

A Comparison of the Costs of Providing Electronic vs. Printed  
Copies of Journal Articles to University of California Users:  
A Conceptual Design Framework for Data Collection and Analysis

Michael D. Cooper

School of Information Management and Systems  
University of California  
Berkeley, California 94720-4600

June 21, 2000

## Introduction

The availability of scholarly journals in electronic form raises a fundamental question: should a library continue to provide and maintain a printed copy of a journal for its users if an electronic copy is available? As electronic journals become pervasive and their use increases, should each individual Campus give up its paper subscription and its backfiles of bound volumes, and instead rely on the electronic form and on one or two bound sets of the journal stored at a Regional storage facility?

This paper presents a methodological framework for examining the problem employing usage and cost analysis. Other aspects of the problem beside cost analysis must be considered: behavioral factors and institutional factors. Behavioral factors include how individual users and librarians may change their perceptions and patterns of use of the collection as a result of the availability of electronic materials. Institutional factors include how the availability of electronic materials changes institutional operations, staffing, client relationships, ongoing budgeting, and capital decisions.

The cost approach develops two general scenarios. In the first, a library provides traditional access to its printed materials, be they on campus or in a storage facility. In the second, the library provides electronic access to some journals, and for those journals for which electronic access is provided, the paper copies are moved to a storage facility. Several other issues are evaluated. The cost trade-off of varying the number of paper copies that are archived is considered. The cost of moving the paper copies back to the source library is considered. And the cost of providing an archived paper copy of a journal or journal article to a user is calculated.

The approach to analyzing these problems is to derive unit costs. They include the cost of acquiring, processing, and storing printed copies and electronic copies of journals;

the cost of storing archival copies of the journals at regional facilities; and the cost of using both printed and electronic copies from both the library and user's standpoint.

## Experimental Design

The basic question to be resolved in this analysis is how the cost of providing access to electronic versus print journals changes for the individual, the library, and the library system as a whole. From the user standpoint, the tradeoff is availability, preference, time, technology, and cost. In what form is the material available, in what form would the user like to access it, how much time will it take to access it in each form, what will be the cost to the user of accessing it in each form, and does the user have the technological skills and equipment to access the item electronically?

From the library's standpoint, the tradeoff is in terms of selection cost, acquisition cost, circulation cost, and storage cost. What is the cost of selecting paper versus electronic journals, how much does it cost to maintain a subscription to each form of journal, how much does it cost to provide circulation services to each one, and how much does it cost to store (or not store) paper versus electronic journals.

For a library system, other issues arise. What is the cost of providing storage space at a regional facility to store paper copies of journals that have been removed from a local library's shelves. And what is the cost saving to the local library of having removed those journals from their shelves? What is the cost to the library system of providing electronic journal access, versus having single subscriptions to paper copies of journals at each campus? What is the reduction in capital costs to the library system (e.g. construction costs) of reducing the storage requirements for paper copies of journals at local libraries and regional libraries?

The answers to these questions can be obtained by accumulating usage data and unit cost data on user and library operations. This data can then be used to perform simulations and sensitivity analyses to arrive at best estimates of the answers. For example, one aspect of the problem is the way in which journal articles are retrieved, and then delivered to the user. For the user, the alternatives include retrieving a printed journal at the library, obtaining a printed journal from a regional storage facility, obtaining a fax or photocopy of the printed journal from the regional storage facility, retrieving a digital image of the printed journal article from the library or storage facility, and retrieving (and possibly printing) an electronic copy of the journal article from a database. Given the way in which cost data will be collected, a cost comparison of these alternatives is possible.

Another basic tradeoff takes the library perspective. A library can provide materials in printed or digital form. It can continue to buy, or stop buying, printed versions of current issues of journals. It can keep, discard, or send to a storage facility, backfiles of printed

journals. The cost of each of these alternatives can be estimated. In addition, a sensitivity analysis can be conducted to see how the costs will vary depending on the number and type of issues and volumes kept or retained in each type of storage facility.

The methodology outlined below allows many specific costs to be ascertained. The following is a sampling to provide perspective:

1. The cost per use of a printed journal article either from the local library or from a storage facility.
2. The cost per use of an electronic journal article.
3. The cost of moving a series of volumes to and from a storage facility.
4. The cost of maintaining electronic access to journals, and, in addition, printed copies of journals on each campus.
5. The cost of maintaining electronic access to journals, and also one or more storage facilities containing printed copies of the journals.

The current pricing structure of electronic journals bundles printed copy costs with electronic copy costs. Purchasers have to take acquire both forms. Should unbundling occur, the unit cost data accumulated here will allow for simulation of alternative scenarios. One scenario could be for one or more paper copies of the journal to be acquired by the University and placed in a storage facility. Meanwhile the electronic copy would be available over the Internet to all members of the University community.

### Processing and Access to Printed Journals

As a precursor to developing a costing methodology, it is important to understand the steps involved in providing user access to both printed and electronic journals.

The decision to acquire a new journal title for the library is made on the basis of the characteristics of the journal, its relevance to the institution's academic program, and the availability of financial resources to support the subscription. Once that decision is made, an order is entered into a computerized serials system and is transmitted to a vendor or jobber. Then a cataloging entry is made for the title, and, hopefully, the serial issues begin arriving. If issues of the serial title do not arrive when expected, a claiming process must be initiated and computer records updated to reflect it.

At some point, payment is authorized for the title subscription and funds transferred to the vendor. At each renewal date, payment is authorized again for the serial, and funds transferred. Each serial issue is checked-in, labeled, and moved to a periodicals room, stacks, or perhaps routed to interested parties. Periodically, a set of individual issues of

the journal are picked from the shelves and sent, with binding instructions, to a bindery. After binding, the volume is returned to the stacks, and cataloging records are updated to reflect the existence of a bound volume rather than individual issues of the title.

In some cases, bound volumes may be moved to a storage facility rather than remain in the main stacks of the library. If this happens, the cataloging record for the library that owns the material is updated to reflect the item's current location. If a title is moved to a storage facility, it is checked-out of the main library, transported to the storage facility, checked-in at the storage facility, cataloged, assigned an accession number, and then shelved.

When a user wishes to review the contents of an issue of a printed journal, multiple steps are involved. They vary depending on the user's approach and the location of the material. In a typical case, the user begins the process by a search of the online catalog to find the library holding the item, its call number, and whether she/he can expect the item to be on the shelf, non-circulating, or checked-out. The user then proceeds to the library, and either enters the stacks or has the item paged for use. Once the user has the item in hand she/he consults the issue and either checks it out or returns it for reshelving.

If the item is to be checked-out, a circulation process begins in which the user presents a library card and the item itself. The card and the bar code in the item are scanned, the item is marked with a due date, and the material and card returned to the user. A computer system maintains circulation records which indicate that the user has checked the item out for a designated period of time. The circulation system allows the user to extend the due date of the item if desired (renewal), allows the library to recall the item if it is needed for another borrower, and perform administrative functions such as assessing fines on the user for not returning the item or returning it beyond its due date. When the item is returned, the library checks the item in, and the circulation records are updated. Then the item is reshelved in the stacks.

If the item is in a storage facility, the user can initiate its recall, or initiate the delivery of a photocopy of the item from storage. If the physical item is to be delivered, the user makes the request which is transmitted by the library to the storage facility. The storage facility receives the request, consults its computer system to find the item's physical location, retrieves the item, checks the item out to the requesting library, and physically transports the item to the library. The receiving library checks the item in, the borrower retrieves the item, and checks it out from the requesting library. When the borrower returns the item the process is reversed.

If a photocopy of the item is to be delivered from the storage facility, the request is transmitted to the storage facility and a work order prepared. The volume is retrieved from the shelves, the article photocopied, and the volume returned to the shelves. The photocopy is then delivered to the requesting library or individual, and records of the activity updated.

## Processing and Access to Electronic Journals

The University of California's California Digital Library and individual University of California campuses offer their users digital images of journal articles delivered to their computer screen through the Internet. The process by which this material is delivered is similar in some respects to that outlined above and also radically different.

The similarity between providing electronic and paper copies of journal articles is in the administrative processes. An electronic journal is developed by a publisher or professional society and is made available on a contractual basis. A staff member or faculty member makes an initial recommendation about the acquisition of the electronic journal. Then staff and lawyers from the vendor and the library are involved in negotiating an agreement and a price. The agreement can take considerable time to complete and is nothing like paying a subscription fee for a printed copy of the journal. Periodically, the agreement comes up for renewal, and its terms and conditions must be renegotiated, usually after lengthy discussions.

After the electronic journal has been acquired, it must be cataloged, and information about its holdings entered into a serials system. Since there is likely to be overlap between the electronic and paper copy of the journal, the holdings statements can be complex and difficult to maintain. As more issues are added to the electronic holdings, the holdings statements must be updated, just as for paper issues. If the way in which the journal is accessed is changed, such as through a different URL, that information must be changed.

There are multiple ways in which the electronic form of the journal can be made available to the library. One is for the library to take the digital form of the journal and store it on its own computer system. Another is for the library's users to access the journal through the vendors computers. Either way, a certain infrastructure is necessary to support access. A computer hardware, software, and telecommunications infrastructure must be in place which allows access to the electronic journal. The hardware configuration must be sufficient to support thousands of user accesses at any one time, the software must be sophisticated enough to provide good data management tools and a good user interface to the journals, and the telecommunications structure must be robust enough to allow a high volume of message traffic. The staff necessary to support this type of enterprise is not insignificant. Thousands of person-hours are involved in developing software to make such systems operational, and staff is needed to maintain the systems as well. Needless to say, there are significant costs of developing and running such a facility.

User access to an electronic journal is through a Web browser running on a personal computer or workstation. The user employs either his/her own computer or one provided by the University. In either case, the computer is connected through dedicated or dial-up telecommunications lines to the University's computer system. Thus the user or the

University must provide a personal computer, and its software, telecommunications connectivity, and perhaps a printer for this access to be viable.

The user can gain access to the full text of the article in many ways, all through the Web browser. One way is to search the the library Web site for the name of the electronic journal. Once the journal title has been found, the Web site indicates the scope of the electronic and hard-copy holdings available, and the user then decides whether the electronic form contains the required article. The user also may be led to the electronic text through an abstracting and indexing database's citation, or by a reference link from another electronic publication. The user can view the article online or can download it to his/her local machine for viewing and/or printing.

In order for the user to have access to the Web site that contains the electronic journal, some form of authentication must take place. If there were no authentication, any user could have access to any electronic journal and publishers would loose control of their intellectual property. Thus, authentication takes place, and authentication server computers and/or administrative procedures must be maintained to provide this control.

#### A Cost Comparison Approach

The goal of a general analysis of the problem is to develop a cost comparison between providing users paper and electronic journals. This comparison should include the costs of archiving the printed version of the journal to a regional storage site, the cost of retrieving the printed volume or a photocopy of a printed article from the storage site, and the costs of storing more than one copy of the journal at more than one storage site.

In the experiment to be conducted here, a more limited strategy is proposed. A set of journals will be identified for which both electronic and paper copies are available. For those journals, the paper copies will be systematically relocated to a regional storage facility, and the electronic copy relied upon to serve current demand. This particular experiment reduces the need for a very general cost analysis to one more limited in scope.

A major problem in the analysis is deciding which cost elements should be included in the analysis. The approach taken here is to assume that there are two categories of costs considered: direct usage-related costs and those capital costs related to the storage of paper copies of materials. Other capital costs are excluded, such as the computer system on which the electronic journal is stored, the computerized circulation system, and the information retrieval software used to locate electronic journals. While some of the capital costs that are omitted can be apportioned between their use for library and non-library- related functions, others can only be divided with difficulty. A good example is telecommunications. It would be very difficult to apportion telecommunications traffic costs between library-specific and University-wide activities.

The alternative is to examine costs directly related to the activities in question. These include the cost of the item itself, along with the cost of its selection, technical processing, and cataloging by the library. Other costs include the cost of the use of the item, including circulation costs for paper journals and licensing costs and per-use costs (if any) for electronic journals.

## Cost Categories

The nature of the methodology proposed here requires accumulating costs for certain operations as discrete activities. For example, in order to conduct the analysis one needs to know the cost of both storing and retrieving paper copies of journals that reside in a storage facility. These discrete cost components can then be combined in a number of ways to derive the economic conclusions. Knowing the cost of retrieving an item from a regional storage facility allows that cost to be used both in calculating the cost of use of an issue and the cost of returning the bound volume to the owning library at the end of the experimental period.

The cost categories that should be accumulated include: library selection costs, library processing costs, library and user circulation costs, remote storage facility storage costs, and remote storage facility circulation and storage (capital) costs.

Library selection costs include the costs of negotiating the purchase of the item and the cost of the item itself. Library processing costs include the cost of cataloging, receiving, marking, and storing the item. User circulation costs for paper journals include the cost of locating the material, retrieving it, checking it out, copying it, and returning it to the library. Library circulation costs for paper journals include checking the item out, checking it in, and reshelving it. The costs of storing materials in a remote facility include selecting the item to be transferred to storage, physical transportation to storage, holdings record updating, receiving at the storage facility, and shelving at the storage facility. User circulation from remote storage includes requesting the item, picking it up, and returning it. Costs at the storage facility for circulation include receiving the request, retrieving the item, checking it out to the requesting library, checking it in from the requesting library once it has been returned, and reshelving it. Costs at the library and storage facility for storing items include the unit cost of keeping a volume on the shelf of the facility. It is necessary to accumulate these costs in order to compute the trade-off between storing copies of the item at multiple libraries versus one or more storage facilities.

A general comparison of costs would involve collecting data in all the categories enumerated above. This would allow a full analysis of the overall cost of acquisition, processing, circulation and storage. But this may not be necessary in the the experimental situation if the results do not need to be widely generalized beyond the University of California system. In the limited case, one need only measure the costs that

will change between in the experimental situation. Thus selection costs and processing costs would be omitted because they do not change for either printed or electronic journals during the experiment. However, circulation costs, costs to process materials at the storage facility, and capital costs to store material at the storage facility would have to be included.

The following table summarizes the need for each category of cost between the base current situation and the experimental situation. The entry 'N/A' in the table indicates the cost is the same for all categories and can be omitted, if desired, in the analysis. The entry 'Y' indicates the cost category should be included, and the entry 'Y\*' indicates the cost category should be included and is likely to change between the base and experimental situations. This could occur either because of differences in the applicable unit costs or significant differences in the level of activity associated with that cost category owing to the experimental situation. The entry 'N' indicates the category should be omitted.

Cost Category	Base Cost		Experimental Cost	
	Paper Journals	Digital Journals	Paper Journals	Digital Journals
Selection costs	N/A	N/A	N/A	N/A
Processing costs	N/A	N/A	N/A	N/A
Circulation costs	Y	Y	Y*	Y*
Storage Facility Circulation costs	Y	N	Y*	N
Storage Facility Storage costs	N/A	N/A	Y	N
Capital Cost to store item	Y	N	Y*	N/A

If a complete cost analysis is conducted as opposed to the more limited one outlined above, data will need to be accumulated for each element within all cost categories. The table below lists these categories and cost elements. It indicates whether the cost element should be accumulated for paper journals, electronic journals, or both.

Cost element	Paper Journals	Electronic Journals
--------------	----------------	---------------------

Library Selection/Acquisition costs:

Decide whether to acquire item	Y	Y
Negotiate contract	Y	Y
Write order/sign contract	Y	Y
Renew subscription	Y	Y
Claim issue	Y	N

Library Processing costs:

Catalog title or revise existing cataloging	Y	Y
Receive issue	Y	N
Verify that electronic issue is accessible	N	Y
Update holdings information	Y	Y
Mark physical item with property stamp and call number	Y	N
Store issue on shelves	Y	N
Bind a series of issues	Y	N

User Circulation costs:

Search catalog for title	Y	Y
Retrieve volume from open stacks	Y	N
Check-out item	Y	N
Make photocopy or print of item if needed	Y	Y
Return item to library	Y	N

Library Circulation costs:

Retrieve volume from closed stacks	Y	N
Issue authentication to allow user electronic access	N	Y
Check-out item to user	Y	N
Record electronic journal item usage	N	Y
Check-in item	Y	N
Reshelve item	Y	N

User Remote Storage Facility Circulation costs:

Request item be retrieved from storage facility	Y	N
Pick up item from library	Y	N
Check-out item	Y	N
Make photocopy or print of item if needed	Y	Y
Return item to library	Y	N

Remote Storage Facility Circulation costs:

Transmit request for item to storage facility (library)	Y	N
Receive request and retrieve item from storage (storage facility)	Y	N
Photocopy or fax item (storage facility)	Y	N
Deliver item to requesting library (storage facility)	Y	N
Receive requested item (library)	Y	N
Check-out item to user (library)	Y	N
Deliver item to storage facility (library)	Y	N
Check-in item (storage facility)	Y	N

Reshelve item (storage facility)	Y	N
Store Materials in Remote Facility:		
Select item to be transferred to storage (library)	Y	N
Pick physical items to be transferred (library)	Y	N
Update holdings information in catalog (library)	Y	N
Transport items to storage facility (library)	Y	N
Check-in items at storage facility (storage facility)	Y	N
Shelve items at storage facility (storage facility)	Y	N
Library Storage costs:		
Cost per volume per year for storage of material	Y	N
Storage Facility Storage costs:		
Cost per volume per year for storage of material	Y	N

Not all the items in this table occur with the same frequency or at all. For example, a user may or may not photocopy an item. When usage data and costs are collected for this element, information must also be accumulated on the proportion of time that the event occurs.

The process of cost estimations for each of the elements should be based on unit cost analysis. Suppose, for example, that the goal is to derive the unit cost of photocopying an article at a regional storage facility. First measure the time required to perform this operation for some number of requests. Then obtain the average direct cost per minute of the photocopy operator's salary. The cost per minute would be multiplied by the number of minutes to perform one photocopy operation. This would result in the labor cost of making the photocopy. One of the basic assumptions in the study is that only direct costs are considered in the analysis. Thus the amortized cost of the photocopy machine is not included. However, photocopy supplies might need to be included depending on the level of detail required in the analysis.

A similar type of analysis needs to be conducted for each of the cost elements.

## Monitoring Electronic Journal Usage

Most of the data needed to perform the cost analysis can be derived by analysts measuring employee and user performance and using published salary and construction data to derive costs. However, this is not the case in measuring the use of electronic journals. It is necessary to enlist the cooperation of the electronic journal vendor in obtaining data on journal usage. This may require making contractual arrangements with the vendor to implement monitoring software.

The vendor should supply the University with one computer-generated log record for each article accessed by a University-authorized user for the journals included in the study. The fields in this log record should include the following:

- Unique ID number of this record
- Date-stamp when record was written
- Vendor ID number
- User authorization code issued by proxy server to access database
- User IP address
- Database ID number
- ID number of article accessed
- Date-stamp of time when article first accessed
- Date-stamp of time when access to article completed (or timeout time)

A date-stamp contains the following elements: four digit year value, month, day, hour, minute, hundredth of second. Database ID numbers should be unique across all databases for which data is accumulated.

## Experimental Methodology

A comprehensive cost analysis of the trade-off of the use of paper and electronic journals for the entire University of California system is out of the question. This would be too expensive. Instead, the goal is to develop a representative sampling methodology which will yield reasonably accurate results given limited time and resources.

Cost information needs to be accumulated for two categories of institutions: libraries and regional storage facilities. In addition cost information needs to be accumulated for two types of materials: paper copies of journals and electronic journals. Finally, there are costs associated with user activities, such as searching for materials and retrieving materials.

How many libraries should be sampled for their cost data? The pragmatic answer is to determine if there are any libraries that already have the data, see if there is any

methodological consistency in it, and use the collected data. If this fails, make a decision on which institutions to sample, such as one large and one small library system. There seems no need to sample more than two libraries.

Costs are needed from regional storage facilities. It appears that cost data already exists from the Southern Regional Library Facility. If this is the case, that data should be used.

The final experimental design issue that must be resolved is the number of electronic journals to be included in the sample. Approximately 20 vendors currently make available about 4800 electronic journals through the California Digital Library. The experimental design is to select a set of these journals for monitoring. For those journals selected, the paper copies of the overlapping issues would be moved to a storage facility or made otherwise unavailable to the users. Then all use would be funneled to the electronic journals.

In order to prevent bias in the analysis, the electronic journals must be selected to be representative of their use across discipline, across campuses, by the number of issues that are available digitally, by language, by publisher, by cost, and by use rate. The best method to derive the sample is to produce a list of the journals with with data about each of the strata described above. Given this information, a subjective decision of which to include can be made. A reasonable sample might be about 100 journal titles.