

NOTES ON A TECHNICAL INFORMATION SYSTEM

Joshua Lederberg

Purposes of an ideal system and how to get one.

First I ask the purposes of a communications system in science--Tukey has summarized them nicely as a switching problem. How can I maintain contact with a relevant sample of the community of knowledge? Three aspects: (1) prompt, detailed and reliable awareness of contributions of other workers in specialized field of interest (2) contact with broader fields of scientific advance, particularly those peripheral to my own specialized interests (3) retrieval at my initiative of specific data in the past archives of recorded knowledge. A single system will not need all these needs.

We should not ignore or distort the casuistic aspects of publication. The responsibility in attaching one's name to an assertion irretrievably in print is indispensable to the integrity of the scientific process. The opportunity of "contributing to human knowledge" formalized in the act of publication is a motivational foundation stone of scientific activity. Retrieval searches are often impelled by a humanistic obligation to understand science, to display the historical antecedents of new discovery, a perspective that may be more important than the inherent instructional value of the prior art. (At least each generation deplores the lack of this perspective among its own students.)*

The present system has generated two responses: the defeat of neurotic frustration for some, the compromise of narrow specialism for others. I feel the survivorship of humanistic science demands a **better solution**. Nor can technical progress coast indefinitely on the progressive narrowing of fields of interest that is the specialist's practical answer. There is no perfect solution, certainly not just retrieval alone. But what energies we have could be used more constructively if we could rely on the system for timely information instead of spending the effort and anxiety we all do now in fighting it with our own personal retrieval systems. As members of a scientific community we have a deeply rooted obligation to interact with the "literature". **Not** so much the size but the dispersion and formlessness of the institution make this an ever more hopeless aspiration. A conventional solution will be to redefine the "literature" as that part of our scientific legacy ~~to~~ which we have

*Citation indexing can make a special contribution here as well as aiding the search for specific items of fact and new approaches to old problems.

routinized access, but the still present possibility of foolish rediscovery will still generate as much uneasiness as our present conceptions of priority and the personal motivation of discovery will insist.

My suggestions are very simple and not at all original. They incorporate several of the ideas that Bill Know discussed in "compacting the literature"--but I can't agree that we can rely upon self-discipline. Page charges would be [^]disciplinary influence if the U.S. did not pay them. Now they work in reverse since an undisciplined author can justify his verbosity to a journal by Uncle Sam paying for it! We have to create an economic market which will so far as possible be self-enforcing to achieve generally desired ends. I can best illustrate my proposal by making it almost too explicit and referring to my own field--equally detailed corrections may be called for apart from the principles illustrated here.

First let us define, for an initial experiment, a large primary community of NIH grantees both generators and users of scientific information and already heavily subsidized for the efficient prosecution of their studies.

Proposal for a depository system together with select journals.

1. The National Library of Medicine (NLM) of the National Institutes of Health (NIH) should distribute its Index Medicus biweekly to this community (if you wish read this "make clear that NIH grant funds are properly used to purchase such subscriptions" as they doubtless now are).

2. The NLM should expand and liberalize its "loan" services--journals should not complain of copyright interference when they are eligible for other forms of implicit subsidy. These procedures will further help to fix the NLM as a central information nexus in the health sciences.

3. The NLM should announce that it will act as a depository for scientific manuscripts from this community. These manuscripts can range from brief technical reports such as might now be submitted to Current Biophysical Communications to carefully written, lengthy and thoughtful reviews.

4. By analogy with ASTIA, NLM will announce titles (with or without abstracts?) promptly in the Index Medicus and distribute copies on request to this community. The depository materials will also be used for intensive experiments in deep indexing, citation indexing, etc. insuring the widest penetration of retrieval operations into the collection.

NLM would also establish regional and international repositories at which copies would be available for local scrutiny. There are obvious **further** possibilities for data links.

5. The ground rules of the depository would include

a. Prompt acceptance and distribution of papers from the community on proviso of critical review of two other members of the same community. Some page or article limitation may be available on the discretion of the accepting office.

b. No paper may be withdrawn once deposited--as with journal publications the author's reputation is permanently attached to it. The author, of course, may submit amendations, corrections, etc., to be attached to a previous submission. The possibility of doing this is already a substantial advantage over present publication means.

c. Deposits would in general not be reviewed centrally in order to speed the availability of the papers to the community. However, if abuses become evident, the NLM office might have the discretion to abbreviate the title and abstract of, say, the third and subsequent contributions each year from a given author to "Contribution No. ". This kind of procedure on the one hand insures that no contribution is ever entirely excluded from the archives, and on the other allows for some degree of discretion in taking abstracting time, and space to broadcast news of a deposit.

d. The author might certify to having read editorial suggestions on format and non-redundancy; in due course he might also be asked to submit descriptors or citations on prepared punch cards to facilitate data processing. Sequential papers should be written as **addenda** to previous ones insofar as requestors can always obtain these concomitantly and **abstracts** will indicate the connections.

e. Authors will attach "MD No. " to their own reprints, taking place of the preprint system. But unlike informal preprints, authors will be formally responsible for them as equivalent to publication and may be quoted or criticized in the corresponding literature.

f. Authors may also submit the same articles to journals to be published at the journals' discretion. Author or editor may annotate the MD No. article to indicate such publication and also to indicate any critical amendations. [If an article is accepted in a journal of wide

currency the NLM might temporarily suspend its own distribution of copies if this is required for reasons of economy.]

g. An updated citation index might be attached to each MD for subsequent distribution; in any case this could be routinely furnished to the author if not comprehensively published.

6. Positive role of depository plan

a. Prompt and widespread availability and indexing service should make this an attractive vehicle for publication of contemporary findings. In a timely but responsible form. Much of scientific advance is sequential and the importance of prompt information to accelerate new discovery and minimize unnecessary duplication is not widely enough understood. That contributions can take a full year to come out in print is an absurdity of modern science. The central depository would facilitate retrieval operations. It would also discourage the redundancy implicit in peripheral publication and the irresponsibility of gossip and "invisible colleges". The connection of NLM with granting functions should further encourage the use of the system especially insofar as the deposits would do multiple duty as project reports (and otherwise obviously facilitate "scientific intelligence" within NIH).

b. The depository would also facilitate the publication of expensive archival documents--e.g. taxonomies--which may be of critical importance but have too *limited* a circulation to justify journal or book publication.

c. Authors prerogatives or abuses? Will the system be abused if deposition is so readily available? The same fact destroys much of the motive for abuse--there is no implied prestige in the deposit of n pieces of paper which have passed no hurdles, and there is negligible waste if they are not broadcast, only listed by title. *Meanwhile a substantial load can be taken off the journals.*

7. The depository and its retrieval system will meet indicated needs where the user must take the initiative--needs which journals fill in the most chaotic way: current awareness and archival retrieval. But the journals continue to play a critical role in scientific culture--they are broadsides on which I would rely to bring me unasked the best or overtly most interesting of contemporary science. The journals should revert to being select journals. They should be few enough that I can hope to scan the ones nearest my own field. They should stress reviews and commentaries (facilitated by better retrieval) that will help guide

me through the literature and find new connections through it. They should discourage useless redundancy. They should be legitimate sources of prestige. They should be attractively and durably produced. One useful journal would be explicitly chosen as a *periodic* reports-reprint series in various fields.

Present journals do not meet these criteria and ^{with happy exceptions the situation is} ~~are~~ rapidly becoming worse. ^{the journals at least more numerous.} The depository system may, however, take much of the moral pressure from them, especially those published by societies. And other constructive measures--through the details of inevitable federal subsidy--can be elaborated. I would prefer to see the overall level of journal output boiled down to about 10 per cent of its present level where I could start to cope with it. At least its further exponential growth--much of it somewhat cynically motivated to exploit present confusion--should be frozen immediately. This in itself, together with competition from the repositories, should gradually upgrade the journals to be "select journals" or at least whatever the readership wants them to be. The scientific societies should be especially sensitive to fulfill the requirements of their members under this definition of journal publication.

8. Economic impact on journals. The federal government is already inextricably involved in the economics of journal publication, if only through its massive support of research, and through indirect payment for advertising and through various hidden as well as overt subsidies. It can hardly withdraw from the field; it should recognize and rationalize its responsibilities. Government obviously can have no direct negative impact on journals by suppression--but it can favor the most useful patterns (a) through the competition of sources and readership and (b) financial subsidies for the qualified journals.

Ideally, a journal should be judged in a market of scientific readership, its own subscribers. The page charge is a subsidy levied at the wrong end and particularly hard to justify in parallel with a depository system. It should be revoked in favor of subsidy to subscribers, namely the grantee community, to facilitate their choice of and influence on the journals meeting their own needs. As a mechanism, say, 3 per cent of grant funds might be automatically available for the purchase of such communications at economic prices.

New profit-making journals pose a difficult problem. If authors (unaccountably) continue to contribute to them, libraries are black-

mailed into buying them almost whatever the price and the vicious cycle begins. One solution is that authors receiving the community services be urged or obliged to agree to submit their contributions to the depository as well where they remain potentially available even if not in so slick a form. The NLM can also get around this by more direct redistribution of the printed papers. There may be copyright problems but they are not insoluble insofar as each grantee has a direct obligation to his scientific colleagues. If they are insoluble under present law we may have to recommend new legislation.

NLM also has the leverage of what it chooses to index and thereby make retrievable. Journals that consist mainly of advertising should hardly qualify for this form of notice and by this convention will be "outside the literature". Individual worthwhile items can still be recouped by deposition.

9. **Extensions.** The NLM-NIH community is a large group but still confined enough to nurture such an experiment. Of course it cannot remain a closed group. The questions of extension are mainly secondary ones of financial policy. The main point is to establish whether the services are useful ones in relation to the costs of the system. Financial policy then becomes a useful tool for the allocation of the services. If the experiment is successful it is bound to spill over into other areas of science. But the present needs are perhaps most acute in biology and medicine.

10. **Centralization.** I am not an advocate of centralization for its own sake, but I prefer it to chaos. Many of the services implied here should be delegated to professional societies in those disciplines fortunate enough to have sufficient organizations. But the government must see that the job is done across the board and the holes plugged one way or another.

The ACS is already initiating a preprint service for "Industrial and Engineering Chemistry"--(why this journal, not JACS? I hope the scale and institutional support of the project is such as to give the scheme a fair test).

11. **International implications.** Our loose program of SIS is no help in trying to organize services and cooperations on an international basis.

A health sciences repository would be a very useful contribution to establish technical good will in other countries as AEC has already done. We should go far beyond NLM in this, but it would be a good next step.

Conclusion. The intended result of these plans would be a dual communication system. A centralized repository would provide the range of materials that I would specify as being required for my immediate and retrospective information requirements. Concurrently, select journals with high standards of selection and editorial quality would maintain my contact with the breadth of scientific culture. Practical means are proposed for establishing such a system on a competitive basis with a minimum of central duress.