DOCUMENT RESUME

ED 046 447 LI 002 464

AUTHOR	King, Donald W.; Brown, A. M.
TITLE	Some Comments on Marketing AIP Information Products and Services.
INSTITUTION	Westat Research, Inc., Rockville, Md.
SPONS AGENCY	American Inst. of Physics, New York, N.Y.
PUB DATE	Jul 70
NOTE	79p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Cost Effectiveness, Costs, *Economics, *Information Services, *Information Systems, *Marketing, Physics, Surveys
IDENTIFIERS	AIP, *American Institute of Physics

ABSTRACT

This study was addressed to marketing considerations for the American Institute of Physics (AIP) information products and services. The general system and its operation in a marketing environment, including promotion, channels of distribution and pricing are covered. Particular emphasis is placed on the cost/demand/price relationship for four information products and services and an attempt is made to develop an approach for allocating fixed costs for these services. Cost/demand and price/demand relationships are estimated or assumed and an optimum allocation is determined, based on net income for six alternative allocation levels for these products. Four recommendations for changes are given and two appendices include a previous marketing research report and reports on site visits made for the study. (AB)

SOME COMMENTS ON MARKETING AIP INFORMATION PRODUCTS AND SERVICES

Prepared for

American Institute of Physics

Submitted by

Westat Research, Inc.
7979 Old Georgetown Road
Bethesda, Maryland 20014

July 1970

TABLE OF CONTENTS

SOME COMMENTS ON MARKETING AIP INFORMATION PRODUCTS AND SERVICES

D. W. King
A. M. Brown

I. INTRODUCTION

The American Institute of Physics has designed an integrated system to enhance awareness of and access to physics information. This system (NISPA) will ultimately incorporate a wide range of new information products and services, including, possibly, repackaging journal articles, availability of articles in microform, a current awareness service, recurring bibliographies, tape sales, a journal index, selective dissemination services, and a retrospective search service. Since these new products and services will be introduced and implemented in a marketlike environment, AIP must carefully consider economic and marketing factors in developing an effective marketing program to maximize the usefulness and availability of these products. This study is addressed to marketing considerations for that program.

The report covers the general NISPA system (past, present, and future) and the way in which this system operates in a marketing environment, including promotion, channels of distribution, and pricing. Particular emphasis is placed on the cost/demand/price relationship for a retrospective search service, Current Physics Titles, tape sales, and recurring bibliographies. An attempt is made to develop an approach for allocating fixed costs, such as those attributable to input, to these four information products and services. Cost/demand and price/demand relationships are estimated or assumed, and an optimum allocation is determined, based on net income for six alternative allocation levels for these products.

Additionally, visits were made to several other national information systems in order to determine if any areas had been overlooked in the design

of NISPA. We found no important gaps in conceptual design or in implementation plans. However, it is recommended that AIP give some consideration to the following:

1. One weakness in the transfer of physics journal articles from author to user results from the fact that articles are available from sources other than the journal itself. The author has reprints, but it may be difficult to locate him after a period of time or, if his supply is exhausted, there may be few recourses for obtaining a copy. A reprint service is expensive and difficult to administer. If AIP prefers not to provide such a service, it is suggested that they, at least, develop a referral service to organizations that do provide reprints.
2. AIP presently has a current awareness service, Current Physics Titles, that is near implementation. Plans now call for this publication to be issued in a single volume and in four sections. We feel that a further subdivision, by subject classification, should be seriously considered in order to reduce the cost to a more reasonable level and to facilitate use by individual physicists. A successful precedent for both format and price of such a publication has been set by the Clearinghouse for Federal Scientific and Technical Information.
3. An important consideration in financing a self-sustaining system is cash flow. AIP should attempt to place as many information products and services under subscription as possible, thereby substantially increasing working capital for operating the system. One way to do this is to provide the physicist with the opportunity to allocate his society registration fee to any information product or service that he desires.

4. One area in which, unfortunately, little research has been done is article format. The format could be modified to facilitate identification, screening, and assimilation of the full-text information.

Appendix A includes a report submitted previously to AIP on marketing research related to Current Physics Titles. Individual reports are given in Appendix B for the visits mentioned above.

II. SYSTEMS CONCEPTS

A. Basic Functions

AIP's information system is designed to accommodate the transfer of information in the form of journal articles and to provide information products and services that enhance this transfer. Basically, the system involves the processes necessary to make its users aware of relevant technical information in the field of physics and to provide the means for obtaining this information.

Six basic functions govern the transfer of articles from author to user. These functions include composition, reproduction, acquisition and storage, identification and location, presentation, and assimilation. They are defined as follows:

1. Composition -- Preparation of an article including writing and editing.
2. Reproduction -- Typing, printing, or taping of an article.
3. Acquisition and storage -- Acquisition, maintenance, and preservation of copies of an article.
4. Identification and location -- Determination of the identity and location of articles to be used.
5. Presentation -- Physically turning over a copy of an article to a user.
6. Assimilation -- Comprehension by the user of the information in an article.

Presumably an article will be put to use following transfer from author to user, and research may be forthcoming, resulting in a new article. A schema of the flow of article transfer from author to user is given in Figure 1. The order presented has no particular significance, but each completed transfer will normally involve all of the six basic functions.

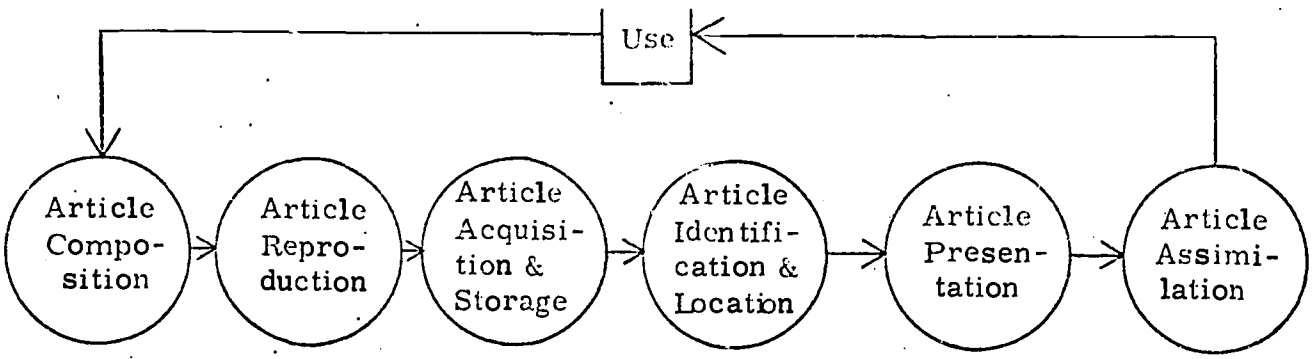


Figure 1. Schema depicting the functions involved in transfer of physics articles from authors to users

These functions are performed at three levels: a central level (the role played by AIP), a local level (such as company, university, or government agency library), and the individual physicist himself. All six functions can be performed at any of the three levels For example, reproduction of articles normally takes place at the central level but could also be performed at a library through copying techniques or by the individual. Also, completed transfer of articles from authors to users may involve scores of possible channels derived from the various functions and levels.

The transfer of journal articles is complicated by certain constraints of time and space. Time presents a two-fold problem. Often the physicist requires an article as soon after discovery as possible. However, composition, reproduction, and other functions may delay transfer of the article so that he does not have access to it soon enough. Another problem associated with time is that certain information may be needed immediately while other may not be needed until some time in the future, making identification and location difficult. Space also presents a dual problem because the over 40,000 potential physics information users are not only widely scattered geographically but highly mobile as well.

B. The AIP System

Because of the complexities of time and space requirements and of the user market, a number of processes have been evolved to facilitate the transfer of physics journal articles. In the past, AIP information products and services have been provided through the processes given in Figure 2.

Past AIP Information Products and Services

Composition	Reproduction	Acq. & Storage	Identification	Presentatio
Individual Author	Local Inst. Copy/Jnl.	Individual Journal	Individual Journal	Other (Indiv., etc
Central (AIP) Editor	Central (AIP) Journal	Local Inst. Journal	Local Inst. Phys. Abst.	User Journal
		Author Reprints	Local Reference	Local Inst. Journal
			Other (Individual, etc.)	Author Reprints

Present or Projected AIP Information Products and Services

Individual Author	Local Inst. Copy/Jnl.	User Journal	Individual	Other (Indiv., et
Central (AIP) Editor	XMicrofilm	Local Inst. Journal	XBibliographies	User Journal
	Central (AIP) Journal Repackaging	XMicrofilm	XCurrent Phys. Titles	Local Inst. Journal
		Author Reprint	XSNI Journal	XMicrofilm read printer
			Local Inst. Bibliographies	Author Reprints
			XJournal index	
			XTapes (SPIN)	
			XCurrent Phys. Titles	
			XSNI Phys. Abst.	
			Central (AIP)	
			XRetrospective search	

Alternatives Being Considered by Others

Author alone	Other forms	Central	SDI Full-text	Central
Reviews	Informal Com.	Reprints		Preprints
	Other modes (Audio-visual)	Preprints		Reprints

Figure 2. Past, present, and alternative AIP information transfer processes

Composition has been performed by the individual author in conjunction with the central (AIP) editor. Reproduction has been primarily by printing in journals, with some copy reproduction performed by local institutional libraries. Acquisition and storage have been performed at all three levels. Authors obtain reprints of their individual articles and store them, to be made available on request; local institutions and individual physicists purchase a journal by subscription and store it on their shelves. Identification has been through the individual identifying articles from his journal subscription, from citations in other articles, and from professional meetings. Institutions provide identification services through Physics Abstracts and their local reference services. Presentation, for the most part, has involved the user's obtaining an article from his own or library bookshelves, from authors by ordering reprints, or from some other source, such as another individual.

To facilitate this transfer of articles, AIP has designed a new system, NISPA, that will provide a number of information products and services to individual physicists directly or to local institutions that serve them. These new information products and services are also shown in Figure 2. New reproduction processes include journal articles in microform. Under consideration is the reproduction of articles in separates so that they may be sorted and packaged to suit the individual's needs. New acquisition and storage processes involve primarily storage of microfilm in local institutions.

Most of the new products and services are aimed at improving the identification function at all three levels. On the individual level, these include bibliographies, a current awareness publication (Current Physics Titles), and selective notification of information (SNI). In addition to these three products, local institutions will be provided with journal index tapes (SPIN) and magnetic tapes for reference searching. Centrally, AIP will also provide a retrospective search service that can be used by both individuals and institutions. With regard to the presentation function, the only significant difference between the old and new systems will be

microfilm viewing and, perhaps, reproduction at the local level from microfilm.

C. Other Systems

The diverse processes being implemented or experimented with by other central information systems all tend to one basic goal -- provide rapid access to relevant information. If a physicist's information needs are met quickly, he will be motivated to use the system repeatedly. Scores of processes could be discussed, but only a few, which may be candidates for future adaptation to the AIP system, will be mentioned here.

A major trend has been toward increasing centralization of acquisition, storage, and presentation. The services provided by the National Lending Library, the Center for Research Libraries, and the Clearinghouse for Federal Scientific and Technical Information are prime examples. Although centralization helps reduce costs at the local level, it also often increases the time lag between identification and presentation. However, the success of CFSTI's efforts indicates that a time lag of as much as three weeks might still be acceptable to users.

Centralization of the acquisition and storage function makes possible the establishment of a central depository of article copies as a backup to the announcement journal. The announcement journal itself can take various forms, listing merely citations or citations with a brief indication of content or citations accompanied by informative abstracts.

Among those using the announcement journal with depository, the group that has gone farthest in its development is the Society of Automotive Engineers. Their service includes a central file of article copies coupled with announcement devices and a modified SDI system. Although the cost per paper is considerably higher than with journal publication, the price, with respect to higher unit page price, still balances out for the user because he is only paying for papers that are relevant to his interests. Furthermore, the cost in his own time becomes less, thus increasing his satisfaction.

SAE has also found that the time lag is much less than it would be with a conventional journal system, the total publication cycle taking about six weeks.[1]

The American Chemical Society publishes a "Research Results" section in its journals, announcing papers submitted and under consideration for publication. Requests for these manuscripts can be processed within 24 hours after receipt. The society and its users consider the system a success,[2] with an average of 250 orders per month for about 300 manuscripts, or almost ten copies per item. ACS also provides a service for ordering copies of individual papers.

The American Psychological Association has a similar system, with a "Manuscripts Accepted for Publication" section in five of its journals. It differs from that of ACS in that only articles accepted for publication are listed, abstracts are not included, and reprint exchange takes place directly between author and user.

The Institute of Electrical and Electronics Engineers, Computer Group, features as part of Computer Group News bibliographic data with abstracts for papers submitted and under consideration for publication. About 100 orders are placed with IEEE per month.

Another approach is that of Academic Press in its publication Communications in Behavioral Biology, which is published in two parts. Part A contains original papers and is published loose-leaf; Part B contains abstracts of all the papers in Part A, plus abstracts of relevant papers accepted for future publication in the journals of the cooperating societies, including the APA, American Pharmacological Society, American Physiological Society, and the EEG Society.

1 F.W. Lancaster and A.M. Brown, "Conceptual Alternatives to the Scientific Journal" (Bethesda, Md.: Westat Research, August 1969), p. 19.

2 Ibid., p. 20.

Another aid to the identification function is the use of SDI systems, which is becoming increasingly widespread and takes many forms. AIP is already considering the implementation of a selective notification of information system involving bibliographic citation with author address, so that copies of most articles may be obtained from the authors. Among the many types of SDI systems are those for the dissemination of full text which have been developed by the American Mathematical Society, APA, and CFSTI. AMS's Mathematical Offprint Service is perhaps the most sophisticated of the three. In the MOS system, when a high-level match occurs between user and document profiles, an offprint of the article is sent. With a low-level match, only a bibliographic citation is sent to the user.

With APA's proposed system, the user can obtain a monthly list of abstracts and request papers from APA, or he can receive full copies of all papers accepted in one of four subject areas--a sort of SDI with broad user profiles. The APA system includes preprints; a paper accepted for distribution in the system as a preprint may later be rejected for publication due to the more stringent refereeing of the journals themselves. CFSTI provides selective dissemination of technical reports in microfiche.

Another device to speed assimilation is rapid scanning. This would involve, among other things, the inclusion of abstracts and index terms with articles submitted for publication. Michaelson[3] has suggested a change in the format of the articles themselves, the text being organized to emphasize the most important aspects of the work. Even the illustrations would be scaled up or scaled down to indicate their relative importance to the text.

We have briefly presented some of the processes being considered or used by other central information systems. In Appendix B, the results of a modest survey of certain major information systems are given in separate reports.

3 M. B. Michaelson, "Achieving a More Disciplined R&D Literature," Journal of Chemical Documentation 8 (November 1968), pp. 198-201.

III. MARKETING CONSIDERATIONS

A. General

It is clear that AIP's information system resides in a marketlike environment and that all of the economic and marketing implications of this environment must be considered. It is also clear, however, that the distribution and sale of information products and services is not like most marketing environments in that these products and services are interrelated and the functions involved in article transfer may be performed in many ways. AIP will be faced with a number of decisions concerning marketing of new services and modification of the old. These decisions include questions of pricing, promotion and advertising policies, and channels of distribution; and they must be based on considerations of cost, income, demand, and the effect of the decisions on other components of the system.

The schema in Figure 3 depicts the functions and processes of the AIP system in a marketing environment. It is shown that processes necessary to accomplish the composition, reproduction, acquisition and storage, identification, location, and presentation functions lead to improvements in such things as accessibility, quality, accuracy, speed, and timeliness of article transfer from authors to users. These improvements are made in order to increase user satisfaction, which in turn motivates the physicists to use the system. This motivation, however, is also partially determined by the price one must pay to use the system and by promotion, sales, and advertising procedures. The price the physicists must pay involves not only AIP's charges for its products and services but also what he must pay in his own time. For example, if a retrospective search results in 5,000 identified titles, he is not likely to be satisfied since he must pay such a high price in his own time to screen out those documents which do not interest him.

In order to make decisions concerning marketing factors, AIP must design and implement an internal costing system, an example of which is given by Helmkamp.[4] This system must be able to identify unit costs that

[4] John G. Helmkamp, Managerial Cost Accounting for a Technical Information Center. (Bloomington, Ind.: Indiana University, 1968).

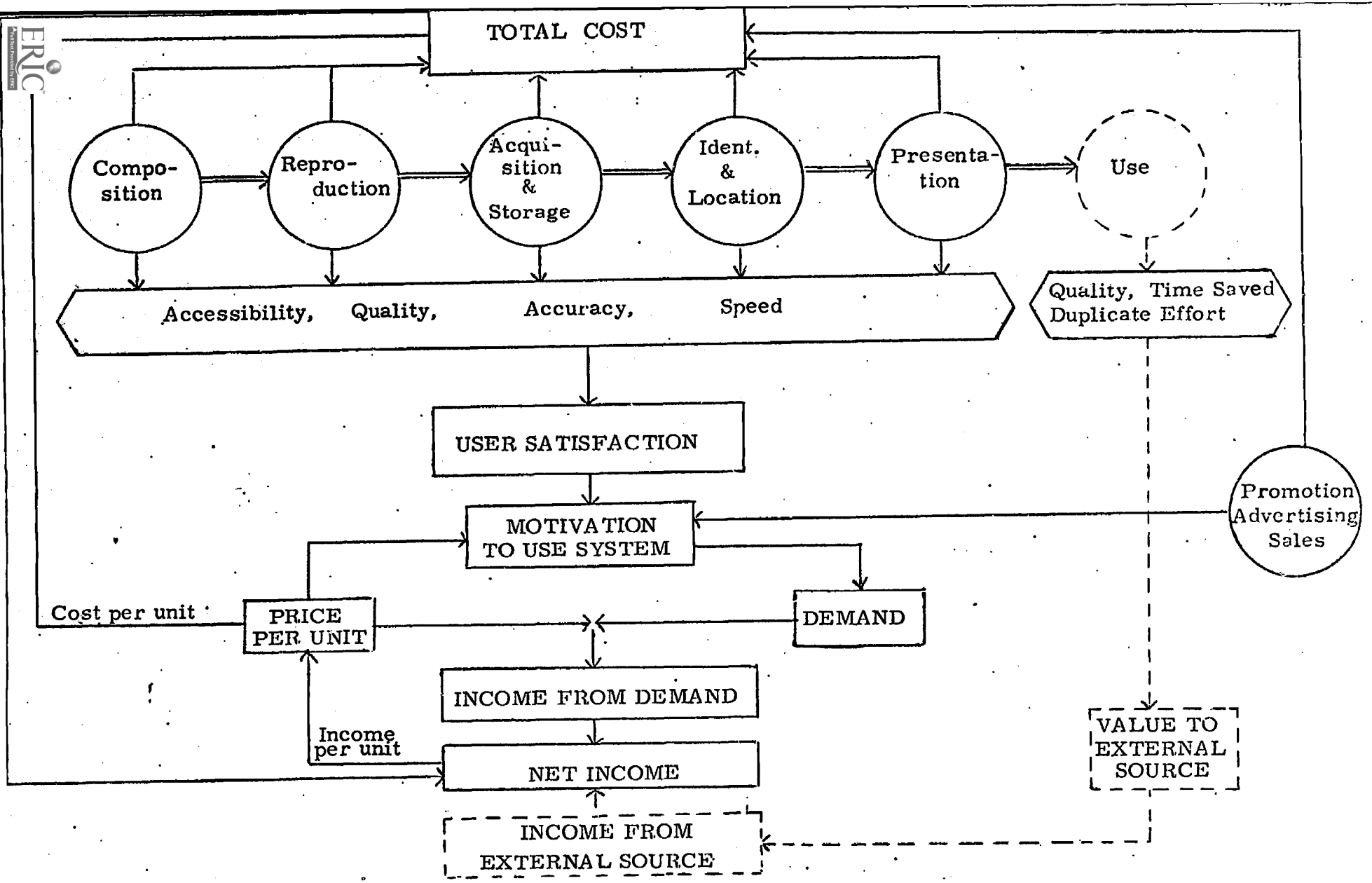


Figure 3. Relationships among functions and measures when users are not a part of the system being evaluated

can be subdivided into fixed and variable costs. Information products and services typically have a high fixed cost and relatively low marginal cost. For example, the fixed cost for producing a journal article may be as illustrated in Figure 4.

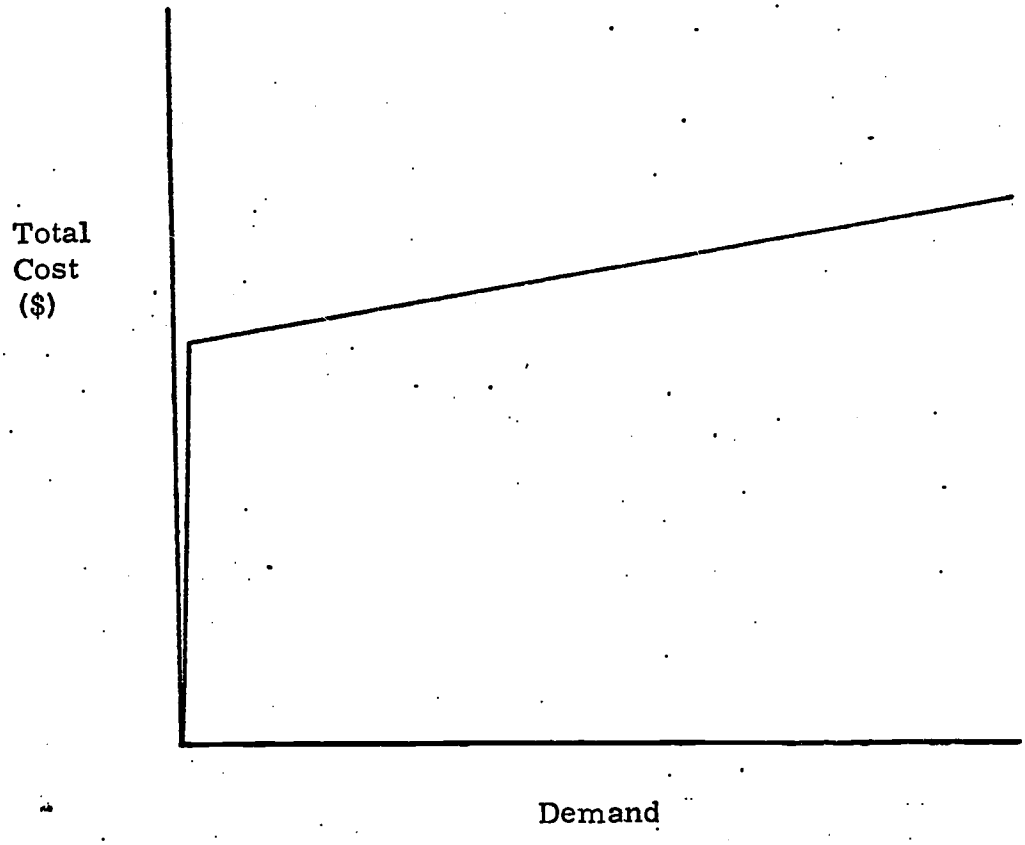


Figure 4. Typical cost curve for information products and services

When one plots the marginal cost against quantity or usage, the curve drops as shown in Figure 5.

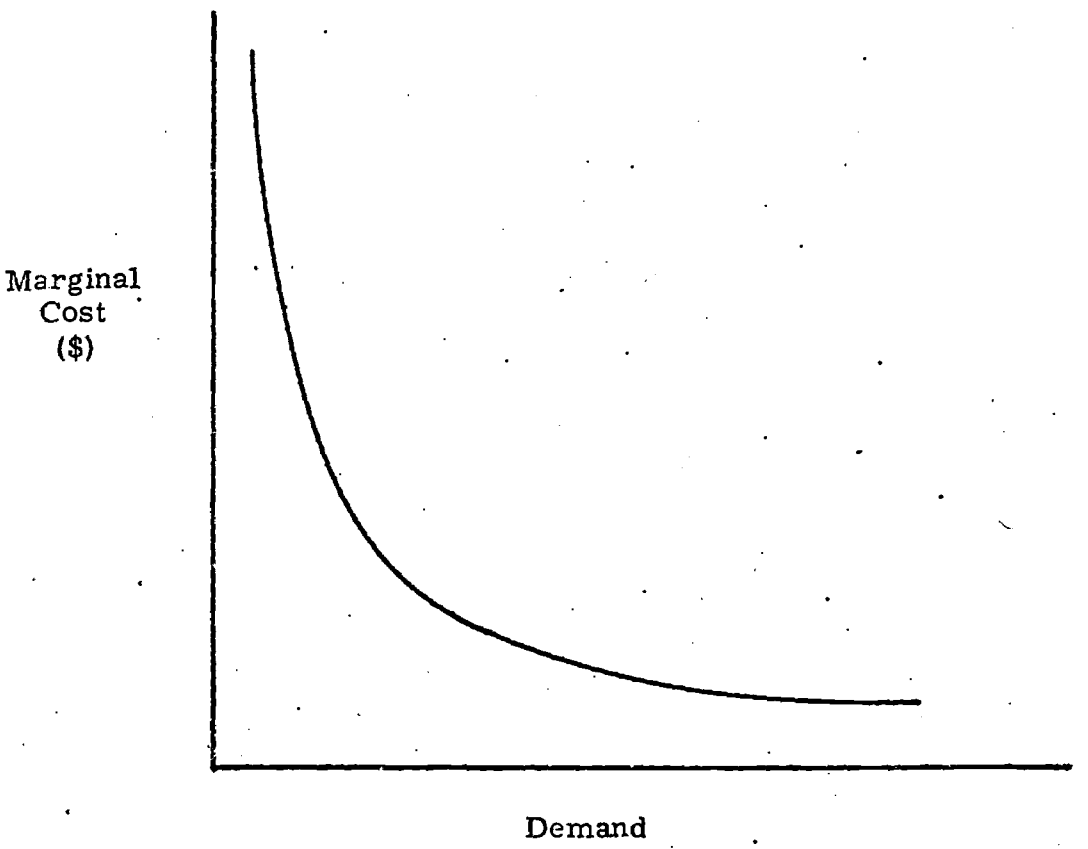


Figure 5. Typical marginal cost curve for information products and services

It is important to establish marginal cost over a likely range of demand for each of the products and services.

AIP's cost accounting system must also be able to identify direct and indirect cost, where the indirect cost involves such items as administration and overhead. Furthermore, it is necessary to isolate indirect costs, such as the preparation of magnetic tapes for computer controlled photocomposition, so that these costs can be allocated to all of the derivative products that will come from these tapes. This cost will be discussed later with individual products and services.

Income is determined by the cost of producing the information products and services and the income derived from demand. The income derived from demand is found by multiplying demand by price per unit. However, since the information products and services provided by AIP yield a direct value to society as a whole and not just to individual users, there is justification for society's partially funding these important operations through such means as the National Science Foundation. This kind of funding can best be accomplished through providing research and developmental capital in order to get a system operational, at which point the system can be self-sustaining. It is clear that a system such as the one envisioned at AIP is not likely to be developed by a private organization since the capital outlay would extend over a long period and the return on investment would probably not accrue in a sufficient time to make the return worthwhile.

As indicated in Figure 3, demand is determined by the influence of the services themselves, promotion, and price. The relative importance of these factors depends largely on the characteristics of the market for the information products and services. There are two classes of market that AIP will serve: individuals and institutions. Each of these two classes has substantially different resources available for purchasing AIP's services. For example, an individual subscriber may be able to spend only $50 to $100 per year, whereas an institution may spend anywhere from $1,000 to $20,000 per year. This means that the two markets may present substantially different demand curves. One would expect the institutional market to have a relatively inelastic (demand not highly sensitive to changes in price) demand curve, as shown in Figure 6.

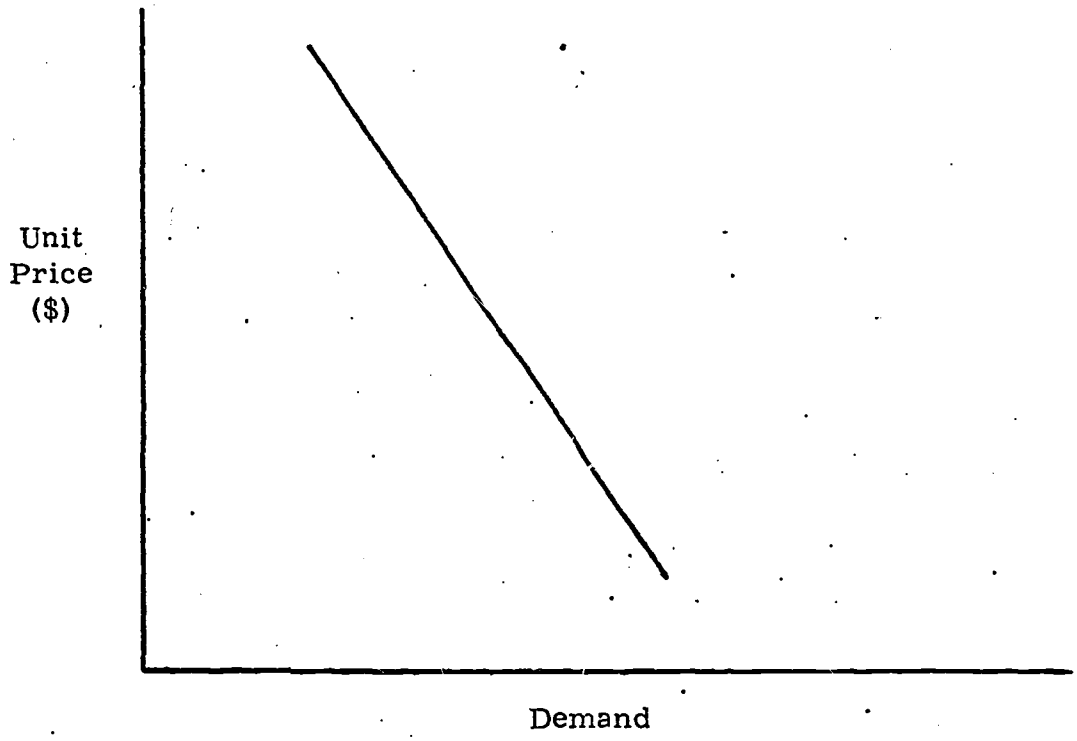


Figure 6. Typical inelastic price/demand curve

On the other hand, the market consisting of individual physicists probably would have an elastic (highly sensitive to price changes) demand curve, as shown in Figure 7.

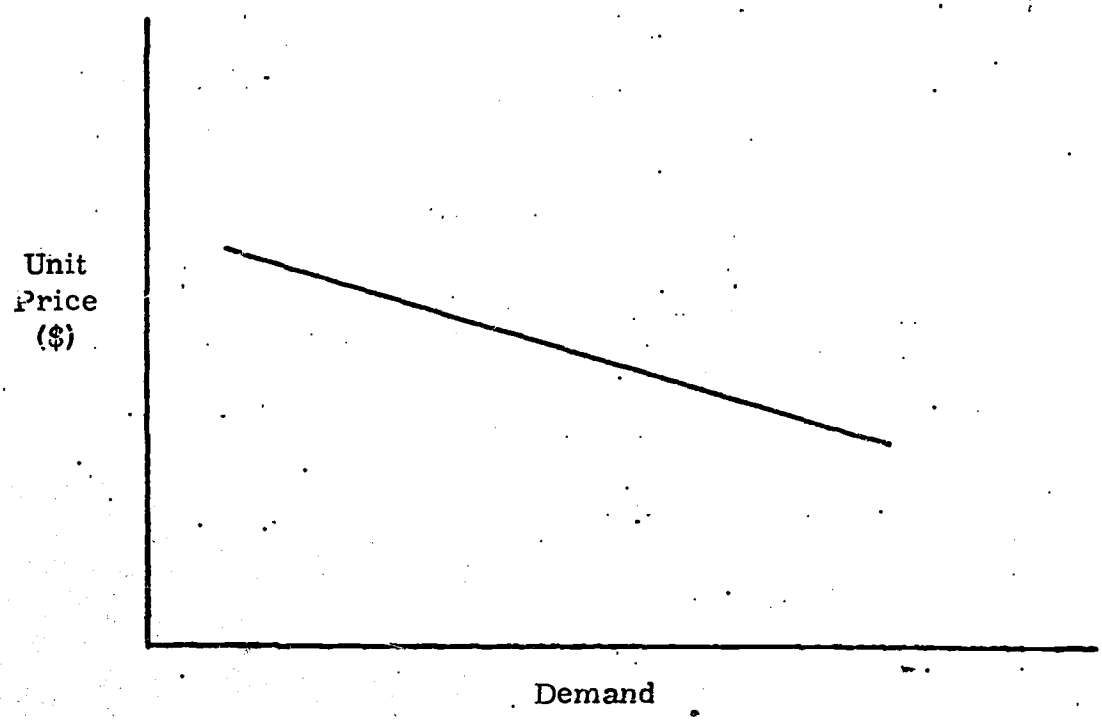


Figure 7. Typical elastic price/demand curve

This means that AIP will have to define carefully the market for each of its products and services and establish a corresponding pricing policy. Fixed and direct cost might be allocated as a component of the price in such a way that a major portion is allocated to those products and services that have an inelastic demand and the remainder to those that have an elastic demand.

One of the weaknesses in marketing information products and services has been a lack of appreciation for promotion, advertising, and sales. One of the reasons for this, among professional societies, is a general feeling that advertising, in particular, is unprofessional. However, it is important that AIP make their new products and services known to those who can benefit from them. On the other hand, AIP should be careful not to create false sales that will not result in repeated use of the system. Another important factor to consider is training in the use of the new services. This training could range all the way from explaining the use of the system through promotion briefs to preparing a training film to be shown at professional meetings and universities.

Finally, AIP must consider the distribution channels involved in the transfer of articles from authors to users. These channels are somewhat difficult to define. For example, articles can be stored in a user's office, in a local library (in the form of hard copy or microfilm), by the author in the form of reprints, or perhaps at a central source such as the Center for Research Libraries. Identification can involve numerous processes, such as Current Physics Titles, Physics Abstracts, the journals themselves, retrospective search, and so on. Almost all combinations of these could occur and probably do occur.

The myriad channels of distribution may be depicted by a matrix consisting of alternative processes and levels for the six basic transfer functions. One could apply this matrix in the form of a stochastic model of market change, determine the relative frequency of each of the distribution channels, and establish the importance of these channels with regard to cost / benefits. Thus, the stochastic model could be used to

diagnose the present system in order to highlight gaps and weaknesses and determine the effect of one product on another. Development of new products and services will undoubtedly affect the demand for and usefulness of other products and services. A highly effective identification system, for example, could result in lower usage of journal subscription channels. The same holds true for local library channels, where it may become easier for the physicist to obtain articles than it would be through his present journal subscription.

Demand is also affected by the amount that an individual or local institution can pay, and this fact must be weighed carefully when each new product or service is designed and marketed. AIP should also consider carefully the possibility of charging a combination fee to physicists and society registrants. For example, instead of limiting the fee to pay for journal publications, the physicist would be encouraged to use this for Current Physics Titles, bibliographies, and even retrospective search.

In order to consider marketing implications in detail, we will limit our further discussion to two services--Current Physics Titles, which probably has an elastic demand, and retrospective searching, which probably has an inelastic demand. Both of these rely on the same magnetic tapes, which constitute a particularly large input cost (approximately $10 per item) due to the sophisticated input requirements, such as those imposed by mathematical equations. The cost equations that we have developed for Current Physics Titles and retrospective search utilize the tape input cost as a fixed input. This cost dominates both equations and will be examined over a wide range of demand in the example given in Section D below.

B. Marketing Current Physics Titles

Current Physics Titles will be an announcement journal covering articles published in AIP's primary journals. The citations will consist of title, author, author's address, and bibliographic information, but no abstract. Page proofs will be produced from photocomposed computer tapes. The publication will appear twice a month; and citations will be current and not cumulative, i.e., an article cited in one issue will not

be cited again two weeks later. Judging from present figures, the journal would handle approximately 24,000 titles per year. Two approaches are under consideration: One would be a single catalog containing all of the subject categories, and the other would be four (or more) editions, each dealing with one subject category. Other alternatives are: whether or not to include the world literature, thereby at least doubling the number of citations; if the journal should appear monthly, bimonthly, semi-monthly, or quarterly; how many items should appear on a page (including size of print and spacing); and how many times each citation will appear. (E.g., since an article can be categorized in a number of different ways, it can appear in several sections. The present average is three sections.)

With regard to marketing, AIP must consider two user subgroups--the individual physicist and the institution, such as a local library. It appears likely that the institutional market may not be a good one for the reference aspect of Current Physics Titles, especially in the single catalog format, since the catalog as now conceived has no reference capabilities other than the very gross identification by subject area. There is no cross-referencing, cumulative index, nor separate identification by author or title. The individual presents a better potential market for the current awareness aspect of Current Physics Titles. However, in this case the single edition would likely prove less popular than the editions broken into categories because the physicist probably will have neither the time nor the patience to search a large catalog and would prefer to subscribe only to those subject categories of interest to him.

A study of CFSTI's U. S. Government Research and Development Reports revealed a problem similar to that faced by AIP.[5] USGRDR was originally designed to do two things: to be a reference tool for librarians and to be a current awareness device for individuals. However, the lack of reference capabilities made it a poor library tool; and its bulkiness

[5] Thomas T. Luginbyhl and Currie S. Downie, "The Poor Man's SDI," n Proceedings of the American Society for Information Science, Vol. 5, (New York: Greenwood Publishing Corporation, 1968).

made it a poor awareness tool, since the user did not have time to wade through the numerous pages, trying to find what he wanted. CFSTI solved the problem by revising USGRDR to make it a better reference tool and by subdividing it by subject categories. Each subject category can then be disseminated, making a better current awareness tool.

Another component of the market environment is the competition. Among AIP's competitors are Current Contents and Physics Abstracts and, to some extent, Chemical Abstracts, Nuclear Science Abstracts, and various NASA publications. AIP must also compete with CFSTI's announcement services, though not necessarily with their document sales.

Finally, attention must be given to the effect that Current Physics Titles will have on other components of the AIP system, particularly other identification modes such as retrospective search. If the current awareness mode is an effective one, the need for retrospective search will be diminished but not eliminated. On the other hand, as librarians and other users become more familiar with the system and with the kinds of information in the file, they will be more competent and better motivated to use the search service.

The cost of producing Current Physics Titles involves such items as the cost of binding each copy, page set-up costs, cover set-up costs, item input costs, subscription fulfillment costs, mailing and handling, and the number of subscribers. A detailed discussion of costs and a cost equation are given in Section D.

Another factor to consider in marketing a product or service is pricing. With regard to the AIP system, pricing is basically a function of the elasticity or inelasticity of the product's demand curve. A current awareness tool such as Current Physics Titles probably would have a relatively elastic demand; that is, it is likely to change substantially with a change in price. This will be discussed more fully in Section D, where a supply-demand curve is plotted at each allocation level along with a price-demand curve.

Promotion and advertising policies must take into account the fact that, in some instances at least, the market can and should be combined. This can be done in two ways: by developing advertising media that can be used with many different groups and by making certain that the media aimed at special groups describe all of the services available. For example, a film might be produced that could be shown to various local professional groups, at professional society meetings, and to university classes. For the special group, such as a particular institution, a single flyer listing all services would reduce mailing costs. AIP should also take advantage of already existing channels to the individual and institutional markets, particularly the specialized journals, for advertising.

Another important aspect of promotion involves training the customers in the use of these products and services. In most cases, the systems will be unfamiliar to the users; therefore, promotion should be aimed, not only at letting them know what the system can do for them, but also at explaining how they can best use it. The universities should not be overlooked in these promotion efforts. A short training film on the use of AIP's information system will familiarize the students with the kinds of products and services available to them in their career activities.

C. Marketing Retrospective Searching Services

Briefly, the retrospective search system being considered by AIP will be an on-line system that searches index terms based on an authority term list and thesaurus. It has not yet been firmly decided if search questions will be answered by letter, telephone, or both; if the output will be screened; or whether a list of titles or abstracts will be sent. For the purposes of the cost analysis below, we will assume that no screening is done and that the user will be sent a listing of abstracts that are available from the abstract input. Physics Abstracts may prove to be an effective reference and searching mode for this system.

With regard to the marketing environment, the user will probably, by necessity, be an institution. Based on the discussion of costs given

below in Section D, even with a yearly demand of 4,000 requests, the cost per search would be approximately $100. This, certainly, is too high for most individual users.

Competition will come primarily from other modes of retrospective searching. A university librarian, for example, might find it less expensive or even less difficult to search manually through back issues of journals than to request the information from a central source. Other information systems will also compete with AIP's searching service. Among these are CFSTI's reference search service; AEC's RESPONSA, which is a search system of literature from Nuclear Science Abstracts; and, in some respects, the EURATOM file and the CIRCOL system at the Foreign Technology Division of Wright-Patterson Air Force Base.

Rather than looking upon these other systems as competitors, however, AIP must consider establishing an interface with them in various ways. One way would be through referral service; that is, when AIP receives a query, it could be returned to the user with an indication that the Clearinghouse, for example, has the search capability and appropriate data base and that the user should contact them. A second way would be to provide combined responses; that is, AIP would respond to the query and then pass it on to CFSTI for their answers as well. Thirdly, AIP could tie into the CFSTI system, conducting searches on their data base from a remote terminal at AIP. (The Foreign Technology Division, with nearly 400,000 foreign articles in the open literature, is also interested in having other groups use their system.) Another way would be to purchase the Clearinghouse tapes, or that portion of them appropriate to the physics community, and search the tapes at AIP.

One other aspect of the marketing environment is the effect that the retrospective search service will have on other components of the AIP system. For the most part, the efficient operation of the other identification processes should reduce the demand for retrospective searching. The reverse, however, is not necessarily true; an efficient searching capability

is not likely to reduce substantially the sales or demand for bibliographies, Current Physics Titles, and the like. On the other hand, the success of the retrospective search capability may produce a demand for a capability not embodied in the AIP system as presently conceived--the handling of hard copies. If AIP does not provide a reprint service, they should at least provide some type of referral service (other than the author, that is). Possibilities would include the Center for Research Libraries, the Clearinghouse in some instances, or even a listing of local libraries that can provide hard copies in some form or other.

Cost considerations are somewhat different for the retrospective search system than for Current Physics Titles. One of the most important is that input costs to the search system can be amortized over, probably, a period of as long as four years because the items input now will still be used. This is not true for the Current Physics Titles input costs, due to their "one-shot" nature.

Allocation of fixed costs will also be different for the two services and, consequently, so will pricing strategy. Since the largely institutional market for the retrospective search probably produces a relatively inelastic price-demand curve, a disproportionate share of fixed costs might be allocated to this service, rather than to a service with a probable elastic curve, such as Current Physics Titles. Section D presents cost equations and a detailed discussion of the various cost and pricing considerations.

The final marketing implication to be considered is promotion and advertising. Here, probably more than with any other product or service of the AIP system, promotion must be geared to training customers in the use of the system. It is very difficult to use a system of this kind, so training might even include a formal training manual such as that developed for the MEDLARS system of the National Library of Medicine.[6] As the training acquaints users with the system and enables them to make better use of it, they will be persuaded to consult it frequently.

---

[6] F. W. Lancaster, The Principles of MEDLARS (Washington, D. C.: Government Printing Office). Report submitted by Westat Research to National Library of Medicine under P. O. No. 467533-9.

D. An Example of Cost and Pricing for Multiple Products

The example given below demonstrates the general relationship among cost, demand, and pricing of AIP's information products and services. In particular, it illustrates the importance of allocating fixed costs optimally among multiple products and services and attempts to establish an optimal strategy from theoretical cost models for retrospective search, a current awareness tool such as Current Physics Titles, magnetic tapes from input, and recurring bibliographies. One of the fixed costs that is especially important to AIP is input cost, especially the cost of the magnetic tapes for photocomposition of bibliographic information ($10 per item input). However, there are a number of bibliographic products and services to which this input cost can be allocated. The question is, How should it be allocated among the four information products and services mentioned above?

Two facets must be considered in allocating costs. One of these is the interaction between cost and elastic or inelastic demand. Generally, the products for the individual market should have an elastic demand, whereas those for the institutional market, such as bibliographies or retrospective search, should have a more or less inelastic demand. In allocating costs, more should be allocated to the products with inelastic demand because their market will not be so sensitive to price changes. For example, if the cost of inputting the magnetic tapes were $10 per abstract and if there were two products, one elastic and one inelastic, to which this cost could be allocated, the allocation should not be $5 to each. A more realistic distribution would perhaps be $7 to the inelastic product and $3 to the elastic. The precise allocation would have to be governed by a study of the demand curve (or, at least, a simulated likely demand curve) for each product in order to determine an optimal pricing system.

The second facet to be considered in allocating costs is amortization. For some of the products and services, set-up costs can be amortized over several years rather than over a single year. For example, the cost of inputting the magnetic tapes for a retrospective search system could be

amortized over a five- or ten-year period because the information may be used for that length of time. On the other hand, amortization would not be feasible for such things as a current awareness service since this product's usefulness (or sales) does not extend over a year.

Another aspect of cost allocation is competition among products, i.e., products in the same system may be competing with each other for markets and in price. An efficient current awareness program, for example, will reduce the necessity for a good retrospective searching program.

A model of the cost of retrospective systems was developed under contract to the American Psychological Association. This model includes the following subsystems:

1. User/system interface
2. Input (full-text versus indexing) and number of items input
3. Search length based on retrieving various levels of recall
4. Search modes
5. Screening processes
6. Method of presentation

For this example we have assumed an on-line system consisting of manual indexing for input, a thesaurus to use for input as well as searching, user requests processed through an intermediary by telephone or in writing, and searches that, on the average, retrieve 80% of the relevant items.[7] No screening is performed on search output, and abstracts of identified documents are sent to the user. We also assume that 25,000 documents are added to the file for search each year and that these will be purged from the files after four years. Thus, the total file will be 100,000 documents, and each year's input can be amortized over a four-year period.

---

[7] The distribution of number of documents necessary to retrieve 80% recall is based on the average observed for a similar type of system.

The generalized cost equation for the retrospective search is as follows:

$$C = C_4 + C_8 + X_1(C_2 + C_3) + X_2\left[C_1 + C_9 + X_3\left(C_5 + \frac{C_6}{T_1} + C_7\right) + X_4 C_{10}\right]$$

where C = total cost per year

X_1 = number of items (abstracts or titles) input per year (25,000)

X_2 = number of searches conducted per year

X_3 = items retrieved per search

C_1 = cost of intermediary per search ($11.25 if search is conducted by letter, $15.00 if by telephone)

C_2 = fixed input costs per item (indexing, keyboarding, other processing) to be allocated among the various services

C_3 = fixed input costs (tape conversion, updating)

C_4 = fixed computer cost per year

C_5 = variable computer cost per item retrieved

C_6 = screening cost per minute

C_7 = computer printing cost per item retrieved

C_8 = fixed cost of screening abstracts ($32,000 per year)

C_9 = fixed mailing cost for titles ($0.002 per item for titles, $10 per item for abstracts)

T_1 = number of minutes to screen (6 titles per minute or 2 abstracts per minute)

Typical costs for a system as described above were established from analysis of a number of similar systems of other professional societies and government.

The following costs are allocated to the four information products mentioned above: indexing $3.50 per item, thesaurus $40,000, and input $10 per item. The percentage of these costs allocated to retrospective

search is amortized over a four-year period. Updating costs at $5,700 and other costs at $0.50 per item are allocated entirely to the retrospective search, with four-year amortization.

The total costs per search are given in Table 1. These costs were found by allocating input costs by 25%, 50%, and 75% to retrospective search. Cost per search for these three levels of allocation is given for 1,000 through 6,000 searches per year in increments of 1,000.

Table 1. Cost per retrospective search over demand of 1000, 2000, 3000, 4000, 5000, and 6000 searches per year with input allocated at 25%, 50%, and 75%

Percent Allocation	1000	2000	3000	4000	5000	6000
25%	$133	$ 90	$76	$69	$65	$61
50%	155	101	83	74	69	65
75%	177	112	90	79	73	69

The number of requests clearly has little bearing on cost-per-search above a demand of 4-5,000. We recognize that these costs may not hold true for all search systems. However, batch processing, on-line indexing, and on-line full-text seem to have cost/demand relationships similar to those shown above.

A crude cost model was derived for the sales of magnetic tapes as a derivative product of the photocomposition tape input. Costs estimated from the model are based on percentage allocation of the tape input, $10,000 other fixed costs, and $100 per annual tape sales of 25,000 items. These costs are given for three levels of allocation (0%, 25%, 50%) in Table 2 over a demand of 10, 25, 50, 75, and 100.

Table 2. Cost per tape sale over demand of 10, 25, 50, 75, and 100 sales per year, with input allocated at 0%, 25%, and 50%

Percent Allocation	Demand 10	25	50	75	100
0%	$ 1,100	$ 500	$ 300	$ 233	$ 200
25%	7,350	3,000	1,550	1,067	825
50%	13,600	5,500	2,800	1,900	1,450

Although these costs have not been considered as carefully as the others, cost/demand curves for tape sales should be approximately as given.

A cost model was also derived for analysis of Current Physics Titles and includes number of items per year, number of times each item appears, number of items per page, frequency of publication, alternative printing policies, and pricing strategy. This cost model is given below.

$$C = C_{12}S + \frac{C_{11}X_3}{X_2} + \frac{C_{10}X_3X_4}{X_1X_2} + C_{13}X_5S + \frac{C_8X_3X_4S}{X_1X_2}$$

$$+ C_9X_5S + C_6X_5S + C_7X_5$$

where

X_1 = item density (i.e., number of items per page)

X_2 = number of sections (1-50)

X_3 = number of items input annually (24,000)

X_4 = number of times each item appears (3)

X_5 = number of issues per year (24)

C_6 = marginal cost of cover ($0.0234)

C_7 = set-up cost of cover ($15)

C_8 = marginal cost of printing each impression ($0.00425)

C_9 = cost of binding each copy ($0.08)

C_{10} = page set-up costs ($5.00 printing, $8.50 photocopy)

C_{11} = input costs per item ($0.30 computer tapes, $0-10 allocated for keyboarding)

C_{12} = subscription fulfillment costs ($1.87 per subscription per year)

C_{13} = copy mailing and handling costs ($0.09 per copy)

S = number of subscribers

Note that C_{11} in this model accounts for the input costs, including the costs of keyboarding magnetic tapes. It is assumed in this example that Current Physics Titles is used as a current awareness tool similar to CAST at the Clearinghouse[8] and that nearly 50 such categories are available and disseminated to users. Table 3 gives the cost for 0% and 25% allocation of the tape input cost for subscription demands of 50, 100, 500, 1,000, and 2,000 demand.

Table 3. Cost per annual subscription for Current Physics Titles over demand of 50, 100, 500, 1000, and 2000 subscribers, with input allocated at 0% and 25%

Percent Allocation	Demand 50	100	500	1000	2000
0%	$19	$12	$6.20	$5.50	$5.10
25%	44	25	8.70	6.75	5.73

These figures are thought to represent accurately AIP's costs for Current Physics Titles. It is noted that the 25% input allocation seems to dominate costs up to 500 annual subscriptions. For that reason, and since the demand for Current Physics Titles may be elastic, only 0% and 25% allocation are considered. Also, it was assumed in the cost calculations that all 50 subject categories will have a similar demand, although this is probably not realistic. However, the cost model can easily accommodate a probable distribution of demand. Furthermore, one can readily determine from the model a break-even number of subscriptions, below which AIP probably would not wish to fulfill.

[8] See the description of CAST in Appendix B.

The same model will be applied to recurring bibliographies, for which a special cover will be used and a total of four categories will be available and disseminated. The cost with 0% and 25% allocation is given in Table 4 for subscription demands of 50, 100, 500, 1,000, and 2,000.

Table 4. Cost per annual subscription of recurring bibliographies over demand of 50, 100, 500, 1000, and 2000 subscribers, with input allocated at 0% and 25%

Percent Allocation	Demand 50	100	500	1000	2000
0%	$194	$101	$26	$17	$12
25%	507	257	57	33	20

These figures were derived from costs for Current Physics Titles (with four sections). Although they may not be highly accurate, they will serve well for this example. Again, the 25% input allocation costs dominate the entire range of demand.

There are two ways in which we can establish optimal allocation to these four information products and services. The first way assumes that we choose a price for each of these products and services such that we maximize net income to AIP. Standard economic theory tells us that this price is the price at which marginal costs and marginal revenue are equal. In order to determine what marginal revenue is, we must assume a demand curve such as shown in Figure 8, with cost curves with 25%, 50%, and 75% input allocation plotted against demand. We find that the marginal cost in each instance is approximately $45 to $50 per search. Assuming the demand shown in Figure 8, we find that the price at which the marginal revenue yields about the same amount is slightly under $100, in which case we should get a demand of 3,000. Thus, we assume $100 to be the optimum price for all three cost curves.

If AIP's interest is not in maximizing the net income but rather in breaking even, we find that the break-even points are at $66, $71, and $79 for allocation of 25%, 50%, and 75%, respectively. The risk associated with each of the two viewpoints is discussed below. The demand curve given in Figure 9

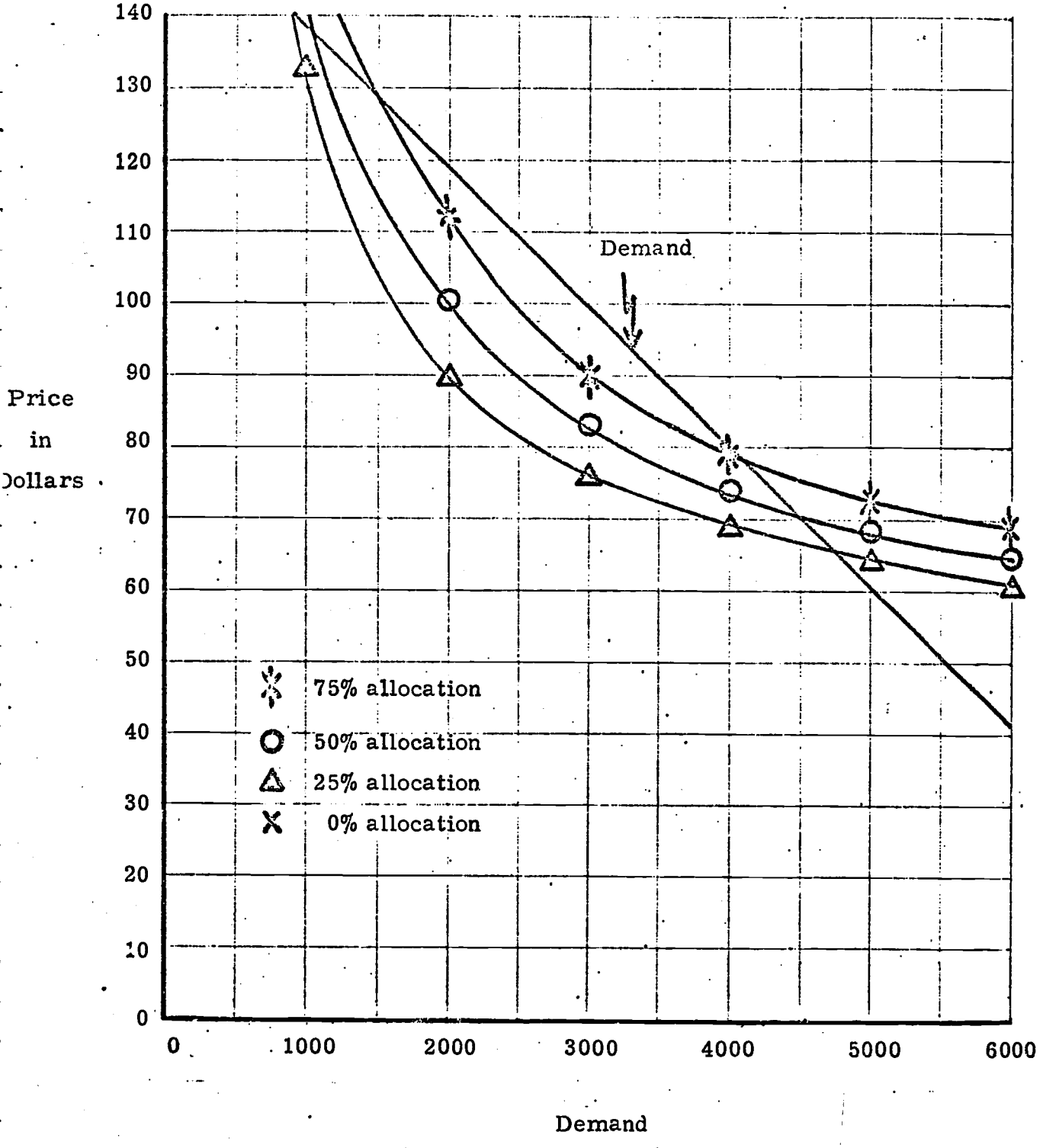


Figure 8. Price/demand and cost/demand (at 25%, 50%, and 75% input allocation) for retrospective search services

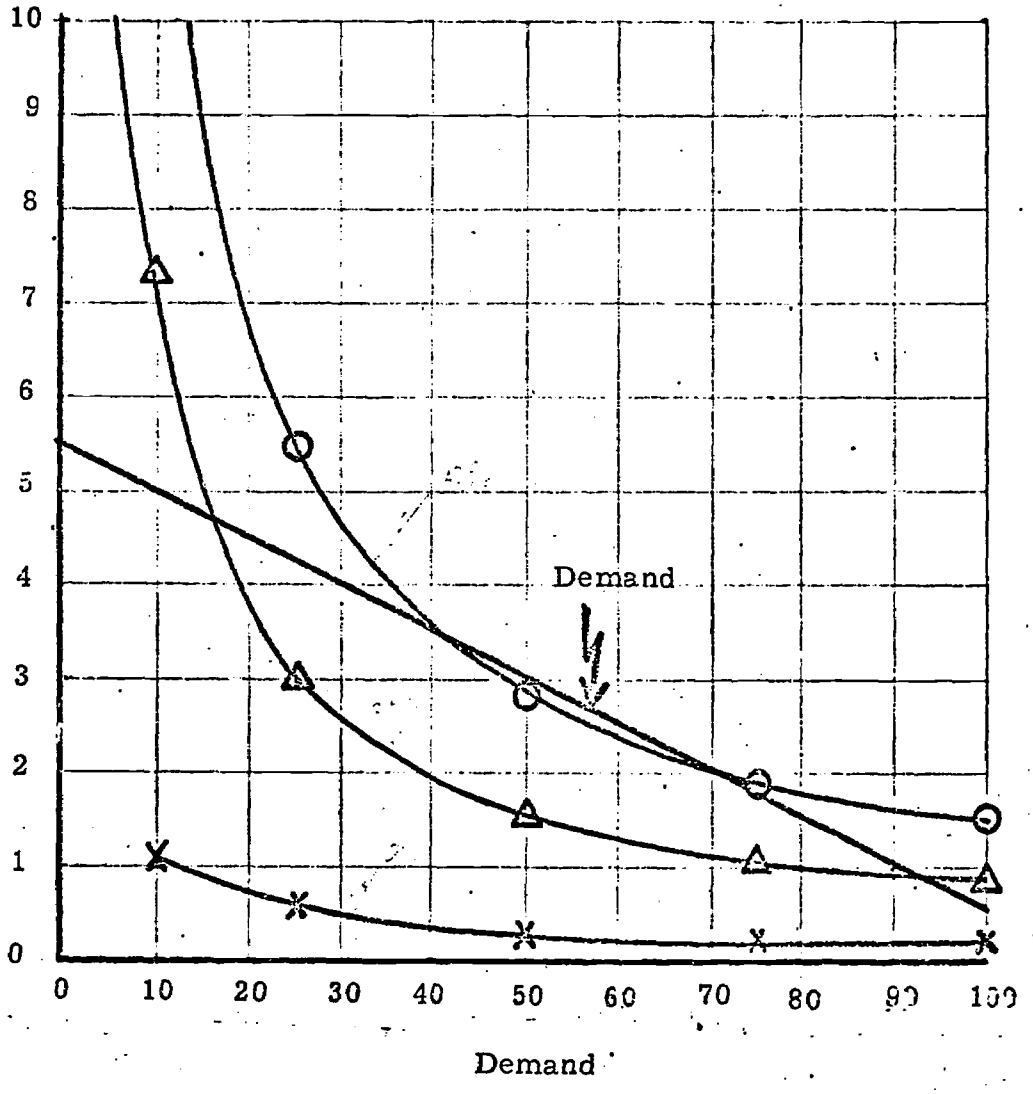


Figure 9. Price/demand and cost/demand (at 0%, 25%, and 50% input allocation) for magnetic tape sales

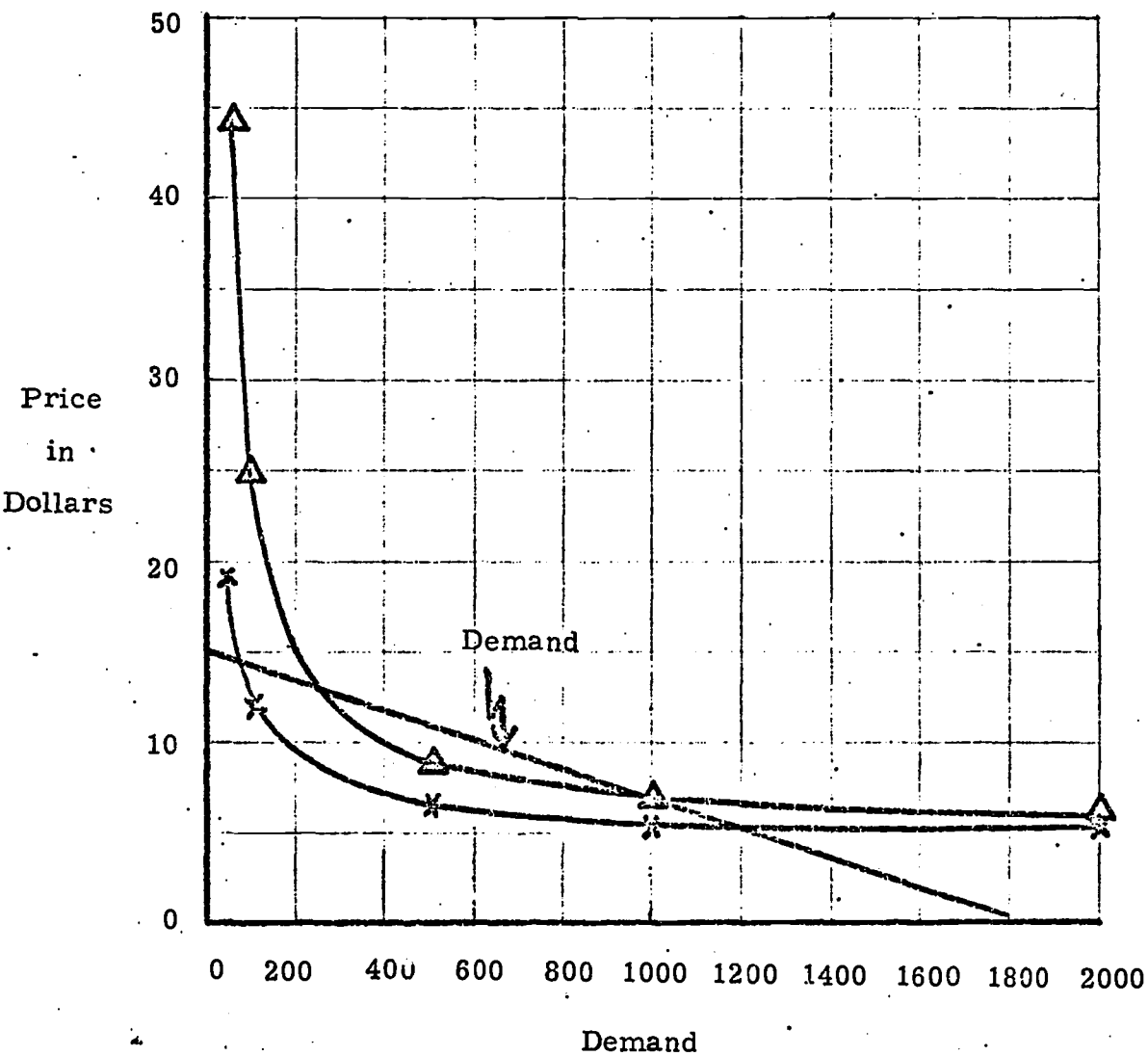


Figure 10. Price/demand and cost/demand (at 0% and 25% input allocation) for Current Physics Titles

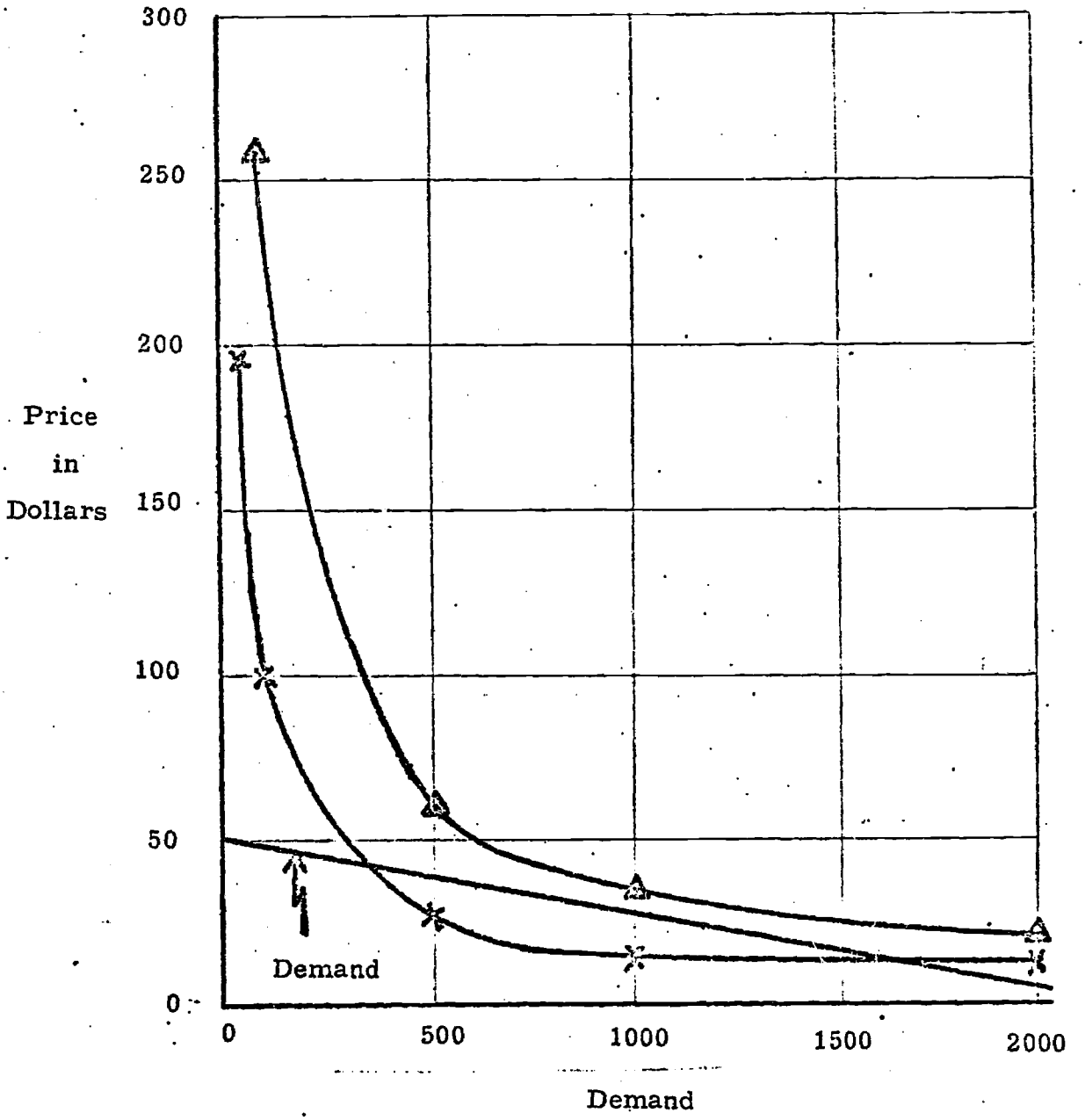


Figure 11. Price/demand and cost/demand (at 0% and 25% input allocation) for recurring bibliographies

assumes a more or less inelastic relationship with price. We have, of course, no way of knowing what the true demand curve is, but will assume the inelastic curve for this example.

Similar curves are plotted for the sale of magnetic tapes. The three curves show cost versus demand at 0%, 25%, and 50% allocation. A rather elastic demand curve is assumed for illustrative purposes. The marginal cost for tapes is found to be in the $1,000 range; the marginal revenue that produces an equivalent amount is calculated to be a price of about $3,000. It is noted, assuming the demand curve above, that the 50% allocation barely yields a positive net income between the range of $2,000 and $3,400.

Figure 10 gives curves for Current Physics Titles with 50 classification sections. Again, a rather elastic demand is assumed since Current Physics Titles may be sent to a number of individual subscribers, who are concerned with price. In this case, the marginal cost ranges from about $4.75 to $5.00, and the price necessary to achieve this marginal revenue is about $9.00 per subscription.

Figure 11 gives the demand and cost curves for recurring bibliographies, where we find that the marginal cost is in the $4.00 to $8.00 range and that a price of $30.00 will achieve an equivalent marginal revenue. The demand curve for recurring bibliographies is considered to be rather elastic since this product will probably not be as popular as some of the others.

Let us assume that we wish to maximize net income and that the prices stated above will hold for the four information products and services. We find that, with six different allocation schemes, we arrive at the widely varying net income shown in Table 5, where the total net income ranges from $176,000 to $307,720 The optimum pricing/cost strategy based on the six allocations in that table is to allocate 100% of the input cost to retrospective search, and none to tape sales, Current Physics Titles or the recurring bibliographies.

Table 5. Net income over various input cost allocations for retrospective search, tape sales, Current Physics Titles, and recurring bibliographies

Percent allocation						
Retrospective search	25	50	50	50	75	100
Tape sales	25	50	25	25	25	0
CPT	25	0	0	25	0	0
Recurring bibliographies	25	0	25	0	0	0
Net income						
Retrospective search	$ 72,000	51,000	51,000	51,000	30,000	9,000
Tape sales	70,000	12,000	70,000	70,000	70,000	134,000
CPT	72,000	126,000	126,000	72,000	126,000	126,000
Recurring bibliographies	-17,600	38,720	-17,600	38,720	38,720	38,720
	$ 196,400	227,720	229,400	231,720	264,720	307,720

However, we should note, as mentioned previously, that one must consider the effect of one product on another. In this case, we may find that the sale of tapes will seriously affect retrospective search demand. This possibility should be carefully considered when making a decision concerning the costing and pricing strategies.

Another criterion for choosing price for cost allocation is based on the risks involved in not choosing the correct price. For example, if we assume 25% allocation to the retrospective search and mistakenly choose a price that is 10% below the break-even price (i.e., instead of pricing at $66, we price at $60, yielding approximately 6,000 demand), we will suffer a loss of nearly $120,000. On the other hand, there is a wide range of prices (from $66 to $140) in which we should have a positive net income. This means that AIP should tend toward a higher price for retrospective searching. It is also noted that, for 75% allocation, the range of profitable prices is $80 to $129, which is still a very loose range; but pricing below the break-even point may yield severe losses. Note also that, if the price is too high, there is a break-even point (on the upper left-hand part of the curve) at which losses

can occur - about $129 for the 75% allocation and about $141 for the 25% allocation. Thus, if the price is higher in either one of these two cases, a loss will also be incurred; but this loss is somewhat less than the loss at the other end of the curve, since a lower demand is involved.

In the other three services with relatively elastic curves, we find that the reverse is true. That is, losses incurred by charging too much are nearly the same as losses incurred by not charging enough. For example, if AIP charges $5,000 at 25% allocation for the magnetic tapes, they would incur a $20,000 loss. On the other hand, if the charge is $500, which is below the break-even point, the severity of the loss is $30,000.

If the shape of the price/demand curve is not known, the best strategy would probably be to choose a price close to the knee of the cost/demand curve, unless there is evidence from other information systems that the price is too high or that the demand necessary to break even at that point is too high. It may be best to price high initially and adjust downward later, if necessary. The reverse may be quite difficult. The cost/demand curves can be very useful in resolving these kinds of problems.

One other decision that can be affected by this kind of analysis concerns the order in which new products are introduced and screening points at which AIP can decide not to continue with a new product, at least in its present form. It is suggested that the least risky products be introduced initially; these are Current Physics Titles and recurring bibliographies. If a risky product such as retrospective searching is not found to be effective from the marketing standpoint, AIP must quickly develop new products to use the expensive input processes or adopt a less costly process.

The example above is not given to suggest specific cost allocations or prices for the AIP information products and services but rather to provide a general framework within which AIP can fit their actual costs and make judgments concerning them.

APPENDIX A

MARKETING RESEARCH FOR CURRENT PHYSICS TITLES

APPENDIX A
MARKETING RESEARCH FOR CURRENT PHYSICS TITLES

AMERICAN INSTITUTE OF PHYSICS

National Physics Information System

NOTES ON SUGGESTED PROCEDURES RELATING TO IMPLEMENTATION

Products

The immediately projected products from the American Institute of Physics National Physics Information System are as follows:

1. Current Titles.
2. Machine-produced bibliographies on particular subjects, probably recurring.
3. An SDI service.
4. Retrospective search on demand.

These services are listed in order of present priority. All will be generated from the same data base. In addition, it is planned that magnetic tapes of the data base will be available for distribution to selected users. Further products and services, needed by the physics community, will be identified at a later date.

Input

Input to the data base is already underway, based primarily on AIP's own journal literature. Plans are also being made to incorporate additional sources, including inputs from Physics Abstracts. Each citation input is identified by a full bibliographic description, and the subject content is expressed by notations selected from a faceted classification scheme. In addition, natural-language descriptors, extracted from the text of the paper, are assigned to give greater specificity. These descriptors are intended primarily as supplementary content indicators but may also be used in searching operations.

The classification scheme is the key to all planned products. It will be used as the basis for retrieving citations in retrospective search, for SDI, and for the recurring bibliographies. It may be used as a basis for organizing Current Titles.

Types of Studies To Be Undertaken

At the present time, we believe that two broad types of study are needed in relation to the projected products and services of the system.

1. For each product, basic decisions have to be made relating to coverage, organization, format, and price. Such decisions will need to be based on accurate cost estimates together with studies of market potential and user preferences. In other words, we are here faced with cost-benefit studies in relation to each product. Because Current Titles appears to be the first service planned, we should begin with such studies applied to this product and later extend to similar studies of the other services to be provided.

2. A study of the classification scheme and the descriptor system as retrieval tools and as indicators of subject content is discussed separately since cost and benefit are both so highly dependent on the adequacy of the information base. Because all of the products depend upon retrieval from the data base, such retrieval being based upon the classification scheme, the effectiveness of the classification will virtually determine the success or failure of the entire program. We therefore recommend that a study be conducted, at this point in time, to assess the capabilities and limitations of the classification scheme as a retrieval device.

These two broad areas of study are discussed in more detail below.

Cost-Benefit Studies On Current Titles

In considering Current Titles, the following system options have to be taken into account:

1. volume of citations to be listed (i.e., coverage),
2. frequency,

3. organization of the listing,
4. contents of the individual unit entry,
5. format, and
6. price.

Decisions concerning each of these options, in turn, must be based on:

1. cost and
2. market potential.

All of these are closely interdependent. The volume will influence format and frequency. Volume, format, frequency, organization, and contents will determine costs of producing and distributing the publication. Price, in turn, is partially determined by cost (to ensure that costs are covered) and income from demand. All of these factors will influence the market potential.

System Options

The first and overriding consideration is that of coverage. The following types of materials could be included:

1. journal articles,
2. technical reports,
3. patents,
4. papers to be presented at forthcoming meetings, and
5. books.

For each of these, estimates of probable volume should be made. Presumably, AIP already has estimates for the majority of these categories. It should be reasonably easy to make the remaining estimates, using inputs from, for example, CFSTI and the U. S. Patent Office.

After considering volume, the next consideration is how these various materials may best be incorporated into the AIP data base and what the unit cost per item input is likely to be. A large part of the journal input will be from AIP's own publications, and presumably much of the remaining journal literature can be acquired through Physics Abstracts. A determination

will have to be made of the unit cost of inputting these items in standard AIP format, including class numbers and descriptors. What proportion of the world's journal literature will be captured through AIP journals and Physics Abstracts? How can the remainder be acquired? Is it worth acquiring?

Similar consideration should be given to the other possible materials -- how can these be acquired and input and at what unit cost. The technical reports relating to physics can be acquired in microfiche form at a reduced cost from CFSTI's new Selective Dissemination of Microfiche. Selection and indexing can then be conducted from these microfiche at AIP. Alternatively, machine-readable tapes containing citations and abstracts of physics reports could be acquired from CFSTI. However, these would be of limited utility because of incompatibility with AIP formats and contents. At some stage, AIP class numbers and descriptors would need to be added. The same problems would apply equally to patent acquisition and input.

The important thing is that fairly accurate estimates be made of volumes and unit costs for input of each type of material. This will allow realistic estimates to be made of the volume of entries to appear in Current Titles. It will also allow preliminary estimates of the unit cost per citation printed and, thus, will establish a price range for the publication.

Frequency

Frequency will be affected by size (volume). A preliminary decision on frequency can be made on the basis of expected volume, production logistics, and the effect of periodicity on cost.

Organization

A number of alternatives should be considered. Possibilities include:

1. Broad subject categories based on the classification scheme.
2. A strict, closely-classified order based on the classification scheme.
3. A keyword arrangement based on the assigned natural-language descriptors.

4. Permut ed titles.

5. A combination of the above.

The most promising arrangements should be incorporated into mock-ups of sample issues. The use of these will be discussed below.

Contents of the Unit Entry

The present plan is to include bibliographic citation, class numbers, and natural-language descriptors. The implication here is that the notations and descriptors will be useful additional content indicators. This has to be tested experimentally (see below). User preferences for contents of unit entry should also be considered in the market survey.

Format

This relates primarily to size, layout, quality of production, and type of binding. Format will depend on the use to which the tool is to be put. If it is purely for current awareness, it can be printed on inexpensive paper, pocket-size for reading on the train and subsequent disposal. If it will be used for retrospective search, and therefore retained for one or more years, different format and quality will be needed. Potential uses and user preferences should be part of the marketing studies on this publication.

Comments on Cost-Benefits

It is clear that each system option must be considered in view of its effect on AIP's costs. However, it is less clear how one must measure the consequence of each option from the standpoint of its benefits. Figure 1 shows the relationships among the system options, costs, and benefits. It shows that an option results in improvement in such things as accessibility, completeness, timeliness, etc., which in turn yields some degree of user satisfaction. The resultant user satisfaction, along with the price and the system's ability to promote and advertise the information product, motivates him to use the system's information product or service. The resulting demand times unit price determines the income from the product. Price, then, is partially determined by what the user is willing to pay and by the net income to AIP.

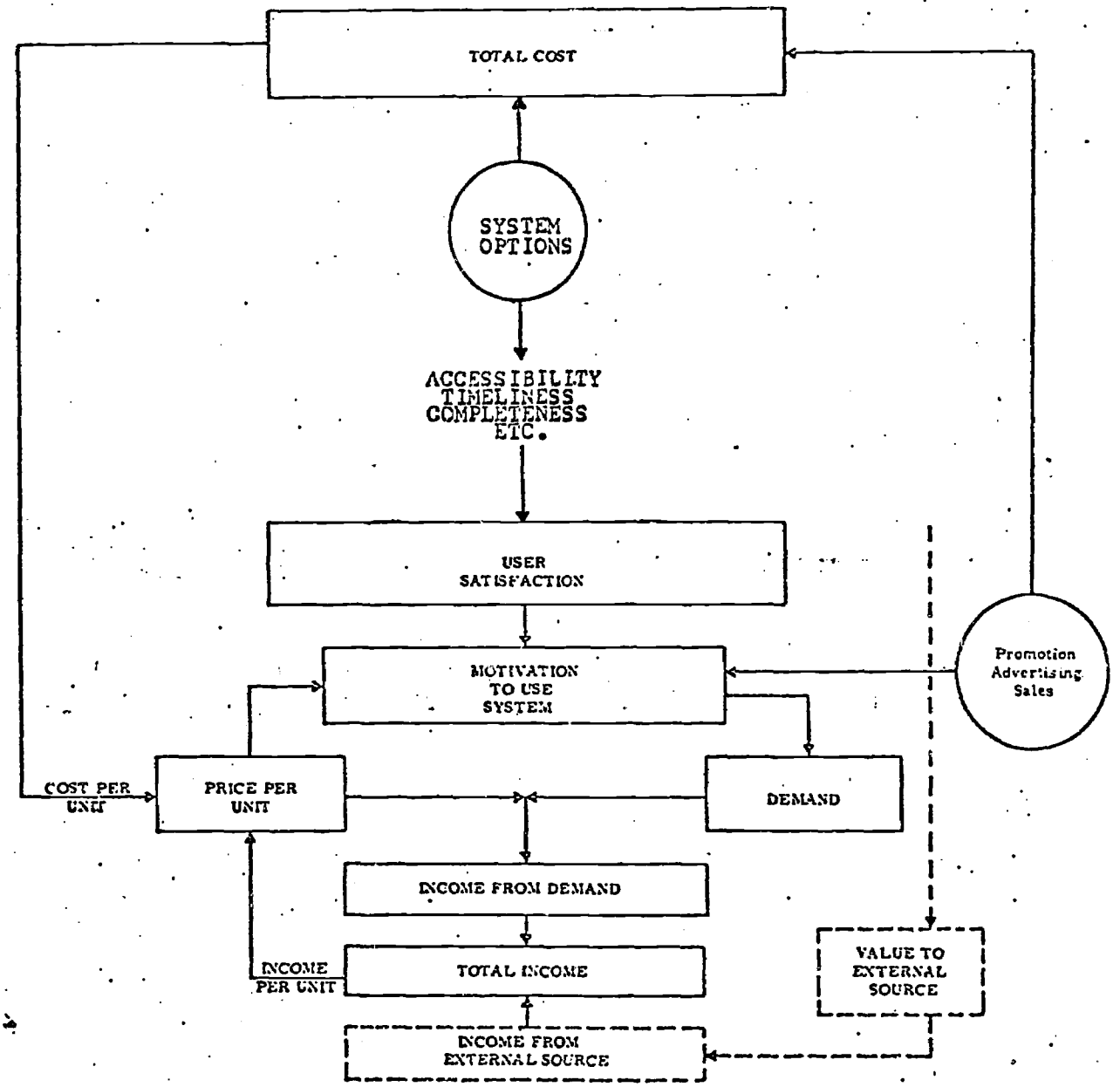


Figure 1. Relationships among system options, cost, and benefits as determined from income.

It is clear that the ultimate consequence of a system alternative must be considered at least in terms of income. To establish the income, one must determine a range of price the user is willing to pay and the market potential over this range. This can only be done by applying marketing research techniques which are not absolute but should yield information for decisions concerning the various system options available. The marketing research should be conducted roughly in the following manner:

1. The cost of the new information product or service should be determined for a specified planning period, say two years. The cost, of course, will vary by a range of possible demand.

2. A price should be estimated over the range of possible demand above, based on a break-even point over the specified planning period.

3. A marketing research survey should be conducted to determine the market potential for new information products and their various options and the user's willingness to pay the range of prices estimated above.

The following sections discuss the marketing studies in more detail.

Preliminary Market Study

We recommend that the preliminary market study be effected by means of in-depth group interviews with selected physicists and librarians in one or more metropolitan areas. In advance, participants would receive two or more mock-ups of portions of sample Current Titles. These samples would incorporate variations in format, contents, and organization. A preliminary questionnaire would be included with the sample issues. The questionnaire and the in-depth interviews would address themselves to the following points:

1. User intentions in relation to the publication. Will it be used for current awareness only or may it be referred to later for retrospective search? This will affect decisions relating to contents and format. For example, cumulations and indexes would be needed if this is to be a search tool as well as an announcement device.

2. User preferences relating to coverage (the implication of a previous AIP study[1] is that physicists will use a current awareness tool if it is comprehensive), format, organization, contents, and frequency.

3. User tolerances to various price thresholds.

On the basis of this preliminary study, specimen formats may be modified. The questionnaire will also be modified on the basis of findings from this pre-test.

Full Market Study

The full market study would be conducted by mailing questionnaires and sample issues (or portions of issues) to a random sample of AIP members, non-AIP members, and libraries. This study will reinforce previous findings as to user preferences and price tolerance. Extrapolations can then be made on the full market potential, and a realistic pricing policy can be established.

Similar studies on market, formats, price structure, etc. should be conducted for the recurring bibliographies, SDI, and demand search, but should await the findings of the the study on Current Titles. Current Titles may be regarded as carrying the bulk of the entire costs of the system. The other services are really by-products that might be offered reasonably cheaply (i.e., Current Titles bears all the input costs, the other products only output costs). Alternatively, input and output costs may be allocated over the entire range of products.

Retrieval System Evaluation

All products will be produced from the same data base and will be dependent on the efficiency of the indexing, the classification scheme, the natural-language descriptors, and the searching strategies.

We therefore recommend a small test to be done as soon as the data base reaches a reasonable size (say 3,000-5,000 citations). This test

[1] M. Slater and S. Keenan, Results of Questionnaire On Current Awareness Methods Used by Physicists Prior to Publication Of "Current Papers in Physics", American Institute of Physics (New York, 1967).

should be designed to tell us as much as possible about the adequacy of the present indexing for retrieval purposes, the capabilities of the classification scheme, searching strategies for using the system, and the utility of the natural-language descriptors as predictors of relevance.

Suggested Test Procedure

Select 100 documents that have been indexed into the data base. These should be selected randomly but might usefully be drawn as two separate random sets: 50 representing the first month's indexing and 50 representing a later month (when presumably the quality of the indexing will have improved).

For the test program it will be necessary to recruit a number of physicists -- say 30 to 50 -- who would be willing to cooperate in the evaluation program. Assume that we have 50 physicists. Each would be given two "source documents" drawn from the files. Preferably the documents should be grouped so that their subject matter is reasonably related to the subject specialty of the physicist. Each physicist would be asked to compose a synthetic question for each source document (i.e., a question for which he regards this source document as a relevant source -- one he would want to see retrieved in response to his request). The rules given to these question compilers would be somewhat as follows:

1. Read or scan the document to arrive at your question.
2. Do not simply re-hash the title.
3. Make the question as realistic as possible. It should represent an information need that you might conceivably have had in the past or may have at some future date.
4. Do not make the question so precise that this is likely to be the only possible document in the literature to be "relevant" to your request.
5. In relation to your question, rate the document on a two-point scale:

A. of major value -- I would not want to miss this in a search on my topic.

B. of minor value -- A relevant citation but there may well be better literature on my topic than this reference.

The questions and source documents will be returned to AIP where they will be examined by physics specialists for "reasonableness". Any doubtful questions will be rejected at this point. For this reason it would be wise to begin with slightly more than 100 source documents.

The questions (but not the source documents) will be given to AIP information staff for preparation of search strategies. Preferably these staff members should know as little as possible about the experiment -- ideally, they should not know that these are synthetic questions based on documents known to be in the file.

Preferably, the search strategies should be compiled at three levels of specificity:

1. a highly specific search designed for high precision,
2. a medium-specificity search, and
3. a broad search designed to get maximum possible retrieval on the subject of the request.

The searcher will vary the specificity of the search (by use of the hierarchies in the classification scheme) and its exhaustivity (by varying the requirements for the number of terms that must co-occur in order to cause retrieval) in order to achieve the three-level strategy outlined above.

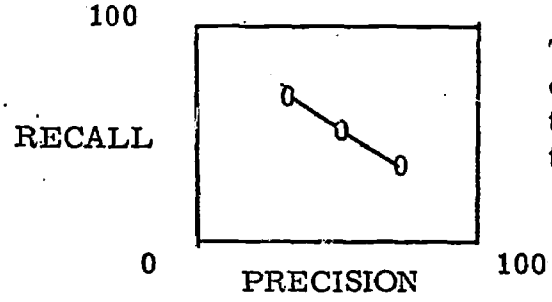
The searches will now be conducted on the machine data base and the results (lists of citations retrieved) will be obtained.

Both the AIP searcher and the requester will be given, in sequence, three surrogates:

1. bibliographic citations only,
2. bibliographic citation + class numbers,
3. bibliographic citation + class numbers + descriptors.

At each stage they will make relevance predictions on the retrieved items in relation to the request. Finally, the requesters will be given the full texts of the documents and will make final relevance assessments on these, judging each document as an A document (as relevant as the source document), a B document, or a C (not relevant) document.

We will now look to see if the original source document was retrieved in each search and at what level it came out -- broad search, intermediate, specific. We will produce an aggregate recall ratio for the 100 searches -- say 85 of the source documents retrieved by the broad strategies, 72 by the intermediate level strategies, and 64 by the specific strategies. We will also derive precision ratios for each search, based on the requester's final relevance assessments, and again at the three levels of specificity. The precision ratios will be averaged over all 100 searches. This will allow us to derive a three-point performance curve to show the recall-precision trade-off at various levels of searching specificity:



These performance characteristics can then be incorporated into models* to determine the relative contribution the input has on the cost/benefits relationships.

We will now do an analysis of the failures uncovered by the test -- all the recall failures and a sample of the precision failures. This analysis will attribute the failures to indexing inadequacy, deficiencies in the classification, or deficienceis in searching strategy, as the case may be. Such an analysis can be expected to tell us a great deal about those parts of the

*Procedural Guide for the Evaluation of Document Retrieval Systems, Westat Research, Inc., (Bethesda, Md., 1968). Prepared for National Science Foundation under Contract NSF-C491.

system that are giving the most problems and will allow us to take corrective action -- before the system design becomes too frozen. Because recurring bibliographies, SDI, and retrospective search will all involve searching strategies based on the classification scheme, the results will be pertinent to all these services.

In addition, we have built in a test of the utility of various surrogates -- citations, citations + class numbers, citations + class numbers + descriptors -- as relevance predictors. Relevance preditions made on these various bases will be compared with final relevance assessments made on the actual documents. We can therefore determine whether, in fact, the extra elements in the citation improve relevance predictability by the searcher or the end user and to what extent. This may have important implications for all products, including Current Titles.

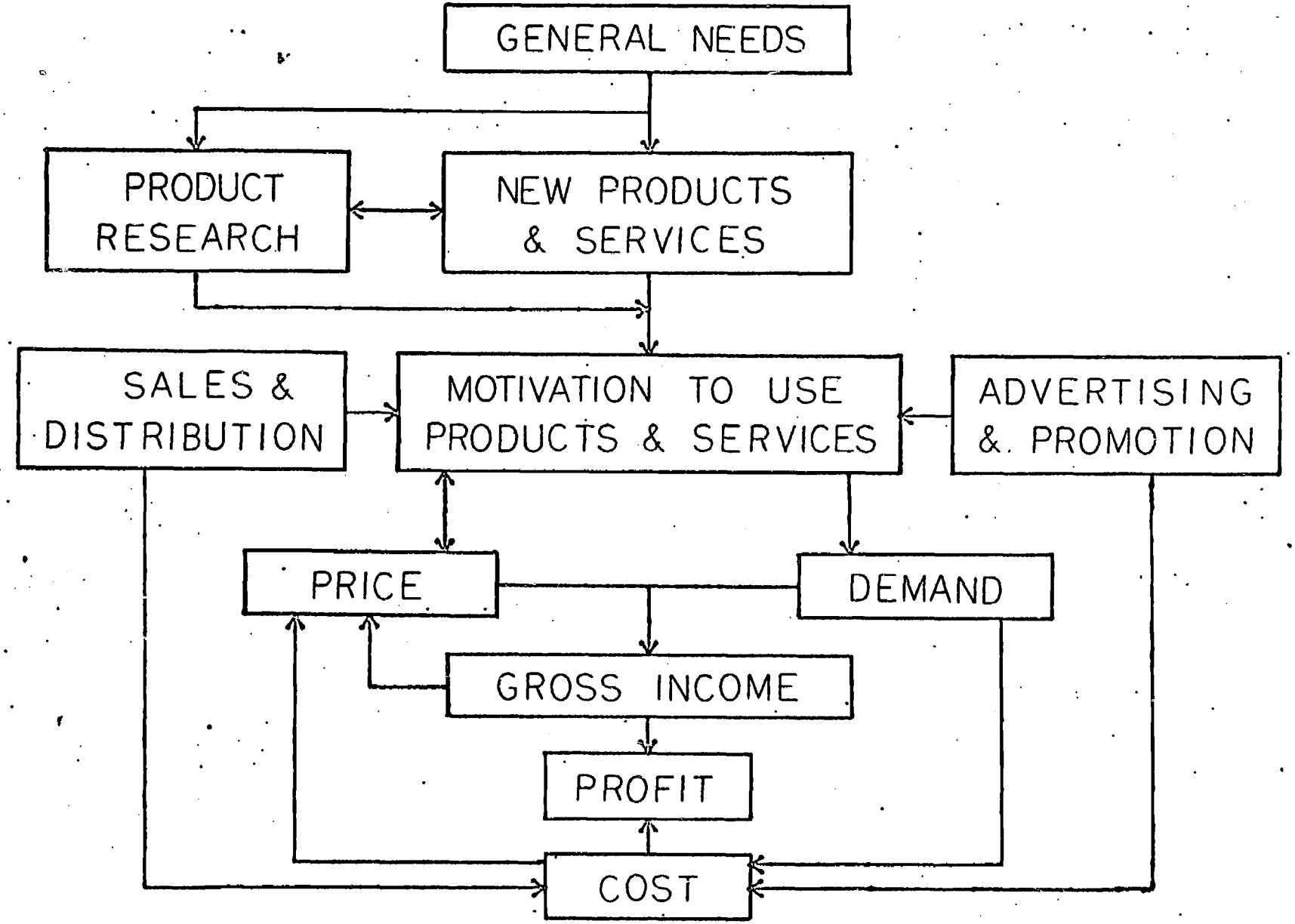
We will then incorporate the entire results into mathematical simulation models (described in Westat's Procedural Guide) that permit us to determine search performance under a variety of simulated search system alternatives. Quality control techniques will also be developed for input such that input accuracy can be established for a satisfactory level of search performance.

PROCEDURES FOR PRODUCT RESEARCH
RELATED TO THE INFORMATION SERVICES PROGRAM OF AIP

The American Institute of Physics is at present involved in the development of a National Physics Information System. This system is designed to produce, from the same data base, a number of different products and services. Many factors are involved and have to be taken into account in the marketing of information products and services. These factors are illustrated in the conceptual model of Figure 1.

The study outlined is directed toward several of these factors as they relate to the proposed publication Current Titles. In particular, the study is intended to establish demand, user preferences, price and related information for the publication. Similar studies may need to be done at a later date for other projected products of the National Physics Information System. However, for the present, we are concentrating on Current Titles as the first output of the system.

Figure 1 depicts several marketing facets that must be considered in the development and marketing of new information products and services. Generally we are interested in the final cost/effectiveness trade-off of new information products and services. Costs to be considered include fixed costs, such as costs of inputting material to the data base and product research, and variable costs. Variable costs will be affected by decisions related to product alternatives found from product research, including scope, packaging, organization and format. As shown in Figure 1, variable costs also depend on sales and distribution alternatives (e.g., frequency of distribution, distribution directly to the scientist versus through a librarian, etc.) and they depend on advertising



CONCEPTUAL MODEL FOR MARKETING OF INFORMATION PRODUCTS AND SERVICES

WESTAT RESEARCH, INC.

FIGURE 1.

and promotion alternatives (e.g., through media such as primary journals, professional meetings, promotion flyers, etc.). Finally, other variable costs, including production and distribution, are dependent on demand.

Income, on the other hand, depends on price and demand where demand is created from user motivation to use the products and services. As shown in Figure 1, this motivation is generated by providing information products and services that satisfy user needs, the price the user is willing to pay for the products and services, sales and distribution procedures, and, finally, by appropriate advertising and promotion. All of these factors are important considerations in development and marketing of new information products and services, and they are thoroughly considered in the marketing research study described below.

The following new product factors and their alternatives must be examined in relation to the proposed publication of Current Titles:

1. Scope. What literature is to be included initially in the publication. This decision will allow an estimate of volume to be made.

2. Packaging. The units in which the publication is to be presented. Alternatives are:

(a) a single announcement journal covering the whole of physics;

(b) separate journals devoted to, say, three or four broad subject areas; or

(c) a large number of separate announcement sheets covering highly specific subject areas of physics.

3. Organization. How the individual entries will be organized in the publication. This is partly dependent on how the publication is packaged. A number of alternatives are possible:

(a) broad subject categories based upon the AIP classification scheme;

(b) a strict, closely-classified order based on the scheme; or

(c) a keyword arrangement based on the assigned natural-language descriptors.

4. Contents of the Unit Entry. The minimum would be full bibliographic citation. However, other items may also be included, such as the AIP classification numbers and natural-language descriptors.

5. Format. Size, layout and type have to be decided upon.

It has already been determined that the publication will at first be restricted to the journal literature. Further types of material (e.g., technical reports) may be included later, depending upon demand and the establishment of viable procedures for capturing the necessary data.

The following distribution factors will also be examined in relation to Current Titles:

1. Frequency. How often should Current Titles be published. This decision is partially based on the desires of the users and the scope of material to be covered.

2. Channels. Should the publication be directed to ultimate users or should it be directed to librarians or other information transfer intermediaries.

Both of these questions are yet to be resolved.

Advertising and promotion will be investigated with regard to:

1. Media. Such media as primary physics journals, promotion flyers and professional meetings will be considered with regard to probable exposure and to costs.

2. Appeals. A number of possible advertising appeals will be considered, such as timeliness, time saving, price, convenience, breadth of coverage, etc.

The final consideration concerning motivation to use the new information product, Current Titles, is price. Price may involve a number of optional strategies that should be considered in the market research study.

The market research study involves the cost associated with alternative new product factors, distribution factors, and advertising and promotion factors, as well as the probable effect of these factors and price on demand. The general research procedure requires determining the general acceptability of various product alternatives. The second step involves estimating costs for each of the alternatives over a range of likely demand for the publication. The third step is to estimate demand related to product factors and price and to establish the best distribution, advertising and promotion strategies to adopt in order to maximize demand at a reasonable cost. From all of this, new product factors, advertising strategies, distribution and price will be determined on the basis of cost/effectiveness trade-off as depicted in Figure 1.

The following steps will be taken in the market research study:

1. The alternative methods of packaging the citations must be considered and accurate estimates made of the costs involved in producing a publication in these various ways over a range of

possible demand. Certain hypotheses relating to distribution and advertising will also have to be made in order that tentative price estimates may be established.

2. Similar preliminary decisions must be made on the contents of the unit entry, and format and organization of the individual issue.

3. A sample issue of Current Titles (mock-up), or a portion of an issue, will be produced for demonstration in user studies.

4. A group interview will be held with selected physicists in the New York area. Participants will be given a copy of the sample issue and will be asked questions relating to their preferences on packaging, format, contents and arrangement and on the price they would be willing to pay for such a publication. The group will be encouraged to discuss freely the publication so that all comments, recommendations and criticisms can be collected. This interview will be recorded for further analysis.

The group interview will be conducted primarily to test certain hypotheses about use of the publication, to assist in the formulation of further hypotheses and to aid in the design of questionnaires. The group interview will also be addressed to questions relating to possible modes of advertising and distribution.

5. A questionnaire will be designed to accompany the sample issue in a mailed user study. Questionnaire design can be expected to benefit greatly from the results of the group interview process. The questionnaire will address itself to the following points:

(a) Given certain alternative ways of packaging the citations, would the recipient be willing to subscribe to one type of package and, if so, how much would he be willing to pay for an annual subscription

(b) How frequently should the publication appear for maximum utility?

(c) For what purposes would the publication be used?

(d) What are the user preferences for method of organization, format and contents of the unit entry?

(e) What other types of materials should be included in the publication to make it most useful?

(f) What other types of information services or products are needed by the user?

Additional data will be collected on personal characteristics of the respondent, including categorization of principal fields of interest, type of work in which engaged, present current awareness activities and major journals read. These data may be of value for a number of purposes, including determination of optimum methods of packaging and decisions on methods of advertising.

6. The questionnaire will be pretested on a small sample of, say, twenty physicists, and will be modified on the basis of the results of the pretest.

7. The mock-up and questionnaire will be mailed to a sample (statistically designed) of 1,000 AIP members and also to a sample of libraries.

8. Follow-ups will be used, and a sample of non-respondents will be contacted by telephone for a bias check.

9. Data from completed questionnaires will be reduced and analyzed. These data will allow extrapolations to be made on probable demand for the publication. Data will also be available on user preferences for organization, content and format. The price that users are willing to pay will also be established.

10. On the basis of this study, final decisions can be made on format, organization, packaging and price.

The study will involve the following schedule (Figure 2) which can be adjusted if necessary.

SUGGESTED SCHEDULE FOR THE STUDY

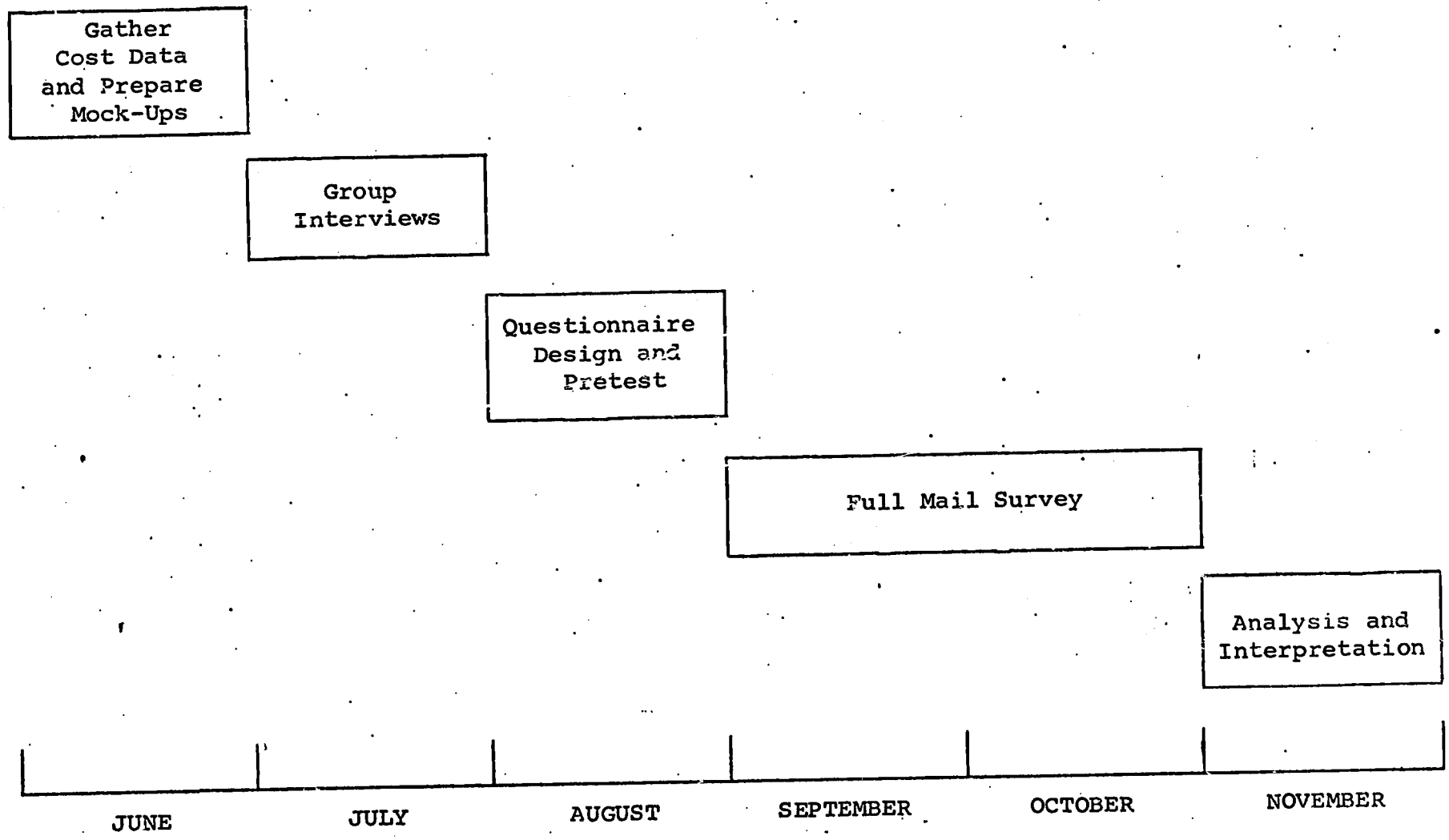


FIGURE 2

Suggested Procedures for Evaluation in Relation to "Demand Search" Requests

The purposes of this test are as follows:

1. To determine how well the AIP information system (including classification scheme, indexing policies and practice, and search methods) responds to specific subject requests of the "demand search" type).

2. To derive recall and precision figures.

3. To identify recall failures and precision failures.

4. To analyze causes of recall and precision failures in terms of:

   (a) inadequacies of the classification: lack of specificity, role problems

   (b) indexing policy and practice: indexer error, indexer omission, inadequate depth of indexing

   (c) searching strategies.

Because of the small size of the present data base, we recommend that the test be conducted by the use of "synthetic" requests based on "source documents" known to be in the collection. The following specific procedures are suggested.

1. Generation of Synthetic Requests

We believe that about 100 requests will be adequate to provide a realistic evaluation of the system at its present state of development. However, it is quite possible that some meaningful results (indicating trends) could be obtained with a smaller set (say 50 requests).

To obtain the requests (and subsequent relevance assessments) we must obtain the cooperation of some physicists. We understand that AIP has a panel of correspondents that could be used for this purpose. The smaller the number of physicists we have to deal with, the easier will be

the entire process. We can reasonably ask each physicist to handle up to four requests. So, we can get by with 25 physicists for 100 requests or 12-13 for 50 requests.

AIP staff will select the necessary group of physicists for participation in this study. Preferably they will be selected to represent the various broad areas of physics (chemical physics, nuclear physics, etc.). Each physicist will be sent some documents or document surrogates from which to generate the synthetic requests. There are two possibilities:

(a) Send each physicist a package of six papers known to be in the data base and known to be related to his area of subject interest. He will be asked to select four of these papers and to compile a subject request for each paper (i.e., a request for which the source document is regarded as a relevant response).

(b) Send each physicist a set of contents pages from a selection of journals relevant to his specialty. He will select four titles, obtain the papers, and formulate his questions as previously mentioned.

Whichever method is used, the end results will be the same. The second method has the advantage of giving the physicist a greater selection of articles to choose from. However, it has a big disadvantage in that it requires more effort on his part (he must obtain some articles himself) and could tend to reduce the likelihood of cooperation. On reflection, I believe that we should adopt the former approach — we will need the articles eventually for analysis purposes anyway.

Once the physicists are selected (we should have a few more than we really need to allow for dropouts) and the set of articles has been assembled, the actual task of obtaining the questions can be conducted by Westat. A set of forms and simple instructions will be prepared and the correspondence and/or telephoning needed to obtain the necessary cooperation can be conducted by Westat.

2. Searching the Data Base

The search requests will be transmitted to AIP for the preparation of search strategies. As previously mentioned, it would be highly desirable if each search could be formulated at three levels of specificity:

1. Broad search - designed for high recall.
2. Intermediate ("compromise") search.
3. Specific search - designed for high precision.

At the present time there is no capability for searching on natural-language terms. However, there is no reason why the natural language terms should not be used in the search strategies. Presumably, these terms will be used in conjunction with the class numbers to give greater specificity and therefore improved precision of search results. By manual comparisons, it will be possible to determine the exact precision ratio that would result from the use of these terms in searching.

This can be illustrated by a simple example. Suppose a three-level search strategy composed as follows:

1. (17 or 18) and 42
2. (17 or 18) and 42 and (A or B or C)
3. (17 or 18) and 42 and A

where 17, 18 and 42 represent class numbers and A, B, C represent natural-language terms. The machine search is conducted only on strategy #1 and the entire document set retrieved is submitted to the requester for his relevance assessments.* Having obtained these assessments, we can determine whether the source document was retrieved (and

*If total retrieval is very large, we can use random sampling for relevance assessment purposes.

thus establish a recall ratio - 0 or 1 - for this search) and what the precision of the broad search was. By examination of the terms and class numbers assigned to the retrieved items, we can derive recall and precision ratios for the level 2 and level 3 strategies. Thus, over 50 or 100 searches, we can derive average recall and precision figures showing the effects of variations in search strategy. These results can then be presented as a three-point performance curve.

3. Presenting Results to the Requester

Having conducted the search, the results will be presented to the requester for his assessment. In order to test relevance predictability on the basis of various types of surrogates, we propose that the results be presented to the requester in a number of stages as follows:

(a) Citations alone.

(b) Citations plus natural language translation of class numbers.

(c) (a) + (b) + natural language descriptors.

(d) Full documents.

To save multiple dealings with each physicist, we recommend that all sets be submitted at the same time, each in a separate sealed envelope. The physicist will be given detailed written instructions as to how to participate in this evaluation. These instructions will be prepared by Westat and all necessary correspondence will be conducted by us.

4. Summary of Performance Figures

Based on the completed relevance assessments, performance figures (recall and precision) will be derived for each search and averaged over the entire group of test searches. Separate recall and precision points for the various strategies — broad, intermediate, narrow — will be derived. In addition, by analysis, we will determine how effectively relevance predictions were made on the basis of the various forms of surrogates. This analysis will be done by Westat.

5. Failure Analysis

All recall and precision failures encountered in the test will be analyzed and sources of system failure will be identified. This will involve an examination of: papers, indexing records for these, requests, and search strategies. The specific source of each failure will be identified and overall conclusions on system weaknesses will be made. Preliminary failure analysis will be conducted by Westat. Detailed failure analysis will be done jointly by Westat and AIP.

6. Report on Test

A full report on test procedures and test results will be prepared by Westat.

7. Supplementary Tests Using Same Corpus

Once the test corpus has been assembled, a number of supplementary studies can be conducted with very little additional effort. Two of these are mentioned below.

(a) Comparison of AIP indexing with that of IEE. By examination of index terms assigned to recall and precision documents by both organizations, a comparative study of the effectiveness of both sets of index terms can be conducted. The results can be expressed in terms of comparative recall and precision figures.

(b) Ability of AIP staff to predict relevance of papers to needs of requesters. One or more members of AIP staff can be given various forms of document surrogates and asked to predict relevance to the stated request. These relevance predictions can then be compared with the actual assessments made by the end user.

Suggested Immediate Procedure

We suggest that we try out the above procedures with three physicists and 12 requests. This will be a pretest to iron out problems, improve the methodology and generally to determine whether or not we are getting meaningful results. If the pretest is a success, we will extend to a total

of 50 requests. Later, if we feel it to be necessary, we will conduct an additional 50 searches, bringing the total up to 100.

Division of Responsibility

AIP	Westat
1. Select physicists and documents. Make copies of documents.	2. Correspond with physicists, prepare forms, collect questions.
3. Prepare search strategies for each request.	5. Devise necessary forms and obtain relevance assessments from physicists.
4. Run test searches and deliver results to Westat, including copies of articles retrieved in each search.	6. Derive recall and precision figures.
5. Conduct, with Westat, final failure analysis.	7. Analyze relevance predictability on various forms of surrogate.
	8. Conduct preliminary failure analysis.
	9. Prepare report to AIP.

Estimated Costs for 100 Searches

AIP	Westat
Impossible for us to estimate exactly. Should be estimated by AIP. Should include duplicating costs.	40 mandays of senior professional 15 mandays of junior analyst Plus visits to New York Estimated Total = $9,300*

*Most of this cost is involved in analysis. The cost could be scaled down proportionately by reduced number of searches - 50% for 50 searches.

Suggested Procedures for Evaluation in
Relation to Current Titles and Other Published Indexes

The purpose of this test is to determine the feasibility of deriving search strategies, based on Boolean combinations of class numbers, that will automatically map citations to appropriate subject headings for organization and printing in Current Titles and other published indexes. We believe that this test can be done largely in-house by AIP, possibly with some assistance from Westat.

The test will be based on the list of subject headings used by Physics Abstracts. AIP will divide up conceptually the field of physics into its broad subfields and will select at random a number of subject headings relevant to each of these subfields (e.g., some nuclear physics headings, some chemical physics headings, etc.). A total of about 50 headings may be enough to get a feel for the way things are going. For each of these headings a search strategy (algorithm) will be prepared by AIP staff based on the existing corpus of citations that have been indexed by IEE. That is, the search strategy will be designed to retrieve as many as possible of the citations that appear under the appropriate subject heading in Physics Abstracts.

Having arrived at search strategies for each of the test headings, these strategies must be validated by application to a second, independent corpus. To do this, we propose the following steps:

1. For each of the test headings, select a number of journal issues likely to contain papers pertinent to this heading. These journal issues should be issues in the AIP data base but should exclude that part of the data base supplied by IEE, from which the search algorithms were derived.

2. Using the journal issues thus selected, choose a number of papers that are deemed to be relevant to each of the test headings. Relevance may be determined by a "jury" of, say, three AIP staff members.

3. Use each search algorithm to conduct a search in the data base. For analysis purposes, exclude the IEE-supplied portion of the data base. For each search, we will then have a set of citations retrieved from the AIP data base (exclusive of the IEE-supplied portion). Each search output will be examined to determine what proportion of the "known relevant" papers (i.e., the set selected by the AIP jury) for each heading were retrieved. This will allow the derivation of a recall ratio for each search. The same AIP jury will examine the complete list of retrieved citations for each heading and will determine which are relevant to the heading and which are not. This will allow the derivation of a precision ratio for each search.

These procedures can be illustrated by a simple example. Assume the heading MAGNETOHYDRODYNAMICS. By examination of various journals in the data base, we find five papers of relevance to this heading (i.e., five papers that should be cited under this heading in a published index). We use the search algorithm for this heading and conduct a search. This search retrieves 20 citations, including four of the five "known relevant" papers. The recall estimate is therefore 4/5 or 80%. We now examine the remaining 16 citations retrieved and decide that 11 are relevant to the heading MAGNETOHYDRODYNAMICS and 5 are not. The precision ratio of the search is therefore determined to be 15/20 or 75%.

4. For each search, then, we will derive recall and precision figures. These figures will illustrate the difficulties involved in using the AIP classification and indexing to map to a scheme of subject headings. Certain headings may be mapped to very successfully while for others the mapping may be quite unsuccessful. Perhaps we will find some relationship between complexity of mapping and subject area, indicating possible deficiencies in the classification scheme in various areas.

5. For each search, we will presumably have discovered some recall failures and some precision failures. The last stage of the experiment will involve an analysis of these failures. Documents, indexing records and search strategies will be examined to determine the precise cause of each failure. Some may be due to inadequate search algorithms, others to inadequacies in the classification, while a further group

of failures may be due to indexing errors. A test of this kind, based on a selection of headings, will not reveal all of the problems of mapping to a scheme of subject headings but it should demonstrate the kinds of problems that will occur and which of these are likely to be most critical.

6. The same test procedure can be used to compare the effectiveness of AIP and IEE indexing. We understand that some of the "non IEE" corpus has been indexed twice, by both AIP and IEE. By examination of the IEE index terms assigned to the test documents (recall set and precision set for each search) we can estimate recall and precision based on IEE indexing and compare this with the results actually achieved on AIP indexing.

Division of Responsibilities

Most of this work will have to be done in-house by AIP staff. Westat could assist in preparing the statistical summary of results, conducting preliminary failure analysis, and undertaking the AIP-IEE indexing comparison. Based on 50 test searches, I estimate that Westat's level of support might amount to about three man-weeks of senior professional time for a total of $3,000.*

I recommend that this test be phased in gradually. Let us begin with 10 or 12 subject headings, go through the entire process, and see which kinds of results we get. We can then modify our procedures before going to the full set of 50 results. This would also allow us to make firmer estimates of the manpower needed to do the more complete evaluation.

*This is based on a detailed failure analysis of 50 searches.

APPENDIX B
REPORTS OF VISITS MADE BY D.W. KING AND F.W. LANCASTER TO NATIONAL INFORMATION SYSTEMS

American Mathematical Society - Dr. Gordon Walker and Sam Whidden

The principal purpose of visiting the American Mathematical Society was to determine from them the feasibility of running an offprint service for users of American Institutes of Physics. The American Mathematical Society has approximately 140 journals that they handle and 300 mathematical reviews. They order offprints from other journal sources by individual articles and can expect about a six week return on obtaining a galley proof of these particular articles. Their process takes the following form: First the authors are coded from the galley proofs and each article is given a journal coden. The title is then keyboarded along with an author code and series, etc. The galley is then Xeroxed and made available upon request of by subscription from the Xerox copies. A mathematician then classifies the articles essentially by 50 classifiers who represent different mathematical disciplines (with some overlap). The classifier then determines a subject classification by assigning one or two primary classes and two or three secondary classes. From this a profile form is established.

A subscriber to the mathematical offprint service gets four lists from which he can choose documents. The first list concerns the 140 journals that comprise the population of journal articles. The second list consists of a number of different foreign languages. The third list is broken down into two sublists of primary and secondary subjects, and the fourth list is a list submitted by the subscriber concerning certain authors or citations that he wishes to get every time that they appear. For example, an author may wish to get every article in which he is cited. Thus, there is a total of six different lists from which a subscriber can obtain articles either by presubscription or on demand.

Orders may be placed by any one of the six sublists depending on the desires of the subscribers. They can subscribe the, by three negative instructions and three positive instructions. The negative instructions are for total inclusion, no more than a title list, or no more than a bibliographic index. The three positive instructions are at least the title listing, at least

a bibliographic unit, or an author. One can ask, for example, for all articles given in Roumanian or for all articles except articles given in Roumanian.

It is obvious that the system can serve as a retrospective search, for selected dissemination by user profile, or for a current awareness or alerting device. For example. the users could ask only for titles or abstracts having certain characteristics, which really amounts to a current awareness service. The system at the present time involves 1200 subscribers who can submit requests at any time. They have approximately 5,000-6,000 offprints involving 20,000-25,000 titles per month. Orders for the galleys for articles to be incorporated in the system are made over two weeks. The cost is approximately $100-$125 per year for each journal that they have in the system. They pay a page price for purchasing offprints. They order a small number of extra copies to handle late orders or orders that are requested on demand. They establish their pricing policy by assessing a price per page plus processing costs that amounts to 17¢ for a total of 47¢ per page. They operate only on a $30 deposit and a computer subtracts new orders from the balance which seems to last approximately one order for most of the subscribers.

I believe that Dr. Walker indicated that they need approximately 2,500 subscribers for the system to break even.
They feel that their marketing efforts may not be as efficient as they might be and ultimately hope to achieve a break even or superior sales market.

If an offprint service of this kind could really be made efficient, it seems that the service could ultimately result in a drastic change in the structure of journal publications and the way that journals go about packaging their articles.

They do feel that their present users are quite satisfied with the service and that most of their subscribers are individuals who are considered leaders in the field.

Other plans for the American Mathematical Society

1. The American Mathematical Society has planned on designing an information center on symposia that will announce titles of papers that are going to be given in symposia. The principal objective of this is to inform people of papers that are being given, or have been given, and to avoid possible conflicts and redundancy of information.

2. They are making several broad attempts to improve library classification systems. For example, they are working with the Library of Congress to develop some improvements on the Dewey classification system and they have asked the Library of Congress to use the same classification that the Mathematical Society uses on its Mathematical Offprint Service. They have also talked to all book publishers to have various assigned subject classes on copyright pages.

3. They have an extensive cooperative program with FID on the universal decimal classification system. Much of this work is in cooperation with Viniti. They hope to develop a system in UDC because they feel that there is less requirement for modification of the UDC classification schedule than one finds in most sciences, such as Physics.

4. They have an interesting new information service in which they produce audio recordings of a number of invited addresses. This system is recorded on a standard cassette and is suplemented with a manual that the listener can use to see space diagrams and formula that the speaker may address himself to. They apparently have found a fairly broad market for this and are selling them at a fixed price.

CENTRAL INTELLIGENCE AGENCY

For obvious reasons we are very limited in the amount of information we can disclose on the CIA information systems. However, Westat is active in consulting with CIA and there is at least one aspect of the Agency's information program that has some relevance to AIP.

There are certain similarities between the CIA indexing and the indexing methods proposed for AIP. Documents in the CIA system are described by three sets of terms:

a. subject codes (selected from a broad classification scheme having about 250 codes in all).

b. area codes (representing geographical region)

c. keywords (uncontrolled)

The area codes are peculiar to CIA, but the combination of subject codes and keywords is equivalent to AIP's use of a classification scheme supplemented by descriptors.

The keywords used by CIA are taken from titles of documents. However, if the title of a document is not sufficiently descriptive of its content it is expanded by the indexer through the addition of suitable keywords. The indexer marks the face of the document to indicate which title words are to be regarded as keywords (obviously only substantive words are thus selected) and also adds subject codes and area codes. The input typist works directly from the document itself.

Lancaster has conducted an evaluation of the CIA system and the combination of subject codes, area codes and keywords proves very powerful in narrowing down the scope of a search. For example, the coordination of the area code for Australia with the area code for the United Kingdom will

retrieve vast quantities of material, but the addition of the single keyword LAMB will cut down the output dramatically and virtually restrict it to the subject of "export of lamb from Australia to the United Kingdom".

The present indexing philosophy is approximately two years old. Previously CIA used a very detailed intelligence classification scheme. The move to the present procedures was a deliberate step intended to reduce input costs. It has been highly successful in this respect. Unfortunately, when we reduce input costs we also tend to increase output costs. This has been true at CIA. Lack of control of keywords puts a much greater burden on the searcher and causes a great deal of duplicative effort. For example, a search is conducted on "petrochemicals in Indonesia" and the searcher must think of all possible keywords that might indicate petrochemicals. How comprehensive this strategy is will depend on the ingenuity and perseverance of the searcher. Once the strategy has been used it is lost for further application. The next time a request is made for information on "petrochemicals in ..." the petrochemicals strategy must be created again. Although lack of complete vocabulary control (i.e., allowing free application of keywords) is economical in indexing it is burdensome to the searcher. Following our evaluation of the CIA system we have recommended free use of keywords by the indexer but a limited form of control of these keywords to aid the search process only. That is, keywords used uncontrolled by the indexer will be grouped together into logical clumps to assist the search process. Lancaster has suggested how data processing techniques may be used to aid the construction of a limited controlled vocabulary of this type. This may be discussed further with AIP if it is a matter of some interest.

Another element in the CIA system that may be relevant is the actual document delivery system. CIA, although they make heavy use of microforms, have not gone the NLM route and invested in the development of expensive special-purpose equipment. The bulk of the document collection is stored in

Sanders Diebold "Power Consoles" on aperture cards. Each console, which costs approximately $5000, will store about half a million aperture cards. An IBM Micro Copier is used to reproduce the cards. The reproduction copy is considered disposable and is handed out in place of a loan copy. Cost of reproduction is approximately one cent per card. Integrity of the master file of documents is fully maintained in this way.

Clearinghouse for Federal Scientific and Technical Information

CFSTI functions primarily as a supply depot to make Government research reports available to the general public. However, one very important activity is the announcement function whereby these reports are brought to the attention of potential customers. The Clearinghouse has gone a long way toward development of efficient announcement devices and it is this aspect that is of particular relevance to the AIP program.

For a number of years the principal announcement device was U.S. Government Research and Development Reports (USGRDR), a semi-monthly abstract journal with a separately issued index (subject, personal author, corporate author, contract number, accession/report number). Although comprehensive and very useful as a retrospective search tool, as well as a selection and ordering tool used by libraries, it was felt USGRDR had definite limitations as an announcement device from the viewpoint of the individual engineer and scientist. Principal reasons are:

1. It is too bulky physically to handle with ease.
2. Subscription price was too high ($30.00 p.a. without indexes) and subject content too broad for individual subscriptions.
3. Format and arrangement made rapid scanning extremely difficult.

In cooperation with the Air Force Office of Aerospace Research, the Clearinghouse began experiments with a new announcement medium in 1967. The experiments, a type of group SDI service based on a number of broad subject fields, was highly successful and led to the introduction of the Clearinghouse Announcements in Science and Technology (CAST) in 1968. CAST is issued in 46 subject categories based largely upon the COSATI subject classification. A number of these categories are physics-related. Subscription to one category is $5.00 p.a. (semi-monthly). Each additional two categories can be purchased for $5.00 p.a. A number of issues of various CAST sections, physics-related, are attached. As a minimum, a full

bibliographic description of the report is given. Usually an abstract is also provided. Documents considered of major importance are marked thus ☐. In this way CAST acts as an evaluative tool as well as an announcement device.

Another current awareness device produced by the Clearinghouse is the Fast Announcement Service (FAS). FAS is issued in 57 subject categories. Each Fast Announcement is a single sheet highlighting recent new R&D reports received by the Clearinghouse in a particular subject area. Subscription to FAS is $5.00 p.a. Fast Announcements are also sent to the trade and technical press for reannouncement.

Selective Dissemination of Microfiche

In late 1968 the Clearinghouse began to plan an SDI system based on the dissemination of microfiche copies of reports. It was planned that copies of all scientific and technical documents announced for public sale by the Clearinghouse would be available for automatic distribution in several hundred selected categories. Customers could subscribe to one or more of these subject categories and would automatically receive a microfiche copy of every report or translation falling into these categories. The projected service, Selective Dissemination of Microfiche (SDM), would be a faster and more economical method of obtaining the latest scientific and technical documents in selected fields of interest. It would eliminate the need for scanning lists and placing orders for individual documents. Moreover, the automatic distribution procedure would allow the new service to be offered at a unit price considerably less than the unit price (65 cents per title) at which microfiche is offered through the regular ordering procedures.

Westat conducted a small study of the market potential for the contemplated SDM service early in 1969. The market potential was studied in terms of the number of organizations having microform equipment available and the number indicating a direct interest in the proposed service. As a result of the promise shown in this study, the SDM service was initiated in July 1969.

The service offers great flexibility to the individual subscriber in how he structures the categories in which he receives the fiche. He can, for example, specify any one of the 22 COSATI fields, or any one of the CAST groupings, or he can specify the agency or agencies required (e.g., NASA only or not DoD). So far there are 80 subscribers but the list is continually growing (CFSTI does not use a "hard-sell" marketing approach). These are institutional subscribers and generally they establish a very broad category in which to receive fiche. Within this service, fiche can be offered at $0.27 each. CFSTI also extends the service through the Defense Documentation Center.

ERIC User Please Note:

Pages B-10 to B-51 are not reproduced here due to marginal legibility of pages in original document. These pages in original document consisted of xerox copies of issues of Clearinghouse Announcements in Science and Technology discussed on p. B-7, B-8, here.