SECONDARY SERVICES: THE SMALLER PRODUCERS

Peter Popper

June 1974 WP-74-13

SECONDARY SERVICES: THE SMALLER PRODUCERS

Talk by Peter Popper during a
Meeting for Discussion

The Future of Information Services for U.K.
Biologists

Tuesday, 26th March 1974

THE ROYAL SOCIETY
London

When I was asked to give this talk I thought it would be easy to pick out a few of the smaller producers and talk about their products. However looking at the remainder of the program I noticed that some very large producers have not been mentioned. I therefore started thinking about which small producers I should offend by either talking about them or not mentioning them and decided that perhaps it would be more interesting to look instead at the reasons and justifications for small producers, and possible future trends.

Secondary services in general can be classified into

1) Encyclopaedic, covering as much of human knowledge as possible (examples are Referatnyi Zhurnal and Bulletin Signaletique)

2) Disciplinarean, covering one major subject as far as possible (examples are Biological Abstracts, Chemical Abstracts)

3) Subject specific, covering one particular sub-field, e.g. Dairy Sciences Abstracts

4) Mission oriented, covering an interdisciplinarean field in support of a specific project, e.g. biodeterioration

It is quite obvious that smaller producers can only fall into the last two catagories, the exception being a totally new field which requires coverage and initially is small and well-defined. A lot of major producers started this way.

It is perhaps appropriate to examine what smaller producers are aiming to do:
in many cases they are societal services which try to meet the specific needs of society members, or they are in-house industrial products which are marketed to recover some of the costs;

or they are tertiary services repackaging output from secondary services. In all cases it is a precarious existence although again in many cases the service provided is essential.

It is quite obvious that the normal user of secondary services will not sit down and read or even browse in Chemical Abstracts on a regular basis. Yet information is required and the small producers are basically trying to meet this need. It is true, that profiles from large data bases would in many cases satisfy the same requirements possibly more speedily as far as current awareness is concerned and more precisely in retrospective search situation than afforded by small secondary services; yet these continue to flourish. Why and how ? It is probably due to the generally higher quality of product:
selection of material for inclusion tends to be based on more critical parameters and the abstracts are usually longer and better, the penalty of course being reduced coverage and an increased time lag. In many cases readers of such services are perfectly satisfied with this since they say the screening reduces their reading and the time delay is no hindrance since they are personally in touch with all their eminent research colleagues and know about what is going through some sort of invisible college: the abstract merely serves as a reminder.

In industrial environments and here I would include say research-association services, the readership tends perhaps to be less sophisticated and does not demand large scale coverage:
most such services tend to reflect the varying interests of their sponsors and have almost a capitive audience. If what I have said so far is not too complementary I must apologize since such services do fill a real need. In looking at any rational set-up for information and abstracting services, this fact should not be forgotten.

How small producers obtain their primary information may be a matter of general concern and require speeding up and modernizing: what they do in producing usually a high quality product meeting a specific need is basically their own concern and should perhaps be encouraged and supported taking each case on its merits. I am sure we all feel that the products of the Biomedical Information Project fill a real need:
whether we should continue to produce them by the scissors and paste method is something that requires detailed investigation. I firmly believe that having spent a large proportion of national resources in research leading to primary publication we can afford a small fraction of this money to make this information available in a form most appropriate to the users' requirement and if users want small services then these should be rationally produced.

If such services cannot meet their mission because of financial restraints then provided a need for the service has been shown it should receive help, the terms of this again varying from case to case. The expensive part of the secondary services these days is input; if we can achieve a reduction of this item through multiple use, then how the input is used may well favour the smaller producer.

An interesting case in point is the new Abstract Journal on Alternatives to Laboratory Animals. There obviously is a real need for a source of information on this topic:
the producers carefully examined existing media and services and found them lacking. At considerable cost therefore a new service was launched:
production involves scanning of a large volume of primary literature for a very specific topic. The scatter in appearance in primary journals is relatively large:

the costs of identification of articles of interest is hence relatively high for each item reported. The product is excellent and no one would doubt its merits or justification. I am however suggesting that there must be a more economic and efficient way of acquiring the primary input.

I think I have already mentioned that small producers may have the advantage of quality of abstract over large producers. Here by quality I mean both the quality of the primary article abstracted because the mechanism of selection is more strict and the actual quality of the abstract. In many instances this approach recognizes the fact that not every primary paper published and abstracted by large producers will be of permanent value. Indeed it is true that only a relatively small percentage of papers and publications are referred to after their publication. Statistics even suggest that large services operating SDI services use less than 2/3 of their input for these and in retrospective searches the utilized percentage may be much smaller than that (I have heard figures of as low as 8 % mentioned in this context with the highest mention being about 70 %). Does this perhaps mean that large producers are wasting effort and that we should only concentrate on smaller producers? This would be a very dangerous conclusion since the identification of the relevant small percentage would require an ability to forecast which we do not possess. This however does not vitiate the argument that the smaller producers do give a higher quality selection at the penalty of missing some major new contributions in their respective fields, particularly where these come from unknown authors or institutes not normally associated with that type of research.

To resolve these conflicting trends and paradoxes would not be appropriate at this meeting. Sufficient that we recognize them and beware of the obvious pitfalls.

It is perhaps sad that having produced excellent abstracts, many of the small secondary services do not provide good and detailed subject indexes to match the quality of the abstracts. In many cases there is not even a notional attempt made to produce such search tools. It is here that the smaller producers must improve the service they offer if they are to justify themselves or expect support.

What then should small secondary producers have as their aim: I would suggest that they should present the highest quality abstracts possible consistent with production costs for primary articles selected for their apparent long life merit and to provide for these the finest possible subject indexing. With such a high quality product the twin needs of large services, promptness and completeness, are not of major importance. Such small services will instead become an elite service which may be the first point of call for researchers in the field covered. Duplication of items between such small services is unimportant provided of course that a service does not duplicate another one with the same subject aim. A small elite service which becomes the main point of reference in its field will gain a further type of support: notification by workers in the field of major articles which might otherwise have been missed.

In order to clear up any misunderstandings, I am not advocating a system consisting of inumerable small services only. The major services must continue and be improved but there is room for small elite services meeting specific parameters and requirements.

The existence of such services with good search facilities (because of the smallness different search strategies and techniques could be applied e.g. full text searching of the abstracts) could well also reduce the pressure for high cost searches on large data banks and result in lower retrieval costs:
I have no figures to back this up but it is just an intuitive feeling.

It is indeed the large encyclopaedic or disciplinarean services which will have to form one of the prime means of notification to the smaller elite type services we are talking about. Whether the feed stock which has to be screened by the small producer originates from an Unisist type scheme or is acquired in some other logical method, is immaterial. The important point is that the subset of information forming the subject of the small elite service is made available for screening, quickly, comprehensively and cheaply. This would rule out screening of primary publications by the small secondary except in really exceptional cases. This subset information should then be subjected to careful scrutiny by experts to select the best:
perhaps we are even talking here of a vetting process which is as stringent as the referreeing process set up for primary publications by societies etc. In this way the elite small service would become a critical one. It may be that this approach might well be the answer to the continuously increasing flood of unreviewed and unrefereed primary publications. This approach might well allow complete freedom of publication, with certainly should not be given up lightly but which in the end is self-defeating by its very volume whilst at the same time ensuring a separation of what could be classed say the upper decile of publication based on value of content in any area. I have already mentioned that there would undoubtedly be mistakes and errors of judgement in selection.

We would however by this means reduce the volume of literature which would be examined as a first step by researchers and students. We would carefully have to balance the expenditure in expert time for selection for the elite service against the time that would undoubtedly be wasted in each retrospective search or even current awareness reading based on an unscreened total bank. This latter would of course have to exist, perhaps not with the same deep indexing, to act as a safety net. It would always be possible to transfer items from the large data bank to an elite service.

If we pursue this idea we might even feel that putting together of the small elite serives with elimination of duplicates would provide a fairly good selective data bank. This might, if suitably composed, cover fairly extensive fields of knowledge, say the whole or life-sciences. Synthesis might perhaps go even further. Certainly such a data bank would only amount to a fraction of a current large service: since presumably it would be computer based it would be ameanable to in-house use by institute etc. and because of its high quality indexing be ameanable to interactive and iterative searches by individual researchers (normally the large data banks require much expertise in handling and costs of live searches are against individual on-line usage). We would therefore, as I said before, facilitate and speed retrieval of the "best".

What I have said would suggest that I can foresee an important new role for small producers, provided these concentrate in giving a high quality service in a narrow or specialist area whether this be disciplinarean or interdisciplinarean. It would certainly be interesting to conduct some tests or at least to sound the opinion of users whether these ideas fall in with their needs and expectations and whether the costs in selection time are justified.

I have talked to several eminent scholars who feel that such a screening process for inclusion in a specialist elite abstract service is essential and to be preferred to some complex referreeing mechanism in primary publication and that they consider it as much of their duty to perform such screening as teaching. I was quite frankly surprised to find this feeling which was quite contrary to the more widely held view that absolute completeness of coverage is the prime requisite and justification of a secondary service: it is true that until a better mechanism can be found even these scholars agreed the need for the safety net of an all embracing data bank.

I seem to have gone a long way from discussing examples of smaller life-science secondary service producers and to have plunged into a highly controversial area. I do not apologize for this but should stress that the views are entirely my own and do not in any way reflect those of OSTIor COBI or indeed IIASA. I do believe that we are faced with a potentially critical situation in the dissemination of knowledge. Technologically, computer science is making enormous advances and 5th generation machines are just around the corner: whether such machines will possess some measure of artificial intelligence or not remains to be seen but they may offer a means of performing the screening function under carefully controlled conditions. We should therefore have some experience with screened elite secondary services so that such machines could be fully exploited. To go on pushing everything into ever-increasing data banks which become more and more unwieldly and expensive to operate will certainly not be the best solution, even if we could be certain that everything is included. To deep index for 100 % retrievability all published documents is a pipedream which can almost certainly never be realized:
to attempt to do it on a partial scale without intellectual screening

of included material to my mind is also foredoomed to failure. Far better to use our resources to make a fairly rough and ready tool which will give us an acceptable degree of retrievability from a general comprehensive data bank and to supplement this by an elite type service either in the form of small subsets i.e. the small producers, and to synthesize these into larger but still managable services.

I hope that the discussion which is to follow will be provocative and lively: the more the better.