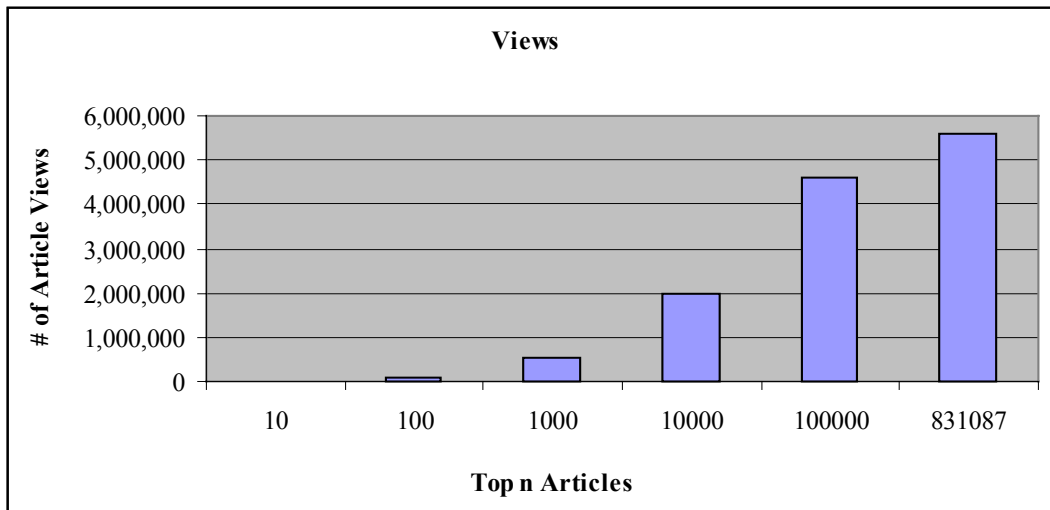
After three years, 430,429 different articles have been viewed, representing 51.8% of all articles in the database. (Many of these articles have been viewed multiple times; the figure above relates to whether the article has *ever* been viewed.) Obviously, the complement to this statement is that approximately half the

---

<sup>6</sup> JSTOR’s search interface requires the user to choose the titles or clusters he wants to search. There is no default list of journals or clusters that is automatically selected.

articles in the database have not yet been seen. 248,683 articles have been printed, representing 29.9% of all articles.

**Figure 2**



Consistent with the importance of the interdisciplinary value of the JSTOR database, one finds that usage of JSTOR is distributed rather widely across all articles. Figure 2 presents the number of article views accounted for by the top n articles. For example, the top 10 viewed articles were viewed 18,149 times, or 0.3% of the total number of times articles have been viewed in the database. The top 100 articles viewed represent 112,072 (2.0%) of the total. And the top 10,000 most viewed articles were viewed 1,987,982 times, or 35.5% of the total article views.

### *Selection Criteria*

Since it is generally accepted that it will not be possible to digitize all journals that have ever been published, an important question is how to select the retrospective content to be made available electronically. In JSTOR a variety of factors are taken into consideration in the selection process, including

surveys of faculty and library professionals in the field in question, library subscription levels, citation impact factor measures, and length of the run, among other things.

Looking at JSTOR usage at the article level, it is evident that citations should not be used as the sole factor in determining what content should be digitized. Table 2 displays the top ten articles in terms of JSTOR use since 1997 (through March 20, 2000) for three Economics titles. The number of citations to each article in the period from 1997 to 1999 is displayed,<sup>7</sup> as are the average number of citations to each article for the period from 1972 through 1999.

**Table 2**  
**JSTOR Usage – Economics Cluster**

<b>Journal Title</b>	<b>Number of Times Cited</b>	<b>Average cites/year</b>	<b>JSTOR Views</b>	<b>Year of Publication</b>
<i>American Economic Review</i>				
Article 1	79	24.1	1,670	1968
Article 2	77	15.7	1,232	1945
Article 3	181	35.9	1,316	1981
<b>Quarterly Journal of Economics</b>				
Article 1	175	32.4	2,426	1970
Article 2	104	26.6	2,400	1992
Article 3	216	50.9	1,583	1991
<i>Journal of Political Economy</i>				
Article 1	4	0.5	1,815	1973
Article 2	8	21.1	1,480	1990
Article 3	93	17.2	1,258	1983

What this table illustrates is that citations do not provide anything like a complete picture of the potential usefulness of a journal article. The most notable example of this point is the number one article for the *Journal of Political Economy*. Even though this 1973 article has rarely been cited (4 times between 1997 and 1999) and only an average of .5 times per year between 1972 and 1999, it has emerged as the most often used article from that journal. This article has been viewed 1,895 times and printed 1,402 times

---

<sup>7</sup> Citation data were determined using the Dialog service to access ISI Social Science Citations. Individual articles were located, and citations to these articles were analyzed by year of publication. Data were determined for 1997-1999 inclusive by simple addition. Average number of cites per year was figured beginning in the year of publication or 1972, whichever is later, through the end of 1999, then dividing that total number by the number of years since publication or 1972, whichever is later.



during the period that it has been accessible in JSTOR. What this example reveals is not only that citation data may not be the most useful measure for determining what should be digitized, but also that citations focus on what might be called the “reference” or “documentation” value of an article, not its usefulness defined more broadly. Articles with four citations may end up, for a variety of reasons, being the most used. Or, alternatively, highly cited articles may not be used very often at all. This is a factor to keep in mind when selecting content for digitization initiatives.

*Age of Useful Articles*

Table 3 shows calculated summary data for the most frequently used articles in each of the 15 JSTOR collections. The purpose of this assessment of our usage was to take an initial snapshot of the relative value of older literature in each of our JSTOR fields. The chart was assembled by first collecting the number of Views and Prints from the JSTOR database, ranking the articles in order of usage, and highlighting the top ten most used articles. When put alongside the publication date, it was apparent that some older articles have truly lasting value. In most of our major fields, older articles were well-represented among the “top ten.”

**Table 3**

	<b>Number of Titles</b>	<b>Num. Of Views from Top 10</b>	<b>Share of Top 10 Views</b>	<b>Avg. First Year of Publication</b>	<b>Avg. Most Recent JSTOR Year</b>	<b>Avg. Age of Top Ten Articles</b>
African American Studies	7	16,637	4%	1959	1996	3
Anthropology	6	12,301	3%	1954	1994	4
Asian Studies	4	5,433	1%	1936	1994	11
Ecology	6	19,293	5%	1943	1996	11
Economics	13	87,711	22%	1936	1994	13
Education	4	13,153	3%	1946	1995	11
Finance	5	13,201	3%	1958	1995	10
History	15	58,365	15%	1934	1995	12
Literature	11	23,992	6%	1946	1995	7
Mathematics	11	7,344	2%	1932	1994	32
Philosophy	10	16,538	4%	1931	1994	16
Political Science	9	52,201	13%	1933	1995	8
Population/Demography	8	15,808	4%	1965	1995	5
Sociology	9	41,387	11%	1945	1994	6
Statistics	11	8,480	2%	1936	1994	9

Again, to use the field of economics as an example, a surprising number of older articles have emerged as the most heavily used. The average age of the articles in the top ten most printed and viewed articles in the economics cluster is 13 years. This is rather surprising.

An even more dramatic example is Mathematics, where the average age of the most used articles in the field thus far is 32 years! This result is consistent with what mathematicians have told us about their field; that is, that older mathematics literature remains valuable. (Mathematicians are some of the most enthusiastic supporters of JSTOR and regularly urge us to include more mathematics titles). However, it is important to point out that usage of the mathematics cluster in JSTOR has lagged behind other fields. With the long runs of its 11 journals, as a cluster mathematics has the highest number of pages in JSTOR, and yet usage of the mathematics cluster represents just 3.3.% of total usage. One reason for making this point here is that there simply is not enough data to make too much of the average length of the article in mathematics. With a small number of total accesses for the field, the actions of a few people can sway the data significantly. As mentioned earlier, one has to be careful about drawing conclusions from the data.

Nevertheless, the apparent contradiction between the qualitative value of JSTOR to mathematicians and the usage of the mathematics journals in JSTOR dramatically illustrates an extremely important point. One must define clearly what one means by “value”. Usage does not necessarily equate to value in the research sense. Older articles may be absolutely vital to the continuation of high quality scholarship and research in the field, but that may not lead to extensive use. Increasingly, one hears that libraries are planning to use electronic usage data to help make subscription decisions. If relied upon exclusively, this could prove to be a very dangerous tool, making it more difficult for lesser used but valuable research journals to survive. Other measures, like citation data, need to be incorporated as well. The nature of these data will also change with the availability of electronic resources. One wonders, for example, if the number of citations to older articles in JSTOR will increase as the older articles become more conveniently accessible. This possibility is worth monitoring, but with the understanding that it will take years before changes in scholars’ behavior will manifest itself in the citation data. Understanding the nature of a field and the way

that research materials are used in the field is essential before making selection and cancellation decisions. It is our hope that, over the long run, JSTOR can contribute to this kind of understanding.

### *Conclusion*

This paper provides a brief overview of preliminary information emerging from JSTOR usage data. As JSTOR usage increases, interesting questions about the way that retrospective electronic collections are used can be asked and investigated. Although it is still too early to draw conclusions (much more data will need to be collected), evidence points to preliminary hypotheses in five primary areas.

1. Electronic access seems to have increased the use of older materials at JSTOR participating sites.
2. The interdisciplinary nature of JSTOR seems to be valued by researchers and students.
3. Citation data alone is not a good predictor of electronic usage, and probably should not be used to make digitization decisions for retrospective content.
4. Older literature seems to remain valuable in many fields.
5. Care should be taken to insure that there is clear understanding of the definition of “value” for research articles. Judging by the nature of the articles that are most used, it is not always those that push forward the research and intellectual understanding of an academic discipline, it may very well be “popular” articles used in larger classes. “Value” needs to be clearly defined as libraries consider acquisition and cancellation decisions for electronic content.

## Appendix A

### *JSTOR Phase I Journal Titles*

African American Review	Journal of the Royal Statistical Society, Series C: Applied Statistics
American Historical Review	Journal of the Royal Statistical Society, Series D: The Statistician
American Journal of International Law	Mathematics of Computation
American Journal of Mathematics	Mind
American Journal of Political Science	MLN
American Journal of Sociology	Monumenta Nipponica
American Literature	Nineteenth-Century Literature
American Mathematical Monthly	Nous
American Political Science Review	Pacific Affairs
American Quarterly	Philosophical Perspectives
American Sociological Review	Philosophical Quarterly
Annals of Applied Probability	Philosophical Review
Annals of Applied Probability	Philosophy and Phenomenological Research
Annals of Mathematics	Philosophy and Public Affairs
Annals of Probability	Political Science Quarterly
Annals of Statistics	Population and Development Review
Annual Review of Anthropology	Population Index
Annual Review of Ecology and Systematics	Population Studies
Annual Review of Sociology	Population: An English Selection
Anthropology Today	Proceedings of the American Mathematical Society
Biometrika	Proceedings of the American Political Science Association
Callaloo	Proceedings of the Royal Anthropological Institute
Contemporary Sociology: a Journal of Reviews	Public Opinion Quarterly
Current Anthropology	Renaissance Quarterly
Demography	Representations
Ecological Applications	Review of Economic Studies
Ecological Monographs	Review of Economics and Statistics
Ecology	Review of Financial Studies
Econometrica	Reviews in American History
Economic Journal	Shakespeare Quarterly
Eighteenth-Century Studies	SIAM Journal on Applied Mathematics
ELH	SIAM Journal on Numerical Analysis
Ethics	SIAM Review
Family Planning Perspectives	Social Psychology Quarterly
Harvard Journal of Asiatic Studies	Sociology of Education
International Family Planning Perspectives	Speculum
International Organization	Statistical Science
Journal of American History	Studies in Family Planning
Journal of Animal Ecology	Studies in the Renaissance
Journal of Applied Econometrics	The American Economic Review
Journal of Asian Studies	The China Journal
Journal of Black Studies	The Journal of Blacks in Higher Education
Journal of Ecology	The Journal of Business
Journal of Economic History	The Journal of Economic Literature
Journal of Financial and Quantitative Analysis	The Journal of Economic Perspectives
Journal of Health and Social Behavior	The Journal of Finance
Journal of Higher Education	The Journal of Military History
Journal of Industrial Economics	The Journal of Modern History
Journal of Money, Credit and Banking	The Journal of Negro Education
Journal of Negro History	The Journal of Political Economy
Journal of Philosophy	The Journal of Politics
Journal of Symbolic Logic	The Journal of Southern History
Journal of the American Mathematical Society	The Journal of the Royal Anthropological Institute/Man
Journal of the American Statistical Association	The Quarterly Journal of Economics
Journal of the History of Ideas	Transactions of the American Mathematical Society
Journal of the RAI of Great Britain and Ireland	Transition
Journal of the Royal Statistical Society, Series A: Statistics in Society	William and Mary Quarterly
Journal of the Royal Statistical Society, Series B: Methodological	World Politics
	Yale French Studies