

The Cost-Effectiveness and Cost Benefits of Commercial Information Services

A. Cawkell and E. Garfield, Ph.D.

Institute for Scientific Information
325 Chestnut Street
Philadelphia, Pa. 19106

Cost-effectiveness and the more difficult subject of cost-benefit are best considered against the historical and sociological background of information services generally, inclusive of chemical services. 'Cost-effectiveness' will be used here in the sense of the cost necessary to achieve some desired performance level, whilst 'cost-benefit' is the value, preferably reduced to monetary terms, of the benefit derived by the user of the service.

Until comparatively recent times libraries were the major information centres for archival scientific information, whilst current awareness requirements were met by journal browsing and to some extent by abstract journals. Most libraries were, and still are, free to users. In 1963, Campbell(1) said 'people who are accustomed to pay the doctor (in the USA) and the lawyer are accustomed to help themselves to many kinds of knowledge free of charge. . . teachers, parents and librarians do not say "the bill will be so many dollars." For some information we pay, for some we do not, and the difference is not based upon the values involved but upon dogmas, conventions and attitudes in our society.' In these circumstances there is little incentive to investigate cost-effectiveness.

The second era of information started in the twentieth century, but did not gain much momentum until the 1960s when it was realized that traditional methods were failing in the face of an exponential literature growth. This realization led to a great increase in activity from professional societies, governments and companies. In general, the tradition of providing 'free' information continued and very large sums of money were spent in developing computer-based systems. Cost information continued to be as scarce as ever, and the literature relating to actual operating cost-accounting systems was scarce, if existent at all. Quite the opposite was true of articles that mention the desirability of having such systems (2).

A number of reasons for the lack of cost information have been suggested:

(a) It may be 'company policy' either because the figures are atrocious (so bad that they cannot be revealed) or because they indicate an efficient operation (the competition cannot be let in on the secret) (2).

(b) Systematic analysis, as developed by economists, has rarely been used in studying information activities, probably because of the substantial communication gap between economists and information specialists (3).

(c) Managers feel that having to measure costs in information activities is degrading. Information can be costed in the same way as any other commodity, but human values are greater than any price tag (4).

(d) Writers prefer not to divulge cost figures for preparing titles and abstracts for fear that they are too high and not justifiable; alternatively they do not know how to measure costs and therefore cannot accurately report them (5).

(e) Development costs are usually totally ignored in any general costing, sometimes because they are embarrassingly large (6, 7).

A new era of critical examination and realism started around 1968. It is suggested in the SAT-COM report (8) that policies for disseminating information are usually formulated by members of the scientific and technical community, who, for the most part, have rejected the view that information must be marketed. The general belief is that if information is valuable, the people for whom it is intended will find it, but there is little basis for such a belief. Elsewhere (2) it is suggested that by charging for the services, even if only 'on paper', then top management, the user and the centre would all become involved in the negotiations. Information services would then no longer be the first choice for cut-backs in lean times.

The change in climate has been accompanied by advice, exhortations and proposals (9) to get

to grips with the cost-accounting, cost-effectiveness and cost-benefits of information. The time is coming when at least some part of costs will be borne by users, and thus all operations should be costed so that involved parties will know what fraction they have to bear. If all costs are not to be charged to users, should they be borne by the government as a public benefit, or by the centre, or recovered in some other way? These questions and the question of the survival of information centres are discussed elsewhere (10). It is becoming evident that those subsidies which were prevalent during the last seven to ten years may soon cease. Future survival of the information centres will depend on their ability to obtain remuneration for their services (11).

Unfortunately the increased publication about costs is unaccompanied by an increase in accuracy. Variations in quoted costs are so great that it is obvious that different costing methods have been used. Often 'costs' mean 'running costs', which are usually only a fraction of total real costs. Attempts to arrive at realistic cost figures from the literature are risky; authors publish on their 'unique' operations (2). However, with careful interpretation, some information can be obtained. Those who are considering or running an information system want to know about the relative virtues of various methods, and the cost of achieving this or that result, whilst those who wish to use a commercial or charged external service obviously need to know the price.

The cost of providing the service and the price charged are, however, as stated earlier, not necessarily related. On the one hand are the commercial services operating at the 'market price'. In the medium term, unless all cost components are covered in the selling price, the service must fail. On the other hand there are the government or institutional sponsored or partially sponsored services, offered to users at no cost, but embodying substantially the same cost components as the commercial services.

Production Costs and Prices of SDI Services

An attempt to equate some published information for four SDI services is shown in Fig. 6.1. Actual or notional 'market prices' are given to provide some basis for price comparison.

In Fig. 6.1, a hit indicates that one or more terms in the retrieved item matches a profile

term. Since all hits are not equally effective, the cost per hit is a cost indicator, not a cost-effectiveness guide. The 'cost per action hit', to be referred to later, is a better guide to cost-effectiveness.

Production Costs and Prices for Other Services

SDI services are usually based upon priced magnetic tapes, so the assessment of costs is relatively easy, assuming that the user is not interested in the real production cost of the tapes. For other services, cost estimates are more difficult to make, and the variation and inadequacy of the published data is so great that it is rarely possible to make comparisons. Some information is given in references (5), (8), (15) and (16). Patent indexing costs vary from \$7.50 to \$15.10, and abstracting production costs vary between \$8.70 and \$33.0; the cost of 'hits which generate document requests' (action hits) for SDI and other service lies between \$0.14 and \$2; 'the retrieval costs per query' have dropped to \$10 to \$20 for time-shared systems, compared to \$20 to \$50 for manual or batched processed retrieval systems.

Cost-effectiveness

A number of proposals have been put forward for measuring cost-effectiveness. In several, the well-known parameters recall and precision are included. Cleverdon has been responsible for extensive investigations in the area of cost-effectiveness and proposed an expression for the cost of retrieving a relevant document (17):

$$C_r = C_{an}/D$$

where C_r = cost of retrieving a relevant document

C_{an} = annual cost of system

D = number of relevant documents retrieved per year.

The cost C_r for MEDLARS in 1967 has been calculated as \$455,000/307,850, which equals \$1.48. The expression is discussed further by Kent (12). For SDI, Cleverdon proposes a similar expression with penalties for noise and miss.

Measures based on relevance are, however, criticized by Tell and Larsson (18), who suggest that anyone who stresses relevance as a criterion must answer the question: relevant to whom, and for what purpose? East (6) states that it would be extremely misleading to compare relevance/recall figures from one system to another unless close comparability existed.

Source of information	Data base	Frequency	Size of operations	Cost or price new pence (cents)/hit	Adjustment	Real or notional price new pence (cents)/hit
ISI price list (ASCA)	ISI source and citation tapes	Weekly	Large	14 (34) (Average for typical profile)*	—	14 (34)
Reference [12]	Chemical titles tapes	Fortnightly	Large	14 (34) for typical profile	Note 1	17.5 (42)
Reference [13]	ISI source tapes	Weekly	Large	7.5 (18) Average	Note 2	12.5 (31)
Reference [14]	Chemical titles tapes	Fortnightly	Medium/small	6.5 (15.5)	Note 3	19 (45)

* Average price per hit for average-sized profile with average mix of terms.
 Note 1 25% added for amortized development costs plus some overheads and marketing costs.
 Note 2 40% added for amortized development costs, marketing costs and some overheads plus 20% profit.
 Note 3 150% added for labour, overheads and marketing costs, plus 20% profit.

FIG. 6.1 Production costs and prices of SDI services.

The most expensive item in any system for information transmission is probably the user's time (19). Only after the inclusion of user costs can a decisive cost advantage over the manual approach be demonstrated. King and Bryant (20) consider that the question of the user being part of a system, or not being part of it, is a prime factor in evaluation. This reference covers many aspects of system evaluation.

A method of assessing cost-effectiveness has been suggested by several writers as 'the price users are prepared to pay'. In some cases, users are not prepared to pay a price comparable with production costs, so the service is subsidized. The sponsors obviously feel that the service is sufficiently beneficial to the users to justify the subsidy, whilst the users do not feel that it is worth that much, otherwise they would be prepared to pay the 'market price' for it. In one case (21) students felt that they would be prepared to pay from \$5 to \$15 (for searching theses), but costs are \$46, indicating that a subsidy will be required for unsupported graduate students. In another case (10), the cost of a unit of service varies from \$7.30 per request (full cost) to \$2.50 (minimum value), but if doctors are reluctant to invest the price of a telephone call to secure desired information, the service is unlikely to receive requests at the full cost of \$7.30 each. Nevertheless the effectiveness as adjudged by 'what users are prepared to pay' has some support.

Turning to a different aspect of user costs, Kent (12) suggests that cost-effectiveness in an SDI service could be improved by increasing precision at each level of recall desired. More comprehensive search strategies might increase search costs by 2% but improve precision by up to 25%. Heaps (22) goes further, proposing a criteria for optimum effectiveness as the tendency, resulting from the information service selection processes, to bring the relevant items to the top of the list. The complexity of cost analysis is well summarized in a recent article by Lancaster (23).

COST-EFFECTIVENESS SUMMARY

E.S. Quade states in a frequently quoted paragraph '... all analysis of choice falls short of scientific research ... we cannot turn cost-effectiveness analysis into a science'. There is no doubt that it is difficult to use cost criteria to compare services *before* implementing them, since the assembling of the data is time-consuming, and a most careful interpretation of it will be

necessary before coming to a decision. However, some proposals are given below:

(a) *Cost per action hit*

A user, confronted with a selected list of document representations, or 'surrogates', can rank-order them by interest level to himself, and will usually satisfy his interest by taking positive action and demanding actual documents above some arbitrary interest level.

The production of action hits is an important function of an information service, whether the action be a request for a library loan, reprint or translation. It is a rough measure of the effectiveness of the service, so the service cost per action hit becomes a rough measure of cost-effectiveness.

(b) *User time spent per surrogate (document representation)*

A surrogate (title/abstract) is at an economic optimum when it contains just sufficient information for the user to make a correct decision about the interest level of the associated document. The unit of measurement is the average time taken per surrogate for a user to read and assess all surrogates, plus the time wasted if the surrogate is misleading.

(c) *Number of documents of interest missed*

The degree of miss (either because of non-coverage or non-access) can be assessed by a user over a period of time. An accurate assessment is difficult because of the obvious problem of discovering the total number of interesting documents.

(d) *Response time*

The time interval between the publication of the primary document and the notification sent to the user of the information service, the response time, can be measured.

The weight users attach to these criteria will vary according to their requirements.

Cost-benefits

As long as the implementation of an information service continues to be an act of faith competing against tangible items which produce a tangible payoff, it will not receive comparable attention from management. Policy is succinctly stated in a recent article (24). Librarians and information officers, particularly in industry, must establish a stronger relationship with management. They need collectively an objective argument based on the economics of information, its handling and its value. Difficulties also arise

from the fact that benefits derived from information are intangible and hard to assign a monetary value (11), and that attempts to estimate benefits do not seem profitable (10). In spite of this gloomy view, the literature abounds with examples of benefits.

Services save time by improving access to the literature, making more literature available in a given time, avoiding duplication of effort, providing specific information for a new project, identifying brilliant ideas and providing state-of-the-art information. It is estimated that the 2.3 hours per week, saved by a particular SDI service, for each of one hundred chemists in a typical organization would be worth \$250,000 a year, (25). In another current awareness service seventy-three people found new research leads, eighty became more aware of the work of others and made additional contacts, twelve found that their current project had already been done, and sixty had more time available (26).

A specific example in the literature of information which provided a noticeable payoff was in the case of the Lightning fighter where an information search about leading-edge devices revealed an effective notch design which saved time and money in basic research (27). In general, it has been shown that high growth companies spend seven times more on 'intangibles', including information, than low growth companies (28).

At least one attempt has been made to break down all aspects of information flow and to modify classical information theory by considerations of the amount, 'informativeness' and costs of information (29). Information theory is, however, only concerned with the probability of the correct reception of messages and not with the meaning of the messages. As Howard (30) points out, if losing all your assets in the stock market and having whale steak for dinner have the same probability, then the information associated with the occurrence of either event is the same. A theory that involves the probabilities of outcomes without considering the consequences cannot possibly be adequate in the area of costing of information activities.

COST-BENEFITS SUMMARY

Scientific information services suffer, when competing for attention or funds, from comparison with products, equipment, or services producing quantifiable benefits. Sensational benefits directly attributable to a specific item of information draw spasmodic attention to the

need for effective services, but a yardstick is necessary. Time saved is probably the most easily measured benefit. It is suggested that after a current awareness service is implemented, managers of services and centres should always request data from users about time saved.

For any new research project a retrospective search will put a researcher in such a position that he need not spend time repeating the work of others. A user should record the activities he has not had to pursue as a result of the information gained from the search. These activities can be roughly translated into 'potential time saved'. The extreme case of the user who finds that all of the proposed research has been done elsewhere is of special interest, because the user should be able to provide an estimate of the money which would have been spent on the still-born project. Cost-benefit information of this kind and information about time saved by SDI is urgently needed to provide quotable data about the positive benefits of information services.

Comparison of Commercial and Non-commercial Chemical Information Services

Little distinction has so far been drawn between commercial and non-commercial services, or between chemical and other services, since most of the foregoing remarks have been applicable to both. It might appear that the non-commercial services, ranging from 'free' to 'cost-recovery' often backed by vast resources, would leave no room for others. A number of companies such as Derwent, IRL, ISI and CCM are, however, operating. Important innovations have come from research done in a commercial environment by people such as Shannon, Luhn, Taube and Garfield.

The most important factor likely to affect the cost-effectiveness and the cost-benefit of commercial services is the incentive provided by competition and the fact that the service will fail unless users subscribe. There is thus a very strong incentive to study users' needs. The difficult history of the commercial services and the problems encountered in trying to reach viability have been frankly described by Hyslop (31). She concludes that subsidies will always be required. However, many opportunities do exist for non-subsidized commercial services. A large non-commercial service may be cumbersome and insensitive to users' needs, and that need can be met by an alternative better service. There may

be areas where no satisfactory service exists and a new service can prosper (for example, Derwent Patent Services, Anbar Management Information). In some cases entirely new methods are developed to provide alternative services (for example, ISI's Science Citation Index).

The 'duty' of a commercial service is to provide users with a specified service for a fee. The non-commercial service may impose upon itself a wider duty. For example, it may feel that it should provide users with 100% coverage of the literature. The pragmatic approach of the commercial services is indicated by considering the following:

- (a) provision of titles alone as the document representation,
- (b) coverage of the literature.

Intuitively one feels that a title, a title plus keywords, and an abstract contain successively more information and increase the probability of a correct interest-level decision by a user. The input and production costs of processing an abstract are probably at least ten times that of processing a title, but what is the payoff? Numerous studies have been carried out in this area, and whilst opinions are by no means unanimous, it seems that titles are often reasonably adequate (18, 32 - 36). In one case titles of ambiguous or non-informative content were very few, i.e. below 4% in a multi-disciplinary field. Title enrichment with keywords did not give any significant improvement, i.e. less than 3% of relevant references. Findings like this are so different from intuition that users find them difficult to accept. We conclude that the provision of titles alone leads to a most cost-effective service.

The adoption of titles alone in a commercial index, such as the Source Index part of the Science Citation Index, is correspondingly compelling. The costs of supposedly increasing the information content would almost certainly result in a non-acceptable cost increase to the user. The publication is already at a point of optimum cost-effectiveness.

When original documents are in difficult foreign languages, an English language abstract may be cost-effective. A user can then decide whether he should buy a complete translation. Scientific Information Consultants Limited operate a successful commercial abstracting service for Russian and Eastern European technical reports and other documents. Another case where abstracts fulfill a special purpose is in ISI's *Current Abstracts of Chemistry and Index*

Chemicus, in which the abstracts have a particularly high information content including structural diagrams to describe new compounds, reaction sequences, etc.

Altogether, there are about 40,000 scientific journals, ideally all of which should be covered in a comprehensive scientific information service. Various studies, for example ref. (37), indicate that everything of importance is published in a core of between 2000 and 3000 world journals. For a service to cover 6000 instead of 3000 journals, a 75% cost increase would increase the information content by, say, 5%. The duty of a commercial service is to cover 90% of the literature at a reasonable price. It is not constrained by a 'duty' to provide 99% coverage in order to produce a marginally better service at an unacceptable price. The non-commercial service may tend to follow the second kind of duty, imposing a heavy burden upon itself and its users.

The Future for Commercial Services

The increasing size and cost of a comprehensive information service is forcing users to consider ways of sharing costs. This pressure may be sufficiently compelling for industrial users to review the privacy usually considered essential for their operations. At the other extreme, such as in a University environment, security problems are minimal but present constraints on expenditure will encourage shared services.

Data bases are and will be operated 'in house' if there is a large enough clientele to justify the cost. Retrospective searching by machine is already cost-effective under certain circumstances, for example in chemical sub-structure searching (38), and it will soon compare more favourably with manual searching generally. The first publicly available commercial service started in the U K in April 1971 (SCISEARCH); access to a central time-shared machine is via a normal dial telephone line. The printed page, its preparation often machine aided, will retain a very strong competitive position, however. An item can be located in seconds from a well-ordered file containing millions of records, and in a well-used library the benefit is shared by many.

For computer-based current awareness systems (and very recently for retrospective machine searching) use has been constrained by costs. Joint operations mitigate that constraint. Mainly for this reason national and regional

centres and information brokers are being set up to operate one or more data bases for sections of the community. Commercial data base producers are likely to find that users who previously directly purchased services derived from the data base will change over to the local service. This loss will be offset by the greater number of users who will be attracted by local special circumstances. The operations at Stockholm conducted by Dr. Bjorn Tell are a good example (18). An SDI service derived from ISI Source Tapes is reinforced by local input on

engineering subjects. In this way a local need is met.

Interactive information retrieval will unquestionably be widely used in due course. A user can formulate questions, observe results and iterate to obtain the required information, thus almost completely removing the present limitations of information retrieval at arm's length from the machine. It has already been claimed (15) that this can be done at a competitive price. If user's time is costed, as it should be, the overall cost drops further.

REFERENCES

- [1] Campbell, W., Feedback From the User—Sine Qua Non, *Information Handling—First Principles* (P. W. Howerton, ed.), Spartan Books, Washington (1963), p. 22.
- [2] Penner, R. J., The Practice of Charging Users for Information Services: a State of the Art Report, *J. Am. Soc. Inf. Sci.* 21 (1), 67 (1970).
- [3] Olsen, H. A., The Economics of Information; Bibliography and Commentary on the Literature, ERIC Clearinghouse, ASIS, Washington (1971).
- [4] Wilson, J. H., Major Trends and Portents Related to Information Costs, p. 249, in *Proc. Am. Soc. Inf. Sci.* 7, 1970 (33rd meeting), (American Society for Information Science, 1140 Connecticut Avenue, N.W., Washington, DC 20036).
- [5] Landau, H. B., The Cost Analysis of Document Surrogates: a Literature Review, *Am. Doc.* 20 (4), 302 (1969).
- [6] East, H., The Development of SDI Services, *Aslib Proc.* 20 (11), 482 (1968).
- [7] Cawkell, A. E., Real Cost of Information Systems, *J. Am. Soc. Inf. Sci.* 21 (1), 100 (1970).
- [8] National Academy of Sciences, Washington, Scientific and Technical Communication (SATCOM), Publication 1707 (1969).
- [9] Helmkamp, J. G., Managerial Cost Accounting for a Technical Information Centre, *Am. Doc.* 20, 111 (1969).
- [10] Lutz, R. P., Costing Information Services, *Bull. Med. Libr. Ass.* 59 (2), 254 (1971).
- [11] Counts, R. W., Marketing Aspects of the User Interface Problem: in an Information Centre, p. 307, in *Proc. Am. Soc. Inf. Sci.* 5 (1968) (31st meeting) (Greenwood Publishing Co., Westport, Conn.).
- [12] Kent, A. K., Performance and Cost of Free-text Search Systems, *Inf. Stor. Retr.* 6, 73 (1970).
- [13] Rowlands, D. G., The Unilever Research SDI System, *Inf. Stor. Retr.* 6, 53 (1970).
- [14] Corbett, L., Using Commercially-available Literature Tapes for a Current Awareness Service, *Inf. Sci.* 2, 83 (1968).
- [15] Landau, R. M., Computer Terminals—The Gateway to Intellectual Infinity, p. 237, in *Proc. Am. Soc. Inf. Sci.* (1970) (see reference 4).
- [16] Freeman, M. E., Determining Costs of Information Systems, *J. Chem. Doc.* 7 (2), 101 (1967).
- [17] Cleverdon, C. W., A Cost-Effectiveness Measure for Information Retrieval Systems, Cranfield, England, conference (1970).
- [18] Tell, B. V. and Larsson, R., Information Retrieval with the A B A C U S Programme, p. 183, in 'Handling of Nuclear Information' (IAEA, Vienna, 1970).
- [19] Dammers, H. F., Computer Use in Information and Data Handling: an Appraisal of its Economic Aspects, *Inf. Stor. Retr.* 6, 17 (1970).
- [20] King, D. W. and Bryant, E. C., The Evaluation of Information Services and Products (Information Resources Press, Washington, 1971).
- [21] Smetana, F. O. and Walker, S. R., Literature Searches for These by Computer—Results of the First Two Years of the Programme, p. 95, in *Proc. Am. Soc. Inf. Sci.* 6 (1969) (2nd meeting) (Greenwood Publishing Co., Westport, Connecticut).
- [22] Heaps, H. S., Criteria for Optimum Effectiveness of Information Retrieval System, *Inf. Control* 18, 156 (1971).
- [23] Lancaster, F. W., The Cost-Effectiveness Analysis of Information Retrieval and Dissemination Systems, *J. Am. Soc. Inf. Sci.* 22, 12 (1971).
- [24] Gilchrist, A., Cost-Effectiveness, *Aslib Proc.* 23 (9), 455 (1971).
- [25] American Chemical Society report 1969, Cost Effectiveness of Information Services.
- [26] Bowman, C. M. and Brown, M. T., The Development, Cost and Impact of a Current Awareness Service in an Industrial Organization, *J. Chem. Doc.* 11 (2), 72 (1971).
- [27] Page, F. W., R and D, and Production, p. 23, in *Proc. Conf. Manag. Decis.* (1969) (Brintex Conference, London).
- [28] Nicolson, D. L., The New Approach to Information Management, p. 10, in *Proc. Conf. Manag. Decis.* (see reference 27).
- [29] Marschak, J., Economics of Information Systems, *J. Am. Stat. Ass.* 66 (333), 192 (1971).
- [30] Howard, R. A., Information Value Theory, *IEEE Trans. Syst. Sci. & Cybern.* SSC-2 (1), 22 (1966).
- [31] Hyslop, M. R., The Economics of Information Systems: Observations on Development Costs and Nature of the Market, p. 301, in *Proc. Am. Soc. Inf. Sci.* (1968) (see reference 11).
- [32] Parker, J. E., Preliminary Assessment of the Comparative Effectiveness of an SDI System Using Controlled or Natural Language for Retrieval, *Programme (Aslib, London)* 5 (1), 26 (1971).
- [33] Saracevic, T., Comparative Effects of Titles, Abstracts, and Full Texts on Relevance Judgements, p. 293, in *Proc. Am. Soc. Inf. Sci.* (1969) (see reference 21).
- [34] Searle, R. M., Human v. Machine Selection for Current Awareness in Mass Spectrometry, *J. Doc.* 26 (3), 221 (1970).
- [35] Cawkell, A. E., Human v. Machine Selection, *J. Doc.* 27 (1), 43 (1971).
- [36] Garfield, E., Methods and Objectives in Judging the Information Content of Document Titles, *J. Chem. Doc.* 10 (4), 260 (1970).
- [37] Martyn, J. and Gilchrist, A., An Evaluation of British Scientific Journals, *Aslib occasional publication number 1* (1968).
- [38] Granito, C. E., Becker, G. T., Roberts, R., Wiswesser, W. J. and Windlinx, K. J., Computer Generated Superimposed Codes for Searching Chemical Substructure Files, *J. Chem. Doc.* 11 (2), 106 (1971).