Is the Literature Worth Retrieving?

by Samuel A. Goudsmit

A FEW YEARS AGO John Maddox, now editor of *Nature*, wrote an article[1] seriously questioning whether the scientific literature was worth keeping. He complained that style and incorrect sentence construction often make scientific articles unintelligible, thus hastening their obsolescence.

Obviously the literature will be kept, whether it is worth keeping or not. Large-scale activities on automatic storage, retrieval and dissemination of scientific information indicate that documentalists will discard nothing. Recent plans to add unrefereed preprints to the mainstream of scientific communication will further increase storage of worthless material.

The problem now becomes whether scientific literature is worth retrieving from storage centers. Although an enormous amount has been written about techniques of storing and retrieving, too little attention has been paid to the user and his needs. A modern computer can make a complete literature search in a remarkably short time and overwhelm the research worker with a bibliography that requires a lifetime to read and study. We hope, of course, that improved coding and indexing will limit the number of items retrieved so that the recipient can use them. Continued explosion of the literature, however, will make this limitation more and more difficult. Moreover the automatic computer cannot judge the quality of retrieved material, and perfect coding finally might retrieve only the preliminary announcement of the article the researcher himself is writing.

The problem is thus to make sure that the retrieved information is useful. I believe that for physics research the solution to this problem is to start automatic retrieval with critical review articles and critical data compilations. It is obvious that this can reduce the items to be studied to a sensible number. Review papers and compilations contain references to original literature from which the user can select judiciously the few he might need.

Old original articles are seldom a helpful source of information. They use notations and techniques that are unfamiliar to the present reader and are thus hard to comprehend. I doubt that many physicists younger than 60 can understand and appreciate Wolfgang Pauli's 1925 paper, in which he introduces the exclusion principle. With the rapid development of physics, nomenclature and methods in theoretical physics often become obsolete within a few years. This is a sensitive point that will be violently denied by many colleagues. I have already been told that the example of the 40-year-old Pauli article is unfair. It contains, however, one of the most important advances in physics of the century. Studies have shown that the half-life for physics-article citation is, on the average, only a few years and will most likely become shorter. Only the articles, not necessarily their contents, become obsolete.

To give recent examples would cause hurt feelings, but I can mention that much of my own work, not long after its appearance, was extensively quoted as "Condon and Shortley, *Theory of Atomic Spectra*, page x or y" where it was presented in a more useful form than in the original papers.

I am not consoled by the citing of the work of many other physicists for several years as "Bacher

---

The author, discoverer with George Uhlenbeck of electron spin, is managing editor of the American Physical Society and editor of *Physical Review Letters*. Born in The Hague, educated at Leiden, he taught at Michigan in 1927 and joined the Brookhaven National Laboratory staff in 1948. He is an officer of the Order of the British Empire, member of the National Academy of Sciences, an APS fellow and a correspondent of the Royal Netherlands Academy of Sciences.

With new techniques rapidly increasing the storage of physics literature, we should limit information retrieval to critical review articles and data compilations despite the loss of author credit that might follow. A good review article makes obsolete much of the literature that it summarizes. But getting good reviews is a problem. To get more good ones, we should organize teams of writers encouraged by attractive honoraria.

and Goudsmit, *Atomic Energy States*, page z." An article that first introduced "fractional parentage" and that according to Giulio Racah was the basis for most of the subsequent work on this subject has been generally overlooked and its results later rediscovered by others. The reason probably is that it appeared just too late to be treated in Edward Condon and George Shortley's excellent book.

Experimental papers can usually still be read fairly easily many years after publication, but here, too, rapid changes in technique make studying details of old clumsy methods often a waste of time. In short, an original paper seldom gives information in a more useful form than a review article, textbook or data compilation.

From this it follows that we must expand and support data collection centers, promote the production of review papers and make them suitable for automatic storage, dissemination and retrieval.

Data collections

Experimental results and measurements are the backbone of physics. No theory is acceptable unless it agrees with the experimental data. Conversely, a systematic study of experimental results can suggest new theoretical approaches. Tables and graphs of numerical data therefore play an important role in the progress of physics. The *International Critical Tables* were for many decades a valuable source of information; the new series of Hans Landolt and Richard L. Börnstein is now widely used. In a specialized field we had Heinrich G. J. Kayser's big volumes of spectral lines, which were essential for the initial successes of analyzing atomic spectra into so-called "series" and "multiplets." At a later stage tables of atomic energy levels were useful in the study of atomic properties. Compilation centers for nuclear data and high-energy particle physics distribute the present-day equivalents of the old spectral tables.

It is thus obvious that specialized data compilations are of great importance and should have the full coöperation of those producing the data. It is also clear that modern computer techniques can handle such data more efficiently than old tabulations could, especially since their number and variety are growing so rapidly.

There are, however, a few significant lessons to be learned from the history of the old tabulations, which should not be ignored even after automation.

- Tables of data or graphs never give complete information. (The great spectroscopist Friedrich Paschen said that looking at the photographic plate gives a better clue to which spectral lines belong in a series or in a multiplet than any amount of arithmetic based on numerical spectral tables.)
- Tables of data, when selected critically, can give enough information so that consulting the original work is not often necessary.
- Tables of data are not final. As soon as theory furnishes a deeper understanding, new measurements become essential. (It is significant that as soon as Kayser's tables gave an initial clue to the analysis of a certain spectrum, one had to disregard the old data and measure that spectrum anew.) At the same time the tables become partly obsolete and can often be effectively replaced by theoretical formulas. In this sense the extensive tables of spectral lines were replaced by the more compact tables of energy levels.

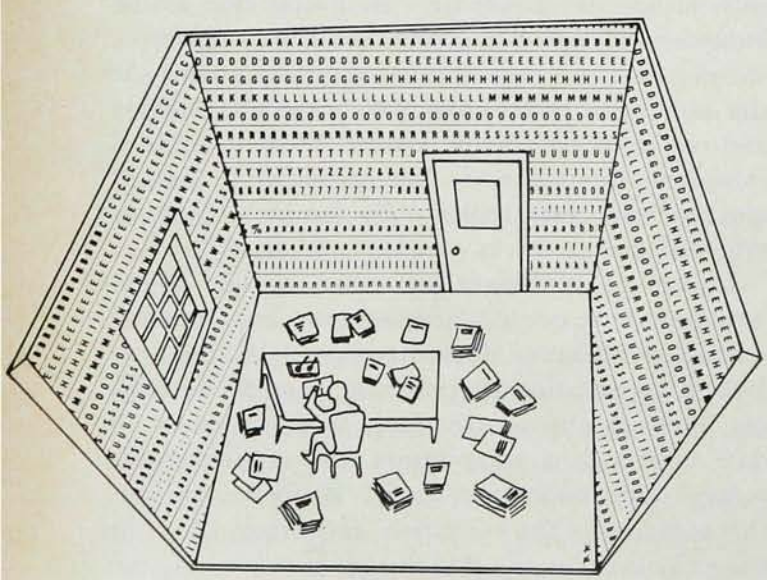
I wish to call special attention to the second item; it points towards anonymity of experimental results. Compilers of data often complain about the lack of coöperation they receive from experimenters. The subconscious reason is that as soon as his data appear in a compilation, the experimenter's name is no longer mentioned in footnotes, only the compiler's. Since citation and recognition are the principal rewards for the physicist's work, this

anonymity is indeed unfortunate. I see no way to avoid it, and it is obviously getting worse. Recent short notes on high-energy experiments with bylines of twenty to fifty authors, however, show that authorship no longer has the meaning it had in the past. These authors probably don't even know each other and their respective roles are not revealed. A film leader and a theater program also give credit to many individuals, but they list who is responsible for costumes, for makeup, for choreography, etc. The multiple authorship of recent physics papers is outright absurd; it might perhaps be helpful if the byline indicated who among them wrote the lyrics.

Theoretical papers still have only one or two authors; no "team" papers have appeared yet. This, however, does not mean that authorship is more significant for theoretical papers than for experimental articles with dozens of authors. The rate at which theoretical papers are published has increased enormously, and with a few brilliant exceptions, most of them contain very little advancement. Many are obsolete in a short time, and there is sharp competition among authors and strong pressure for rapid publishing.

Need for review articles

Retrieval of nonnumerical information seems more difficult than that of measurements. One can define the meaning of numerical data sharply so that information automatically retrieved can be just what is wanted. When dealing with concepts, however, their classification, indexing and coding are less easily defined and may change considerably when a field develops. Fortunately special methods are being developed to make such retrieval more meaningful than what can be achieved by subject classification alone.



For theoretical subjects, the equivalent of numerical tables is the critical review article or the specialized book. When the theory of atomic structure was just beginning to be developed, Arnold Sommerfeld's famous *Atombau und Spektrallinien* appeared. It summarized in a clear and concise way all that was worth knowing up to the time of its publication. It was of tremendous help to research workers, even beginners, and it stimulated an enormous amount of further development and discoveries in atomic structure, experimental as well as theoretical. Such a book should not be confused with an advanced textbook, which covers a subject after and not before it is fully understood.

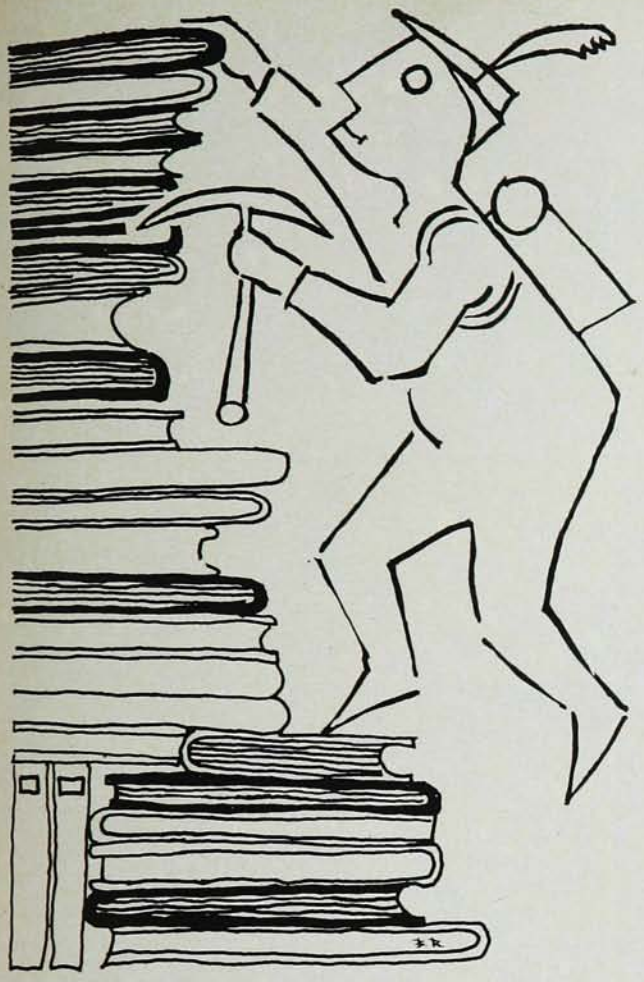
A similar example is the articles by Hans Bethe, M. Stanley Livingston and Robert Bacher on nuclear physics in *Reviews of Modern Physics* (1936-37). These and several other articles have in common that their authors pass judgment on what is worth saving in the literature; they make a study of original articles almost unnecessary, and, most important, they unify nomenclature, notation and method of approach. All this makes them ideal starting points for further work. Books and articles that periodically report recent "progress" do not usually exhibit such characteristics. The now popular collections of selected original papers are only a partial approach to review articles though some of them contain excellent clarifying introductory articles.

Authors often complain that critical review articles or books have not fully reported their work or have omitted a reference completely. If the review article were just an enumeration of the complete literature instead of a critical and clarifying exposition, however, it would be valueless.

It is, therefore, very desirable to promote the writing of books and articles of the kind mentioned here. With the growth of physics such an undertaking will have to be organized; one can no longer rely upon inspired volunteers to do this work. Unfortunately, as mentioned above, good review articles and textbooks doom authors of original articles to anonymity.

How to produce reviews

From the foregoing it is obvious that initial automatic retrieval of original physics articles by subject matter can seldom be helpful to the user. The number of papers retrieved will often be too large. (A recent review article on the highly specialized subject of SU (6) symmetries[2] contains almost 300 references to very recent literature on the subject. Though the article is more than 40 pages long, it is certainly more useful than a computer-produced



reproduction of all the original articles.) Moreover many retrieved articles will be incomprehensible. Therefore automatic retrieval should be started with finding review articles, relevant chapters in books and data compilations and should include only subsequent original papers. Older original papers should be stored so that they can be called for by journal reference or retrieved by other methods whenever the study of a review article or compilation makes this desirable.

This brings us to the crucial problem: How can we promote the writing of reliable critical review articles? We know from experience that merely paying for them is not the solution. As a trial, *Reviews of Modern Physics* a few years ago received funds from the National Science Foundation to pay $3000 each for four articles. Nevertheless the editor had difficulty finding physicists willing to give some of their research time to write suitable reviews. It has also been suggested that an award be created to honor the best review article every year. Such an award will in time raise the quality of these articles and enhance their prestige, but in attracting authors it can hardly do better than outright payment of a considerable honorarium.

The most useful review article is one that will be cited in the literature instead of the original articles it summarizes. Many review articles do not fill this requirement though they are excellently composed and helpful in studying a new field of research. We do not know just what qualities a review article must have to answer our specific need. The authority of its composer or editor and proper timing are probably important factors.

We suggest, therefore, that the writing of review articles be organized in a manner similar to data compilations. Several centers with a permanent staff of competent full-time and part-time writers prepare background material and drafts for a review article. Guidance and the final article will be the responsibility of one or two recognized authorities in the particular field, who do not have to be located at the center. In this manner it may be possible to get outstanding physicists to participate in this necessary project. Part-time writers could best be selected from teachers of advanced courses in the field under consideration. They too do not necessarily have to be working at the center, but a permanent staff is essential for continuity. Subject matter can probably best be selected by journal editors because they know in which area a review article is needed and who the best writers and authorities are. It would be advantageous if such review centers were located at large universities or research institutes. The American Institute of Physics could be the coördinating center for these activities.

I have been told, of course, that such an organization will fail; only the master can write a good critical review. With popular books as well as high-school and college texts, however, coöperative effort and assistance of professional science writers have given very successful results. There is no reason why this mode of operation can not be extended to the higher level of advanced research subjects. It is recognized that some physicists are excellent teachers, lucid writers and expert organizers of material whereas others are definitely not. A better appreciation and remuneration for good writing can focus attention on those physicists whose great contributions will be in communication. They can form the nucleus of the review-writing staff and can also play a vital role in improving the communication value of research articles. □

References

1. J. Maddox, *Is the Literature Worth Keeping?*, Rockefeller Institute Review, Feb. 1963; reprinted in The Graduate Journal of the University of Texas, Winter 1964.
2. A. Pais, Rev. Mod. Phys. **38**, 215 (1966).