

## **SUBSIDIZING END USER ACCESS TO RESEARCH DATABASES: FROM CARD FILE TO THE WORLD WIDE WEB**

**By Joseph A. Busch and Angela Giral\***

**Abstract:** This paper reviews work of the Getty Art History Information Program (AHIP, recently renamed the Getty Information Institute) over the past decade to subsidize end user on-line access to scholarly research databases. A brief history of the formation of AHIP is followed by a summary of a study of subsidized access to Dialog databases by Getty Center resident scholars which was reported in a series of papers by Marcia Bates, professor of Library Science at the University of California, Los Angeles. Dr. Bates found that searching patters of humanities researchers differed substantially from previous studies in the sciences and social sciences disciplines. The second part of the paper presents a model of the relative merits and opportunities associated with the various contractual arrangements and incentive strategies employed by AHIP with vendors such as Dialog, consortia such as the Research Libraries Group, and a CD-ROM publication program, compared to print publications and the experimental offer to access to some of the AHIP databases over the World Wide Web.

### **Introduction**

Since its inception, the Getty Art History Information Program (AHIP, recently renamed the Getty Information Institute) has worked to create scholarly information resources and data standards. These resources and standards are fundamental and critical to the cultural disciplines. They are also building blocks and enabling mechanisms for information networks. All of AHIP's activities are grounded in the belief that networks will eventually connect arts and humanities information across national boundaries; and that this networked information must maintain scholarly depth and perspective for research and education while offering content and form that appeal to broader audiences.

In the early 1980's the Getty Trust agreed to undertake the costs of operating the *Avery Index to Architectural Periodicals* and the *International Repertory of the Literature of Art/Répertoire International de Littérature de l'Art* (known as

---

\* Paper presented at the Workshop of the IFLA Art Libraries Section on the topic „Pay or Profit: Fee or Free“ at the 62<sup>nd</sup> General Conference of IFLA in Beijing 1996

RILA). In 1989, RILA merged with the *Répertoire d'Art et d'Archéologie* (known as RAA) to form the *Bibliography of the History of Art/Bibliographie d'Histoire de l'Art* (known as BHA) which is jointly operated with the Institut National de l'Information Scientifique et Technique (INIST). Together with the *Provenance Documentation Collaborative*, a consortium of libraries and archives in Europe and North America, AHIP's current Research Database Program is producing seminal resources that support the study and practice of architecture, art history, and associated disciplines. In partnership with affiliated institutions and individual researchers more than ten year's dedication to building these resources has brought them to maturity just as information networks have advanced to a state where they can exploit them.

Today, our continued investment in and commitment to the resource databases is taking advantage of increased inter-connectivity not only in shaping user access to the resources but in creating decentralized and participatory mechanisms for adding to those resources. In the future, these resources may be created in cooperation with their users. They will contribute to the selection, description, and indexing of relevant research materials. Users will add links to related information that exist in the many and various digital libraries, archives, and museums. The role of the Research Database Program will be to orchestrate, coordinate, and edit information about our shared cultural heritage worldwide, and the research databases will appeal to a broad audience of specialists as well as the public.

### **Case Study 1: Changes in Pricing of the *Avery Index to Architectural Periodicals***

The *Avery Index to Architectural Periodicals* was founded in the mid-1930's as an extension to the reference services of Avery Architectural and Fine Arts Library. Originally maintained as a card file available only at the library, it was first published and made available to the outside world by G. K. Hall in 1968.

In 1974 Columbia University became one of the founding members of the Research Libraries Information Network (RLIN). The process of computerization, designed to enhance worldwide accessibility to the *Index*, had the immediate impact of slowing down the creation of records, on the one hand, and decreasing accessibility to the users of the Avery Library for whom the *Index* had originally been created. To recover costs, RLIN charged its members for deriving a cataloging record from an original record contributed by another institution. It also levied a fee for searching the growing bibliographic database. When the special databases were created (the *Avery Index* being the first of these) it was

known that no one was going to derive records from them. Thus the contributing institution, the Avery Library in this case, paid a fee for creating its own records in RLIN. A surcharge of 50% per connect hour was then charged to searchers to compensate the creators for the fees they paid for creation as well as for the intellectual effort of creating the analytic records. Online searching being what it was in those days, the income did not cover for the costs. Thus, as a result of computerization, the creation of *Avery Index* records was costing Columbia more while it was serving its users less than before computerization.

In October 1983 the *Index* became an operating program at the J. Paul Getty Trust and in May 1984 it became one of the constituent activities of the newly created Art History Information Program. One of the first things the Getty Trust did was to lower the cost of searching the *Avery Index* online by eliminating the surcharge charged for searching it on RLIN. Thereafter users of the *Avery Index* online on RLIN paid only \$60 per hour of connect time instead of the \$90 they had been paying up to that time. Also, thanks to Getty support, the Avery was able to expand the indexing staff so that the accumulated backlog was eliminated, and the *Index* was able to be kept current. Finally, the Getty subsidized the preparation of camera-ready pages from the online database, so that G. K. Hall could print annual „supplements“ to the *Index* and sell them to subscribers at an affordable price. The first computer-generated volumes were published in 1985 and titled *The Fourth Supplement (1979-1982)*.

Then the costs of technology changed. The cost of computer storage decreased as the cost of telecommunications with dedicated lines increased, making it less cost-effective for RLIN members to create records on a central mainframe computer located in California, and more cost-effective to create records on local cataloging systems. This threatened to diminish the value of the central RLIN bibliographic database as a source for „copy cataloging.“ At the same time, the use of the RLIN bibliographic database for verification, reference, and interlibrary loan was increasing.

RLIN decided to change the cost-recovery formula and began charging on the basis of the number and kind of searches performed instead of on the number of connect hours. For example, a search by the record identification number (which is very efficient) cost half or a quarter as much as a search by any other field. To encourage institutions to contribute good quality records to the database, a system of searching credits granted for the contribution of original records was devised. Lastly, the cost per search became lower when an institution purchased a block of searches in a year. But all of these pricing strategies were designed for a bibliographic service aimed at a professional market of libraries and librarians.

The full implementation of local area networks with gateways to the Internet were just emerging.

In 1987, the Avery conducted an end user pilot study to assess the feasibility of providing users direct access to the *Index* without the intervention of a librarian, much as they had had when the *Index* existed only in card form. The project was advertised to faculty and students in the Department of Art History and the Graduate School of Architecture at Columbia University and flyers were posted throughout the Avery Library to capture potential alumni users. The flyer offered a week of free access to the database if users agreed to come for a two-hour training session, to have their transactions recorded, and to answer a questionnaire. Of a potential user population estimated to be 800 students and faculty, only 14 applied to participate in the pilot project. Half of these received instruction and access information, including a password, but only three continued with the project by actually logging in search time from their home or office.

In spite of the small population of this study, the Avery was able to learn some useful things, reported fully by Janice Woo in the „Final Report“ [1]. Perhaps the most significant findings were that the transaction logs showed successful searching whether the user had received the two-hour training or just the written instructions, and that all three were willing to pay for access from home or office - but no more than \$10.00 per hour.

However, it was not until 1992 that two important developments enabled enhanced direct user-access to the *Avery Index* online: RLIN designed a more user-friendly interface, named „Eureka,“ and the *Avery Index* moved from the special database environment to a new service entitled „CitaDel.“ This is a subscription service where an institution pays an annual subscription fee for wide access through their campus or local area network. The subscription charge was initially based on the total number of potential users, but was changed to the current formula based on the number of simultaneous users.

Columbia University keeps statistics on the use of the various databases offered on its Clio Plus campus network bibliographic system. The *Avery Index* receives an average of 4,000 searches per month. It is unknown how many searches are performed at other institutions, but at least 60 are offering such access (through RLIN) to their faculty and students throughout the world.

The Research Libraries Group (RLG) is a non-profit consortium that has to recover the costs of the services it provides. RLG has made creative use of pricing policies to encourage certain behaviors while discouraging others. It is clear that

for arts and humanities databases (among which Avery is counted) are not commercially viable, let alone profitable. The Avery Index exists because first Columbia University, and then the Getty AHIP have subsidized the creation of the information it contains. Access to this information has traditionally been purchased by institutions for the free use of their selected users: first through the purchase of the printed books, and more recently through the purchase of either the CD-ROM version or subscription to online access through RLIN. Never has the individual user paid directly for such information and those who were asked did not volunteer to pay a large fee.

### **Case Study 2: Getty On-line Searching Project**

From 1989-90, AHIP conducted a study of subsidized access to Dialog databases by resident scholars at the Getty Center for the History of Art and the Humanities. The visiting scholars were given unlimited on-line searching access to Dialog databases from a workstation in the Getty Center Library. Before each search, the scholars typed their research question in their own words, and during the search a log of all entries and system responses was captured by the computer. With the scholar's permission, this transaction log was analyzed by Marcia Bates, professor of Library Science at the University of California, Los Angeles, with the assistance of students and Getty staff, and reported on in a series of papers [2-6]. This log of the on-line search activity of 28 humanities researchers over a two-year period is a unique data set representing the perspective of humanities researchers that can be compared to similar bodies of data representing on-line searchers in the science and social science disciplines.

The statements of research questions by the scholars in the Getty project has been particularly valuable in documenting differences in the search behavior of humanities researchers compared to the search behavior in other disciplines. As illustrated in *figure 1* approximately half of the humanities search questions included personal names, a quarter geographical terminology, a quarter „discipline“ terms, and a sixth contained chronological terminology. (Note that the total percentages are greater than 100 because search question statements may have included more than one type of search term). Discipline terms are the names of disciplines such as „art history“ or „rhetoric.“ These observations confirm that scientists and social scientists largely search by common subject terminology, while humanists make extensive use of formal names.

Type of Subject Term	Humanities	Sciences & Social Sciences
	<i>Getty Study</i>	<i>NSF Study [7]</i>
Work or publication	5%	3%
Personal name	49%	0%
Geographical name	25%	8%
Chronological term	17%	3%
Discipline term	23%	0%
Other proper term	7%	11%
Other common term	57%	100%

*Figure 1 - Comparison of types of terms in statements of research questions.*

While humanists use different conceptual models in formulating their research questions, the principles of databases design and of on-line search services such as Dialog are specific to the needs of users in the bio-medical and technological disciplines. Dr. Bates reports that --

- The free Dialog search service was not used very extensively; most used it for less two hours total.
- Search statements were very simple; approximately two thirds were Select statements.
- Boolean searching is a problem for humanists; more than one quarter never used a Boolean OR.

These results do not diverge significantly from other research into studies of end user on-line searching behavior regardless of discipline.

The most interesting of Dr. Bates' papers is no. 4 [2]. In it she speculates on the intrinsic nature of humanities research, how it differs from other disciplines in

terms of, for example, the objects of study and the relationship between research and the literature which documents it. She observes from the interviews

conducted with scholars who participated in the Getty Project that the value of on-line (or manual) searching for these scholars was to identify publications that were outside their normal research areas. Finally, she discusses what database services might be of interest to humanists, how databases might be designed to address the specific needs of humanists, and search interface and functional design considerations for this special class of end users.

Taken together, the reports of the Getty Project provide many insights into the potential for on-line searching in the humanities. They also provide a comprehensive justification of AHIP's research database program and a road map for its improvement in the future.

### **Comparison of Distribution Models**

The Getty's involvement with supporting the creation of research resources bridges the transition from printed card and book indexes to on-line databases and CD-ROM's. The early computer systems and editorial standards and policies were designed to support centralized data collection and editorial work, and the automation of typesetting and distribution of printed books. In the mid-1980's, AHIP began to experiment with commercial and non-profit on-line search services such as Dialog and the Research Libraries Information Network (RLIN). In 1994, AHIP began publishing on CD-ROM, and last year initiated limited World Wide Web access to some databases on an experimental basis.

From the beginning, the costs of collecting, editing, and publishing high quality research information has far outstripped the revenues generated from their use. Supporting these costs has amounted to a huge subsidy to the institutions who purchase and provide access to the products. While AHIP has always assumed that success could only be measured in terms of our effectiveness in reaching the ultimate beneficiaries of these resources, the researchers; it is very difficult to measure the actual number of users of a printed index or CD-ROM, or the number of actual end users of an on-line search service. While some have argued that the information would reach the most users by simply giving it away at no charge, the inability to measure the *real* end user audience has been an obstacle to making this argument. Like other organizations, the Getty requires accountability from its projects, and as everywhere else there is competition for funds. The research databases have experienced level funding for the past five years, and the Getty would like the operating costs in some cases to decrease substantially.

There are relative merits and opportunities associated with the various methods for distributing research databases. The following model suggests some of the factors to be considered in measuring their relative value.

1. Information Quality - *the quality of the information in terms of the searching functionality that is supported.* Printed books require look-ups in each quarterly or annual volume, while a CD-ROM or on-line service supports searches across the entire database. In a printed book the complete bibliographic information exists only once in a volume, while a CD-ROM or on-line service provides access directly to the complete search results. Searching databases at AHIP's web site supports only keyword searching, while CD-ROM's and on-line services support field-specific searching with index lists. Web searching is stateless, while CD's and on-line search services are interactive, that is, search results can be combined, sorted, and saved.
2. Number of Accesses - *the number of accesses that can be physically supported as well as the total number of users.* Printed books and CD-ROM's are essentially single user products. While it is possible to network our CD-ROM products or copy them to a network hard disk, this is done infrequently for the purpose of providing simultaneous access. Commercial services while capable of supporting many users at the same time, have relatively few users. Non-profit services such as subscriptions to the RLIN CitaDel service provide desk top access through campus-wide information services, for example, to every desk top at every University of California campus. Accessible at no charge through any Internet provider, the AHIP web site has the largest measurable number of database accesses (over 5,000 per month).
3. Royalties Received - *the amount of royalties earned as a function of the number of sales or accesses and the royalty rates.* Royalties from books and CD's are relatively high but the number of units sold is small since these are marketed principally to institutions rather than individuals. Since the manufacturing costs of CD-ROM's is low, there is the potential to realize higher royalties if sales are equal to printed volumes, (particularly with BHA which is self-published). The highest royalty rates are from the commercial on-line search services, but AHIP has generally foregone royalties in order to keep the access charges as low as possible. The Avery Index realizes the greatest royalties from RLG CitaDel searching. Using the same strategy as with commercial services, AHIP has agreed to forego RLG royalties for BHA for 3 years in order subsidize access at a lower cost. There are no royalties from Web access.

4. User Charges - *the amount charged to purchase or access a resource (a negative value)*. Print and CD-ROM publications are priced for institutional purchasers at approximately \$1,000 initially, plus \$500 per year for update subscriptions. Commercial services are the most costly to use at approximately \$50 or more per hour, plus charges for each citation that is viewed or printed. Non-profit search subscriptions cost approximately \$1,200 per year for the first unlimited access logon and \$750 per year for each additional logon. Rates for large consortia are negotiated individually. Web access is as inexpensive as an Internet connection plus telecommunications charges.
5. Producer Subsidy - *the amount of subsidy provided by the Getty (aside from the content development costs)*. Printed volumes are expensive to produce, market, and distribute; CD-ROM's less so. Data for commercial and non-profit services is inexpensive to produce but royalties are generally foregone to subsidize access. Web access is totally subsidized and potentially undercuts sales through other distribution products.
6. User Information Contribution - *the opportunity for users to contribute to the selection, description, and indexing of relevant research materials*. Print, CD, and commercial services are one-way communications channels. Non-profit and Web-based services can be enabled for two-way communications.

Each of these factors has been given an approximate value (from *low* to *high*) in figure 2 below. This subjective analysis indicates that the non-profit distribution method provides the best mode of overall access. AHIP concluded an agreement with the Research Libraries Group in March to foster broader information access and contribution by the international cultural heritage community. While this analysis supports this decision, we will monitor the results of the AHIP/RLG partnership to determine how best to extend access to research databases in the future.

	<i>User</i>						
	<i>Information Quality</i>	<i>Number of Accesses</i>	<i>Royalties Earned</i>	<i>User Charges</i>	<i>Producer Subsidy</i>	<i>Information Contribution</i>	<i>Total Rating</i>
<i>Print</i>	Medium	Low	Medium	(Medium)	High	None	Low
<i>CD-ROM</i>	High	Low	High	(Medium)	Medium	None	Low
<i>Commercial</i>	High	Low	Low	(High)	Medium	None	Low
<i>Non-Profit</i>	High	High	Medium	(Medium)	Medium	Medium	Medium
<i>Web</i>	Low	High	None	(Low)	High	Medium	Low

*Figure 2 - Comparison of Distribution Models.*

**Notes**

1. Woo, Janice. „The Online Avery Index End-User Pilot Project: Final Report.“ Information technology and Libraries (September 1988): pp.223-229.
2. Bates, Marcia J. „The Design of Databases and other Information Resources for Humanities Scholars: the Getty On-line Searching Project Report Number 4.“ On-line & CDROM Review 18 (1994): pp.331-340.
3. Bates, Marcia J. „Document Familiarity in relation to Relevance, Information Retrieval Theory, and Bradford’s Law: the Getty On-line Searching Project Report Number 5.“ (manuscript under review).
4. Bates, Marcia J., Wilde, Deborah N., and Siegfried, Susan. „An Analysis of Search terminology Used by Humanities Scholars: the Getty On-line Searching Project Report Number 1.“ The Library Quarterly 63 (January 1993): pp.1-39.
5. Bates, Marcia J., Wilde, Deborah N., and Siegfried, Susan. „Research Practices of Humanities Scholars in an On-line Environment: the Getty On-line Searching Project Report Number 3.“ LISR 17 (1995): pp.5-40.
6. Siegfried, Susan, Bates, Marcia J., and Wilde, Deborah N. „A Profile of End-User Searching Behavior by Humanities Scholars: the Getty On-line Searching Project Report Number 2.“ Journal of the American Society for Information Science 44 (June 1993): pp.19-291.

7. Saracevic, Tefko, and Kantor, Paul. „A Study of Information Seeking and Retrieving. II. Users, Questions, and Effectiveness.“ Journal of the American Society for Information Science 39 (May 1988): pp.177-96.

Joseph A. Busch  
Program Manager, Standards and Research Projects  
Getty Information Institute  
401 Wilshire Blvd.  
Suite 1100  
Santa Monica CA 90401  
USA

Angela Giral  
Avery Architecture & Fine Art Library  
Columbia University  
Morningside Heights  
New York 10027  
USA