4 The PEAK Project: A Field Experiment in Pricing and Usage of a Digital Collection

Maria S. Bonn, Wendy P. Lougee, Jeffrey K. MacKie-Mason and Juan F. Riveros

Electronic access to scholarly journals is now an important and commonly accepted tool for researchers. The user community has become more familiar with the medium over time and has started to actively bid for alternative forms of access. Technological improvements in communication networks, paired with decreasing costs of hardware, create greater incentives for innovation. Consequently, although publishers and libraries face a number of challenges, they also have promising new opportunities.[1] Publishers are creating many new electronic-only journals on the Internet, while also developing and deploying electronic access to literature traditionally distributed on paper. They are modifying traditional pricing schemes and content bundles, and creating new schemes to take advantage of the characteristics of digital duplication and distribution.

From 1997 to 1999, researchers in economics at the University of Michigan worked in collaboration with the University of Michigan Library to design and run a project called Pricing Electronic Access to Knowledge (PEAK). This project was both a production service for electronic journal delivery and an opportunity for experimental pricing research that provided access to the more than 1,100 journals then published by Elsevier Science—journals that include much of the leading research in the physical, life and social sciences. The project provided an opportunity for universities and other research institutions to have electronic access to a large number of journals. This access provided fast and sophisticated searching, nearly instantaneous document delivery, and new possibilities for subscriptions. The University of Michigan Library Digital Library Production Service (DLPS) provided a host service consisting of roughly three and a half years of content (January 1995–June 1999) of the Elsevier Science scholarly journals. Participating institutions had access to this content for over 18 months, after which the project ended and access through our system ceased. Michigan provided Internet-based delivery to over 340,000 authorized users at twelve campuses and commercial research facilities across the U.S. On top of

1 See MacKie-Mason and Riveros (2000) for a discussion of the economics of electronic publishing.

this production system we implemented a field trial in electronic access pricing and usage.

Our primary experimental objective was to learn how additional value can be extracted from existing content by means of innovative electronic product offerings and pricing schemes. We sought to determine how users respond to different pricing schemes and to assess the additional value created from different product offerings. We also analyzed the impact of the different pricing schemes on producer revenues. To a limited extent, we think our results generalize to various business models, customer populations and information goods. Finally, we compared our empirical results with the current conclusions of the economic literature on bundling of information goods.

4.1 PEAK in context: Electronic journal publishing and the University of Michigan Library

The scholarly journal has a tradition of purpose and structure dating back several centuries, with little change. Despite the combined effects of price inflation and fluctuations of currency exchange that libraries weathered in the 1970's and 1980's, the basic construct of journals and subscriptions remained stable and, in fact, the journal has continued to flourish in a world of scholarly publishing that is increasingly global and conglomerate. In contrast to this tradition-laden history, the rapid change stimulated by information technologies in the 1990's was remarkable and unprecedented.

Early efforts to harness the potential of digital technology for journals focused primarily on distribution and access. A far more gradual and separate process of re-engineering editorial review and production processes emerged somewhat later. Major publishers undertook an array of projects with heightened activity evident at the dawn of the Web. Efforts such as Springer Verlag's Red Sage project and Elsevier Sciences' TULIP (The University LIcensing Program) initiative broke ground in testing the limits of Internet distribution and catalyzing the development of more robust access systems. TULIP involved nine institutions and addressed a broad set of issues, including both technical and behavioral concerns. The four-year project achieved significant progress, but failed to address issues of economics and pricing for the new electronic

Math6X9/2005/11/02:10:25 Page 47

media (Elsevier Science 1996).

In the aftermath of this early experimentation in electronic journal publishing, a number of inter-related issues emerged that stimulated interest in the economic questions surrounding journals and their electronic versions. Nearly every major publisher launched electronic publishing initiatives and, typically, tackled issues of price, product, and market in a manner that extrapolated from print practices. Early pricing models tightly coupled electronic and print subscriptions. Often electronic versions were available as a companion to the print version, at a surcharge of 15% or more. Almost simultaneously, the phenomenon of electronic preprint services emerged. These factors—plus a growing appetite for enhanced journal functionality—have contributed to the heightened interest surrounding pricing and product models for scholarly journals.

The University of Michigan was one of the institutional participants in TULIP, with a joint project team drawing from Engineering, the School of Information and Library Studies (now the School of Information), the Information Technology Division, and the University Library. Michigan was the first site to implement the 43 journals in materials science offered through TULIP and was also the first to move the service to the Web environment. TULIP's outcomes included a far better understanding of the distribution and access issues associated with electronic journals, but also underscored the inadequacy of an experiment offering too few journals to attract users on a regular basis.

The TULIP experience, coupled with an early history of standardized markup language (i.e., SGML) development in the 1980's, provided a unique environment for digital library development and contributed to Michigan's selection as a technology service provider for the Mellon Foundation-funded JSTOR project (Guthrie 1997, this volume). The unique organizational collaboration begun with TULIP was expanded in 1993 and institutionalized in a campus-wide digital library program that today encompasses a full production service and development capability (Lougee 1998). Within this new program the TULIP legacy was pursued with an eye toward better understanding the value, price, and product for electronic journals.

In 1996, an agreement was reached with Elsevier Science to launch PEAK in an attempt to address issues left outstanding in the TULIP process. Through PEAK, Michigan hoped to gain a better understand-

ing of large-scale management of electronic journals through the development of production systems and processes to accommodate the large body of content published by Elsevier Science. While this goal was important, PEAK also provided a large-scale testbed in which to explore issues of pricing and product design for electronic journals.

4.2 Issues guiding the design of PEAK

Information goods such as electronic journals have two defining characteristics. The first and most important is low marginal (incremental) cost. Once the content is transformed into a digital format, the information can be repackaged and distributed at almost zero cost. Nevertheless, the second feature is that information goods often involve high fixed ("first copy") costs of production. A production facility and distribution server must be in place in order to take advantage of the low costs of distribution. For a typical scholarly journal, most of the cost to be recovered by the producer is fixed.[2] The same is true for both publisher and distributor in an electronic access environment. With the cost of electronic "printing and postage" essentially zero, nearly all of the distribution cost consists of system costs for hardware, administration, database creation and maintenance—all costs that must be incurred whether there are two or two million users. Our experience with PEAK bears this out: the only significant variable operating cost was the service of the user support team who answered questions from individual users—a small part of the total cost of providing the PEAK service.

Electronic access offers new opportunities to create and extract value from scholarly literature. This additional value can benefit readers, libraries, distributors and publishers. For distributors and publishers, additional value can help to recover the high fixed costs. Increased value can be *created* through the production of new products and services (such as early notification services and bibliographic hyperlinking). Additional value that already exists in current content can also be delivered to users and, in part, extracted by publishers through new product bundling and nonlinear pricing schemes that become possible with electronic distribu-

2 Odlyzko (1995) estimates that it costs between $900–$8700 to produce a single math article. 70% of the cost is editorial and production, 30% is reproduction and distribution.

tion. For example, journal content can be unbundled and then rebundled in many different ways. Bundling enables the generation of additional value from existing content by targeting a variety of product packages for customers who value the existing content differently. For example, most four-year colleges subscribe to only a small fraction of Elsevier titles. With innovative electronic bundling options, this and other less-served populations may be able to access additional content.

4.3 System design and implementation

Our research team worked with the University Library to design an online system and market this system to a variety of information clients. We primarily targeted libraries, focusing on academic and corporate libraries. Contacts were made with institutions expressing interest and institutions already invested in digital library activity. Over thirty institutions were contacted as potential participants, of which twelve agreed to join the effort. Decisions not to participate were frequently driven by budget limitations, or by the absence of pricing options of interest to the institution.[3] The resulting mix of institutions were diverse in size and information technology infrastructure, as well as in organizational mission. PEAK participants were the University of Michigan, the University of Minnesota, Indiana University, Texas A & M, Lehigh University, Michigan Technological University, Vanderbilt, Drexel, Philadelphia College of Osteopathic Medicine, University of the Sciences in Philadelphia, Dow Chemical, and Warner-Lambert (now part of Pfizer Pharmaceuticals).

The PEAK system provided full-text search of and retrieval from the entire body of Elsevier content for the duration of the experiment, including some content from earlier years that Elsevier provided to help establish a critical mass of content. Several search and browse options were available to users, including mechanisms that limited searches to discipline-specific categories designed and assigned by librarians at the University of Michigan. Any authorized user could search the system, view abstracts, and have access to all free content (see below). Access to "full-length articles" (a designation assigned by Elsevier) depended on

3 This assertion and others concerning user preferences are based on our analysis of the ethnographic records compiled during the marketing process by Ann Zimmerman, and the results from a series of user surveys we administered during the project.

the user's institutional subscription package. With this access, articles could be viewed on screen or printed.

The delivery and management of such a large body of content (over 11,000,000 pages at the conclusion of the experiment) and the support of the PEAK experiment required a considerable commitment of both system and human resources. In addition to the actual delivery of content, project staff were responsible for managing the authentication mechanisms, collecting and extracting statistics, and providing user support for, potentially, tens of thousands of users.

PEAK ran primarily on a Sun E3000 with four processors, and was stored on several different configurations of RAID (redundant, fast access drive systems). User authentication and subscription/purchase information was handled by a subsidiary Sun UltraSparc.

Searching was conducted primarily with a locally-developed search engine called FTL (Faster Than Light). Bibliographic search used the OpenText search engine. The authentication/authorization server ran Oracle to manage user and subscription information. Several other types of software came into play with use of the system. They included

- Cartesian Inc.'s compression software, CPC, which allowed us to regain a significant amount of disk space through compression of the TIFF images;
- Tif2gif software developed at the University of Michigan, which converted images stored in CPC to GIFs;
- CPC, printps (for generating Postscript), and Adobe Distiller, which were used in combination to deliver images to users as PDF files; and
- The Stronghold web server, which provided SSL encryption for the security of user information.

Project staff at the University of Michigan Digital Library Production Service (DLPS) wrote middleware to manage the interoperation of the tools discussed above.

Designing and maintaining the PEAK system, as well as providing user support and service for the participant institutions, required significant staff resources. Once the system was specified by the research staff, design and maintenance of the system were undertaken by a senior programmer working close to full time in collaboration with a DLPS interface specialist. DLPS programming staff contributed as needed, and

the head of DLPS provided management. A full time programmer provided PEAK database support, collecting statistics for the research team and the participants, as well as maintaining the database of authorized users and the transaction database. Two librarians provided about one full-time equivalent of user support (one was responsible for the remote sites, the other for the University of Michigan community). Other UM library staff put in considerable time during the setup phases of PEAK to market the service to potential participants, some of whom required substantial education about the methods and aims of the experiment, and to formalize the licensing agreements with participants.

In order to facilitate per-article purchases, PEAK also needed to have the capacity to accept and process credit card charges. In the early months of the service, this billing was handled by First Virtual, a third-party electronic commerce company. This commercial provider also verified the legitimacy of users and issued virtual PINs that were to be used as passwords for the PEAK system. Less than half way through the PEAK experiment, First Virtual restructured and no longer offered these services. At that point, DLPS took over the processing of applications and passwords. Credit card operations were transferred to the University of Michigan Press.

We designed the system to support the planned research as well as to serve the daily information needs of a large and varied user community. Accommodating both purposes introduced a number of complexities. We sought to balance conflicting demands and to adhere to some fundamental goals:

- Providing meaningful intellectual access via a Web interface to a very large body of content.
- Balancing the aims of the research team with the library's commitment to making the content as easily accessible as possible.
- Enabling and supporting a number of different transaction types, taking into account that not all users have access to all types of transactions and that the suite of transaction choices may change over time, depending on the manipulation of experimental conditions.
- Enabling and supporting a number of different access levels, based on whether the user authenticates by password, the location of the user, the date of the material, and the type of material (e.g., full-length articles vs. other materials).

Tensions were exacerbated by our reliance on content from just one large commercial publisher and by the specific requirements for the research experiments. John Price-Wilkin, Head of DLPS, compared the production system problems to those of a standard service (Price-Wilkin 1999):

> The research model further complicates these methods for access, where all methods for access are not available to all institutions, and not all institutions choose to take advantage of all methods available to them. This creates a complex matrix of users and materials, a matrix that must be available and reliable for the system to function properly. Independence from Elsevier was critical in order for us to be able to test these models, and the body of Elsevier materials was equally important to ensure that users would have a valuable body of materials that would draw them into the research environment. The ultimate control and flexibility of the local production environment allowed the University of Michigan to perform research that would probably not have otherwise been possible, or could not have been performed in ways that the researcher stipulated.

4.4 Economic and experimental design

The PEAK system built upon existing digital library infrastructure and information retrieval mechanisms described above, but its primary purpose was to serve as a testbed for economic experiments in pricing and bundling information goods. Central to the PEAK experiment were the opportunities that electronic access creates for unbundling and re-bundling scholarly literature. A print-on-paper journal is, in itself, a bundle of issues. Each issue contains a bundle of articles, each of which is again a bundle of bibliographic information, an abstract, references, text, figures and many other elements.[4] In addition, the electronic environment makes possible other new dimensions of product variations. For example, access can be granted for a limited period of time (e.g., day, month, and year) and new services such as hyperlinks can be incorporated as part of the content. Permutations and combinations are

4 One pragmatic and unresolved issue for applying bundling models is the treatment of journal content other than articles. Many notices and reviews, as well as editorial content integral to a journal's identity, cannot be categorized as articles. How and when these items are indexed (thus becoming searchable) as well as how they should be priced are still open questions in electronic journal delivery and pricing.

Math6X9/2005/11/02:10:25 Page 53

almost limitless.[5]

Choosing what to offer from the different bundling alternatives is not an easy task. In the PEAK experiment, we were constrained by the demands of the experiment and the demands of the customers. Given the limited number of participants, bundle alternatives had to be limited in order to obtain enough experimental variation to support statistical analysis. The products had to be familiar enough to potential users to generate participation and reduce confounding learning effects. The economic models were designed by the University of Michigan research team, then reviewed and approved by a joint advisory board comprised of two senior Elsevier staff members (Karen Hunter and Roland Dietz), the University of Michigan Library Associate Director for Digital Library Initiatives (Wendy Lougee), and the head of the research team (Professor Jeffrey MacKie-Mason).

After balancing the different alternatives and constraints, we selected three different bundle types as the products for the experiment. We refer to the product types as traditional subscriptions, generalized subscriptions and single articles (sometimes called the "per-article" model).[6] We now describe these three product offerings.

Traditional subscription: A user or a library could purchase unlimited access to a set of articles designated as a journal by the publisher for $4/issue if the library already held a print subscription. The value-based logic supporting this model is that the content value is already paid in the print subscription price, so the customer is only charged for an incremental electronic delivery cost. If the institution was not previously subscribed to the paper version, the cost of the traditional subscription was $4 per issue, plus 10% of the paper version subscription price. In this case, the customer is charged for the electronic delivery cost plus a percentage of the paper subscription price to reflect the value of the content. The electronic journals corresponded to the Elsevier print-on-paper journal titles. Access to subscribed content continued throughout the project. The traditional subscription is a "seller-chooses"

5 See MacKie-Mason and Riveros (2000) for a more complete discussion of the economic design space.

6 All older materials (i.e., pre-1998) were freely available to all project participants, as were bibliographic and full-text searches, with no charges levied for viewing citations, abstracts or non-article material such as reviews and notices. The remaining content (essentially, full-length articles published after 1997) could be purchased through one of the product types.

bundle, in that the seller, through the editorial process, determines which articles are delivered to subscribed users.

Generalized subscription: An institution (typically with the library acting as agent) could pre-purchase unlimited access to a set of any 120 articles selected by users. These pre-purchases cost $548 for the bundle, which averages about $4.50 per article selected. This is a "user-chooses" bundle. With a generalized subscription, the user selected which articles were accessed, from across all Elsevier titles, after the user had subscribed. Once purchased the article was available to anyone in that user community.

Per-article: A user could purchase unlimited access to a specific article for $7/article. This option was designed to closely mimic a traditional document delivery or interlibrary loan (ILL) product. With ILL the individual usually receives a printed copy of the article that can be retained indefinitely. This is different from the "per-use" pricing model often applied to electronic data sources. The article was retained on the PEAK server, but the user could access a paid-for article as often as desired. This was a "buyer-chooses" scheme, in that the buyer selected the articles purchased.

The per-article and generalized subscription options allowed users to capture value from the entire corpus of articles without having to subscribe to all of the journal titles. Once the content is created and added to the server database and the distribution system is constructed, the incremental cost of delivery is approximately zero. Therefore, to create maximal value from the content, it is important that as many users as possible have access. The design of the price and bundling schemes affected both how much value was delivered from the content (the number of readers), and how that value was shared between the users and the publisher.

Institutional generalized subscriptions may be thought of as a way to pre-pay for individual document delivery requests. One advantage of generalized subscription purchases for both libraries and individuals is that the "tokens" cost substantially less per article than the per article license price. By predicting in advance how many tokens would be used (and thus bearing some risk), the library could essentially pre-pay for document delivery at a reduced rate. However, unlike commercial document delivery or an interlibrary loan, all users within the community have ongoing unlimited access to the articles obtained with generalized

subscription tokens. Thus, for the user community, a generalized subscription combines features of both document delivery (individual article purchase on demand) and traditional subscriptions (ongoing shared access). One advantage to a publisher is that generalized subscriptions represent a committed flow of revenue at the beginning of each year, and thus shift some of the risk for usage (and revenue) variation onto the users. Another is that they allow access to the entire body of content to all users and, by thus increasing user value from the content, provide an opportunity to obtain greater returns from the publication of that content.

A simplified example illustrates why a library might spend more purchasing generalized subscriptions than traditional subscriptions. Consider a traditional subscription with 120 seller-specified articles, selling for $120, and a generalized subscription that allows readers to select 120 articles from the larger corpus, also for $120. Suppose that in the traditional subscription, users get zero value from half of the articles, but something positive from the other half. Then, the library is essentially paying $2 per article for which users have positive value. In this example, a cost-benefit oriented library would only purchase traditional subscriptions as long as the average value for the articles users want is at least $2 each. In a generalized subscription, however, users select articles that they actually value (they are not burdened with unwanted articles), so the average cost is $1 per article the user actually wants to read. The library then might justify a budget that continues buying generalized subscriptions to obtain the additional articles that are worth more than $1 but less than $2 to users. The result is more articles and more publisher revenue than with traditional subscriptions. Of course, the library decision process is more complicated than this, but the basic principle is that users get more value for the dollar spent when they—not the sellers—select the articles to include, and thus, since additional spending creates more user value wth the generalized subscription, over time the library might spend more.

The twelve institutions participating in PEAK were assigned randomly to one of three groups, representing three different experimental treatments, which we labeled the Greeen, Blue and Red groups. Users in every group could purchase articles on a per-article basis. In the Green group they could also purchase institutional generalized subscriptions; in the Blue group they could purchase traditional subscriptions; in the

Red group they could purchase both traditional and generalized subscriptions in addition to individual articles.

Regardless of treatment group, access was further determined by whether the user had logged in with a PEAK password. Use of a password allowed access from any computer (rather than only those with authorized IP addresses) and, when appropriate, allowed the user to utilize a generalized subscription token. The password authorization protected institutions from excessive usage of pre-paid generalized subscription tokens by walk-in users at public workstations. The password was also required to purchase articles on a per-article basis (to secure the financial transaction) and to view previously purchased articles (to provide some protection to the publisher against widespread network access by users outside the institution).

The password mechanism was also useful for collecting research data. Usage sessions were of two types: those authenticated with a PEAK password and those not. For unauthenticated sessions, the system recorded which IP addresses accessed which articles from which journal titles. When users employed their PEAK passwords, the system recorded which specific users accessed which articles. Uses that resulted in full-article accesses were also classified according to which product type was used to pay for access. Because we recorded all interface communications, we were able to measure complex "transactions". For example, if a user requesting an article (i.e., clicked on its link) was informed that the article was not available in the (traditional or generalized) subscription base, we could distinguish between whether the user chose to pay for the article on a per article basis or decided to forego access.

An important consideration for both the design and analysis of the experiment was the possibility of learning effects during the life of the project. If present, learning makes it more difficult to generalize results. We expected significant learning about the availability and use of the system due to the novelty of the offered products and the lack of user experience with electronic access to scholarly journals. To decrease the impact of learning effects, we worked with participating institutions to actively educate users about the products and pricing prior to implementing the service. Data were also collected for almost two years, which enabled us to isolate some of the learning effects.

To test institutional choices (whether to purchase traditional or generalized subscriptions, and how many), we needed participation from

Math6X9/2005/11/02:10:25 Page 57

several institutions. Further, to explore the extent to which electronic access and various price systems would increase content usage, we needed a diverse group of potential (individual) users. Therefore, we marketed the project to a diverse group of institutions: four-year colleges, research universities, specialized research schools, and corporate research libraries. A large user community clearly improved the breadth of the data, but also introduced other complications. For example, user behavior might be conditioned by the characteristics of the participating institutions (such as characteristics of the institution's library system, availability of other electronic access products, institutional willingness to reimburse per-article purchases, etc.).

4.5 Pricing

Pricing electronic access to scholarly information is far from being a well-understood problem. Contemporaneous with the PEAK experiment, Prior (1999) reported that, based on a survey of 37 publishers, when both print-on-paper and electronic versions were offered 62% of the publishers had a single combined price, with a surcharge over the paper subscription price of between 8% and 65%. The most common surcharge was between 15% and 20%. Half of the respondents offered electronic access separately at a price between 65% and 150% of print, most commonly between 90% and 100%. Fully 30% of the participating publishers changed their pricing policy in just one year (1999). In this section, we will describe the pricing structure we implemented in the PEAK experiment and our rationale for it.

For content that can be delivered either on paper or electronically, there are three primary cost categories: content cost, paper delivery cost and electronic delivery costs. The price levels chosen for the experiment reflect the components of cost, adjusted downward for an overall discount to encourage participation in the experiment.[7]

To recover the costs of constructing and operating the electronic delivery system, participating institutions paid the University of Michigan an individually negotiated institutional participation license (IPL) fee,

7 If the scholarly publishing market is competitive, then we expect long-run prices to reflect incremental component costs. Whether this market is competitive is a contentious issue; see McCabe (this volume) for one view on this question.

roughly proportional to the number of authorized users. To account for the content cost, institutions or individual users paid the access prices associated with each product type described above (traditional subscriptions, generalized subscriptions, or individual articles)

Arbitrage possibilities impose some constraints on the relative prices between access options. Arbitrage arises when users can choose different options to replicate the same access. For example, the PEAK price per article in a per-article purchase had to be greater than the price per article in a generalized subscription, and this price had to be greater than the price per article in a traditional subscription. These inequalities impose the restriction that the user could not save by replicating a traditional subscription through purchasing individual articles or a generalized subscription, nor save by replicating a generalized subscription by paying for individual articles. Alternatively, arbitrage constrains publishers to charge a price for bundles that is less than the sum of the individual component prices.

The mapping of component costs to price levels is not exact, and in some cases the relationship is complicated. For example, although electronic delivery costs are essentially zero, there is some additional cost to creating the electronic versions of the content (especially at the time of the PEAK experiment because Elsevier's current production process was not fully unified for print and electronic publication). Therefore, the electronic access price might be set in a competitive market to recover both the content value and also some amount of incremental electronic delivery cost.

Based on the considerations above, and on negotiations with the publisher, we set the following prices: per article at \$7; generalized subscriptions at \$548 for 120 articles; and traditional subscriptions at \$4 per issue plus 10% of the paper subscription price. A substantial amount of material, including all content available that was published two calendar years prior, was available without any additional charge after an institution paid the IPL fee for the service. We refer to this as "unmetered". Full-length articles from the current two calendar years were "metered": users could access them only if the articles were paid for under a traditional or generalized subscription, or purchased on a per-article basis.

4.6 Results

In this section we report some descriptive results from both the experiment and the production service. See Gazzale and MacKie-Mason (this volume) for a more detailed study of the economic experiment results.

Revenues and costs

In Table 4.1 we summarize PEAK revenues. The actual total was over $580,000 (the sum of total revenue in the first two rows).[8] The first and third rows report annual revenues, with 1999 adjusted to reflect an estimate of what revenues would have been if the service were to run for the full year (it ended in August 1999, but only six months of content were included, and prices were adjusted accordingly). On an annualized basis, two-year revenues were about $712,000.

Between the first and second year of the service, the number of traditional subscriptions substantially decreased: this occurred because two schools cancelled all of their (electronic) subscriptions. By reducing the number of journal titles under traditional subscription, the users of these libraries needed to rely more heavily on the availability of unused generalized subscription tokens, or they had to pay the per-article fee. We see from the table that the annualized revenues for per-article purchasing are seventeen times higher in 1999 than in 1998, and that the 1999 generalized subscription revenues (annualized) are 8% lower than in 1997–1998.

A full calculation of the costs of supporting the PEAK service is difficult, given the mix and dynamic nature of costs (e.g., hardware). We estimate that total expenditures by the University of Michigan were nearly $400,000 during the 18 month life of the project. Of this cost, roughly 35% was expended on technical infrastructure and 55% on staff support (i.e., system development and maintenance, data loading, user support, authentication/authorization/security, project management). Participant institution (IPL) fees covered approximately 45% of the project costs, with vendor and campus in-kind contributions covering another

8 Due to delays in starting the project, the first revenue period covered content from both 1997–98, although access was available only during 1998. For this period, prices for traditional subscriptions were set to equal $6/issue, or 1.5 times the annual price of $4/issue, to adjust for the greater content availability.

Table 4.1
PEAK Revenues

	Traditional Subscriptions		Generalized Subscriptions		Individual Articles		Total Access	IPL****	Total
Year	Quantity	Revenue	Quantity	Revenue	Quantity	Revenue	Revenue	Revenue	Revenue
1997–1998	1939	$216,018	151	$82,748	275	$1,925	$300,691	$140,000	$440,691
1999*	1277	$33,608	92	$50,416	3186	$22,302	$106,326	$42,000	$148,326
Annualized 1999**	1277	$78,996	138	$75,624	4779	$33,453	$188,073	$84,000	$272,073
Total 1997–1999***	3216	$295,014	289	$158,372	5054	$35,378	$488,764	$224,000	$712,764

NOTE: * Partial year results, January to August 1999; new articles availabe only if published before June.
** Annualization done by scaling the quantity of generalized subscriptions and per article puchases. Traditional subscriptions priced at the full year rate.
*** Annualized.
**** The "IPL" is the Institutional Participation License, an annual fee charged to each participating institution.

Table 4.2
Distribution of users with passwords by status and academic division

Division	Status					Total
	Faculty	Staff	Graduate Student	Undergrad	Other	
Engineering, science and medicine	408	214	1032	211	38	1903
Architecture and urban planning	103	11	47	16	19	196
Education, business, information/library science and social science	91	43	287	46	2	469
Other	178	240	350	176	34	978
Total	780	508	1716	449	93	3546

20–25%.[9] The UM Digital Library Production Service contributed resources to this effort, reflecting the University of Michigan's desire to provide this service to its community, and also its support for the research.

4.7 User demographics

In the PEAK project design, unmetered articles and articles covered by traditional subscriptions could be accessed by any user from a workstation associated with one of the participating sites (authenticated by the computer's IP address). If users wanted to use generalized subscription tokens or to purchase individual articles on a per-article basis they had to obtain a password and use it to authenticate.[10] We have more complete data on the subset of users who obtained and used passwords.

In Table 4.2 we report the distribution of the more than three thousand users who obtained passwords and who used PEAK at least once. Most of the users are from engineering, science and medicine, reflecting the strength of the Elsevier collection in these disciplines. 70% of these users were either faculty or graduate students (see Figure 4.1).

9 The University of Michigan production service retained IPL annual participation fees. The publisher received the content charges, minus a modest service fee for the credit card service provided by the University of Michigan Press.

10 Through an onscreen message we encouraged all users to obtain a password and use it every time in order to provide better data for the researchers. From the data, we concluded that only a small fraction apparently chose to obtain passwords based solely on our urging; most who did apparently obtained passwords because they were necessary to access a specific article.

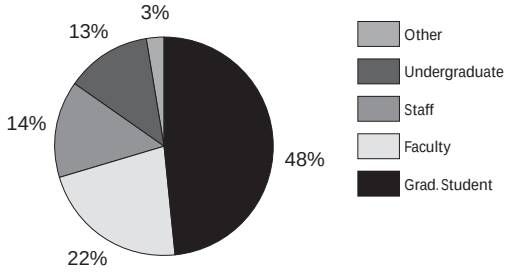


Figure 4.1
Distribution of users who otained passwords and used them to access PEAK

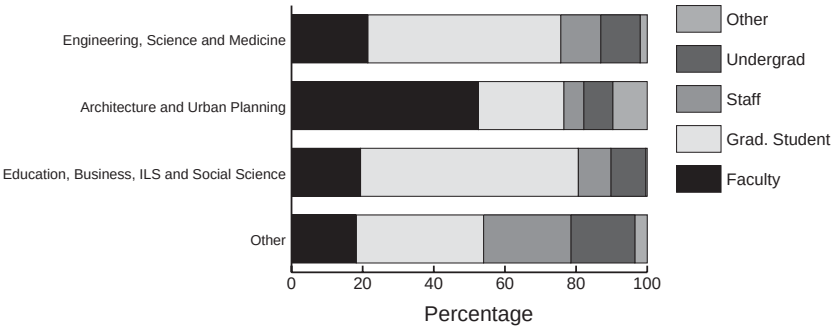


Figure 4.2
Users with Passwords Who Accessed PEAK

The relative fractions of faculty and graduate students varies widely by discipline (see Figure 4.2). Our sample of password-authenticated users, while probably not representative of all electronic access usage, includes all those who accessed articles via either generalized subscription tokens or per-article purchase. It represents the interested group of users, who were sufficiently motivated to obtain and use a password. Gazzale and MacKie-Mason (this volume) discuss the effects of passwords and other user costs on user behavior.

In Table 4.3 we summarize usage of PEAK through August 1999. Authorized users joined the system gradually over the first nine months of 1998. There were 208,104 different accesses to the content in the PEAK

Math6X9/2005/11/02:10:25 Page 63

Table 4.3
Total number of unique content accesses by treatment group and type of access (Jan 1998–August 1999)

Access Type	Treatment group			
	Green	Red	Blue	All Groups
Unmetered	24632	96658	13911	135201
Traditional subscription articles				
1st use	N/A	27140	2881	30021
2nd or higher use	N/A	11914	597	12511
Generalized subscription articles				
1st use	8922	9467	N/A	18389
2nd or higher use	3535	4789	N/A	8324
Individually purchased articles				
1st use	194	75	3192	3461
2nd or higher use	108	26	63	197
Total accesses	37391	150069	20644	208104

NOTE: See definitions of treatment groups at p.55, above.

system over 17 months.[11] Of these, 65% were accesses of unmetered material (not-full-length articles, plus all 1998 accesses to content published pre-1997, and all 1999 accesses to pre-1998 content).[12] However, one should not leap to the conclusion that users will access scholarly material much less when they have to pay for it, though surely that is true to some degree. To correctly interpret the "free" versus "paid" accesses we need to account for three effects. First, to users much of the metered content appeared to be free: the libraries paid for the traditional subscriptions and the generalized subscription tokens. Second, the quantity of unmetered content in PEAK was substantial: on day one, approximately January 1, 1998, all 1996 content and some 1997 content was in this category. On January 1, 1999, all 1996 and 1997 content and some 1998 content was in this category. Third, the nature of some unmetered content (for example, letters and announcements) is different from metered articles, which might also contribute to usage differences.

Generalized subscription "tokens" were used to purchase access to 18,389 specific articles ("1st use"). These articles were then distinctly ac-

11 We limited our scope to what we call "unique accesses," counting multiple accesses to a given article by a single individual during a PEAK session as only one access. For anonymous access (i.e., access by users not entering a password), we define a "unique" access as any number of accesses to an article within 30 minutes from a particular IP address. For authenticated users, we define a "unique" access as any number of accesses to an article by an authenticated user within 30 minutes of the first access.

12 See the definition of unmetered material in the text at p. 58, above.

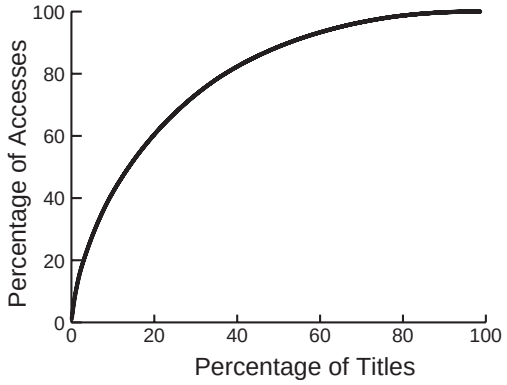


Figure 4.3
Concentration of article accesses across different journal titles

cessed an additional 8,324 times ("2nd or higher use"), for an average of 1.45 accesses per generalized subscription article. Traditional subscription articles had an average of 1.42 accesses per article. A total of 3461 articles were purchased individually on a per-article basis; these were accessed 1.06 times per-article on average. The difference in the number of accesses per article for articles obtained by generalized subscription and by per-article purchase is likely due to the difference in who may access the article after initial purchase. *All* authorized users at a site could access an article once it has been purchased with a generalized subscription token, while only the individual making a per-article purchase has the ability to re-access that article. Thus, we estimate that for individually purchased articles (whether by generalized subscription token or per-article purchase), the initial reader accessed the articles 1.06 times, and additional readers accessed these articles 0.39 times. That is, there appears on average at least one-third additional user per article under the more lenient access provisions of a generalized subscription token.

In Figure 4.3 we show a curve that reveals the concentration of usage among a relatively small number of Elsevier titles. We sorted articles that were accessed from high to low in terms of how often they were accessed. We then determined the smallest number of articles that, together, comprised a given percentage of total accesses, and counted the number of journal titles from which these articles were drawn. For

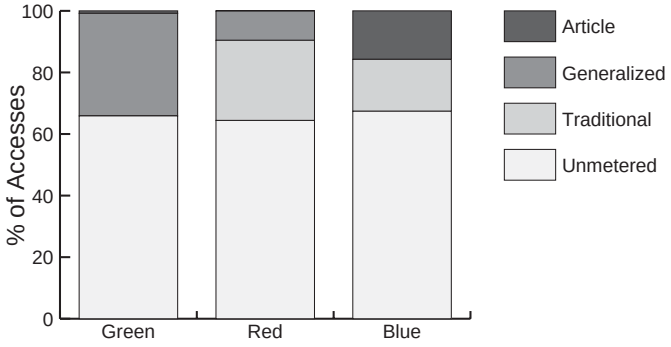


Figure 4.4
Percentage of model used by experimental group: Jan 1998–Aug 1999

example, 37% of the 1200 Elsevier titles generated 80% of the total accesses. 40% of the total accesses were accounted for by only about 10% of the journal titles.

In Figure 4.4 we compare the fraction of accesses within each treatment group that are accounted for by traditional subscriptions, generalized subscriptions and per-article purchases. Recall that the Green and Blue groups only had two of the three access options.[13] When institutions had the choice of purchasing generalized subscription tokens, their users purchased essentially no access on a per-article basis. Of course, this makes sense as long as tokens are available: it costs the users nothing to use a token, but it costs real money to purchase on a per-article basis. Indeed, our data indicate that institutions that could purchase generalized subscription tokens tended to purchase more than enough to cover all of the demand for articles by their users; i.e., they didn't run out of tokens in 1998. We show this in aggregate in Figure 4.5: only about 50% of the tokens purchased for 1998 were in fact used. Institutions that did not run out of tokens in 1999 appear to have done a better job of forecasting their token demand for the year (78% of the tokens purchased for 1999 were used). Institutions that ran out of tokens used about 80% of the tokens available by around the beginning of May.

Articles in the unmetered category constituted about 65% of use

13 Individual article purchase was available to both; Green institutions could also purchase generalized subscriptions, and Blue could purchase traditional subscriptions.

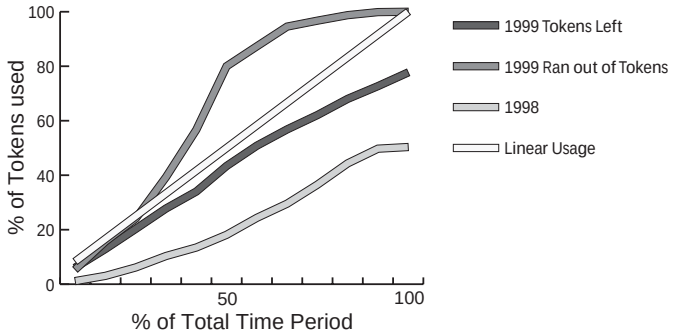


Figure 4.5
Percentage of pre-paid tokens used as a percentage of tiem available

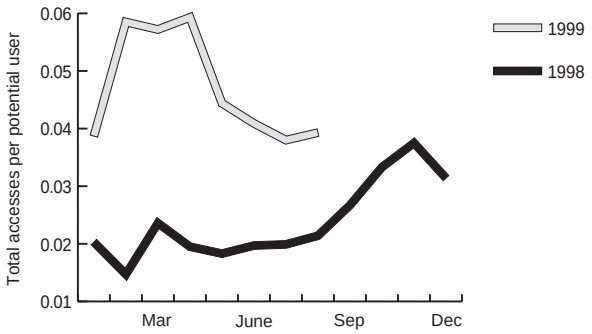


Figure 4.6
Total accesses per potential user: Jan 1998–August 1999

across all three groups, regardless of which combination or quantity of traditional and generalized subscriptions an institution purchased. The remaining 35% of use was paid for with a different mix of options depending on the choices available to the institution. Evidently, none of the priced options choked off use altogether.

We show the total number of accesses per potential user for 1998 and 1999 in Figure 4.6. We divide by potential users (the number of people authorized to use the computer network at each of the participating institutions) because different institutions joined the experiment at different times. This figure thus gives us an estimate of learning and

Table 4.4
Learning: usage comparison across two years (March–May averages)

	1998	1999	Percentage Change
Unmetered	19291	55745	189%
Traditional	6374	10560	66%
1st Token	1648	4805	192%
1st per-article purchase	1	1288	N/A
2nd or higher Token	3060	8166	167%
2nd or higher per-article purchase	8	472	5800%
Total	30382	81036	167%

seasonality effects in usage. Usage per potential user was relatively low and stable for the first 9 months. However, it then increased to a level nearly three times as high over the next 9 months. We expect that this increase was due to more users learning about the existence of PEAK and becoming accustomed to using it. Note also that the growth begins in September 1998, the beginning of a new school year with a natural bulge in demand for scholarly articles. We also see pronounced seasonal effects in usage: local peaks in March, November and April.

To see the learning effect without interference from the seasonal effect, we calculated usage by type of access in the same three months (March–May) of 1998 and 1999; see Table 4.4. Overall, usage increased 167% from the first year to the second.

We considered the pattern of repeat accesses distributed over time. In Figure 4.7 we show that about 93% of articles accessed were accessed no more than two times. To further study repeat accesses, we selected only those articles (7%) that were accessed three or more times between January 1998 and August 1999 (high use articles). We then counted the number of times they were used in the first month after the initial access, the second month after, and so forth; see Figure 4.8. What we see is that almost all access to even high use articles occurred during the first month. After that, a very low rate of use persisted for about seven more months, then faded out altogether. Thus, we see that, even among the most popular articles, recency was very important.

Although recency appears to be quite important, we saw in Table 4.1 that over 60% of total accesses were for content in the unmetered category, most of which was over one year old. Although we pointed out that the monetary price to users for most non-unmetered articles was still zero (if accessed via institution-paid traditional or generalized

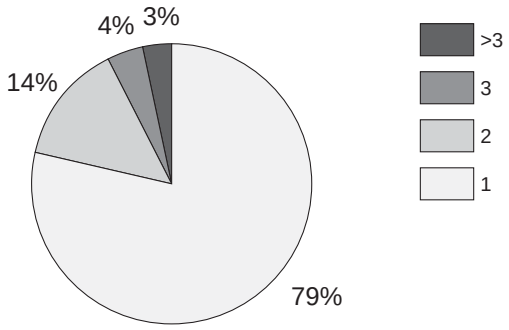


Figure 4.7
Percentage of Articles by Number of Times Read

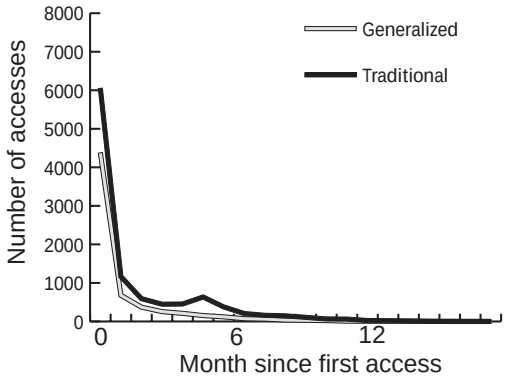


Figure 4.8
The distribution of usage for high use articles

subscriptions), there were still higher user costs for much of the more recent usage. If a user wanted to access an article using a generalized subscription token, then she had to obtain a password, remember it (or where she put it) and use it. If the article was not available in a traditional subscription and no tokens were available, then she had to do the above plus pay for the article with hard currency. Therefore, there are real user cost differences between the unmetered and metered content. The fact that usage of the older, unmetered content is so high, despite the clear preference for recency, supports the notion that users

respond strongly to costs of accessing scholarly articles.[14]

4.8 Discussion

PEAK was a rather unique project. During a relatively early stage of the transition to digital scholarly collections, we delivered a large-scale production service containing several years of content from over 1100 journal titles, for a total of about 10 million pages. On top of the commitment to a production-quality service we also implemented a field experiment to test usage response to several different pricing and bundling models, one of which (generalized subscriptions) was quite novel. We summarize our most important general conclusions here:

- Of the purchasing options we offered, the generalized subscriptions—our innovation—was the most popular and generated the most interest. Libraries saw the generalized subscription as a way of increasing the flexibility of their journal budgets and of tying purchasing more closely to actual use. The generalized subscription provides fast and easy access to articles in demand from the complete corpus, not just from a subscribed subset of titles.
- We observed great interest in the monthly statistical reports on local article use that we generated for participating libraries. Participants were eager to use these to help assess current subscription choices and to further understand user behavior.
- The user cost of access—comprised not just of monetary payments, but also of time and effort—has a significant effect on the number of articles that readers access. (See Gazzale and MacKie-Mason (this volume) for further detail.)
- There was a substantial learning period during which users became

14 In another preliminary test of the impact of user cost on usage, we compared the usage of the Red and Blue groups. Red institutions had both generalized and traditional subscriptions available; Blue had only traditional. We calculated the number of paid articles accessed (paid by generalized tokens or per-article) for each group, after normalizing by the number of traditional subscriptions, and the number of potential users at the institutions. We found that when generalized subscriptions were available, which have a much lower user cost since the library pays for the tokens, three times as many articles were accessed as at institutions which had to pay for each of these articles on a per-article basis. See Gazzale and MacKie-Mason (this volume).

aware of the service and accustomed to using it. It appears that usage was increasing even after a year of service. By the end of the experiment, usage was at a rather high level: approximately five articles accessed per month per 100 potential users, with potential users defined broadly (including all undergraduate students, who rarely use scholarly articles directly).

- It has long been known that overall readership of scholarly literature is low. We have seen that even the most popular articles are read only a few times, across 12 institutions. We did not, however, measure how often those articles were being read in print versions during the same periods.
- Recency is very important: repeat usage dropped off considerably after the first month. (This was also reflected in user comments, not reported above.)

PEAK had a limited life by design, and today most of the major publishers of scholarly journals have implemented their own PEAK-like service. The environment is far from stable, however, and service options, pricing and bundle offerings continue to evolve. Our results bear on the economics and usage of digital collections today and in the future, and provide some support for particular design choices.