THE MISINFORMATION EXPLOSION:

Is the literature worth reviewing?

...And can we increase the number of useful papers?

by LEWIS M. BRANSCOMB, Joint Institute for Laboratory Astrophysics, Univ. of Colorado, Boulder, Colo.

It is clear, everyone agrees, that a serious problem in information storage and retrieval faces scientists today, and some splendid and bold suggestions have been made for dealing with this crisis. Thus, Leonard H. Marks, Director of the U.S. Information Agency, advocated recently a satellite information grid that "would eventually make possible the ready retrieval and transfer, to any point on earth, of any single item of the total sum of man's million-year accumulation of knowledge."

But it appears to me that evidence on every hand indicates that the real problem is that the reader does not know what to believe in the surfeit of so-called "information" that retrieval systems may disgorge.

Consider the plight of the reader who wants to make use of some of the scientific gems of knowledge contained in the 30,000 technical journals published every year. He pushes the appropriate key word-buttons on his computer retrieval program and what does he get? A bibliography of papers, reports, conference proceedings and other scraps, each of which is certified to say something about the desired datum. How is he to find out what information these items contain? How is he to decide what to believe? What kind of a design margin should he leave for the possibility that the data he selects are in error?

What is the solution? We must find a way to throw out, for good and for all, a good fraction of the published scientific literature, and after a short time virtually all of the unpublished report literature.

We must increase the density of useful information in the literature. Information that is wrong is not useful. And much of the data in the literature is wrong.

Information that is presented in such a way that no one can tell whether it is right or wrong is not useful. That is a correct description of too many scientific papers.

We need better quality control in science.

Quality of reviews. A step that has often been proposed is to strengthen substantially the review function in the information system. The idea is that if teams of qualified scientists can be persuaded to anticipate the reader's needs—let's say for data on properties of matter—they can tackle the retrieval and digestion of the stored material in a comprehensive way and prepare tables of recommended values of such properties once and for all.

I think it is fair to say that the need for far more extensive reviews to reduce the raw material in the information system to a more concentrated form is now universally accepted.

It is just as absurd for the user to tap the total collection of raw material for his data as it would be for the jeweller to order six tons of gold-bearing ore when he wants to make a cuff link.

In the area of data on physical properties, the Federal Council for Science & Technology has re-
We must find a way to throw out a good fraction of the published scientific literature

tached to it in specified units. A statement about its precision and accuracy may or may not be at-
tached. It may be obvious to the informed reader that the result is wrong. It is still reference data.
There’s a lot of it.

Standard reference data, on the other hand, are data that have been subjected to critical evalua-
tion. There are some who may be satisfied by letting critical evaluation mean well-informed quality
judgment by an expert. For example, you can imagine tables of reference data from the literature
arranged by the reviewer in columns like the Sears Roebuck catalogue, with the headings “good,”
“better,” “best.”

But such an evaluation has several grave deficiencies. First, it fails to give the reader a quantita-
tive estimate of the amount by which the value given as “best” may be in error. Without this esti-
mate, no meaningful comparison with theory can be made, no design can be carried out with a prediction
of performance at a specified level of confidence.

Second, such a review cannot be demonstrated, by any a priori argument, to be superior to a review by
any other “expert,” whose opinions might differ sharply.

Thus, I prefer to define critical evaluation as requiring that a meaningful quantitative statement can be made about the probable
presence of systematic errors in the data. This statement must be
based on a set of objective criteria for assessing the likely presence and effect of systematic errors.

If there can be agreement on the criteria based on rational scientific
analysis, hopefully all competent scientists, evaluating the data from
this point of view, would come to the same conclusions about the
reliability of the results.

My definition of standard reference data does not ensure that such
data will not be found to be in error by much more than the stated
limits.

It is perfectly true that, in principle, one can never show that all the possible sources of system-
atic error have been identified and studied; to do so would require
that all of physics be known in advance. But, one does have a right
to ask that if the data are to be called “standard”, all of the well-
known sources of systematic error have been delimited.

Our experience shows that standard reference data are hard to find.
I question whether the existing re-
view literature contains an ap-
preciable amount of such data outside of atomic constants, such as the speed of light, the fine structure constant, and the Rydberg.

What is even worse, it appears that much of the reference data are not susceptible to such an analysis because of the failure of the authors of the papers even to mention the study of helium ionization in the measurement. And, if the data in the literature cannot be critically evaluated because the minimum of essential information is not present, then one must conclude that much of the literature is not worth reviewing.

Let me give a few illustrations taken from the three critical reviews that Kieffer's center has completed, and several others, which are under way.

They were written by scientists who are known in this field, some of them senior research staff of our Joint Institute for Laboratory Astrophysics (JILA), others well-established people at other universities, here and abroad. The reviews took from one to two years to complete with about half-time effort. The first dealt with experimental determination of ionization cross sections by electron impact (2). This seemed to be the easiest kind of electron cross-section measurement to make. Perhaps, though authors Kieffer and Dunn, it would prove to be a good place to find some potential standard reference data. After all, in 1927 a method for measuring total ionization cross sections was developed, and its application by P. T. Smith to the study of helium ionization is a classic in the field. Smith also approached the problem correctly by developing a theory for how his apparatus worked, and attempted to demonstrate experimentally that the data were consistent with this model.

Perhaps it is not surprising that it turns out 40 years later that the theory was defective, and there are known sources of systematic error not properly controlled in this first work. But the shocking thing is that after nearly 40 years of work, which produced about 30 independent reports of the helium ionization cross section, the two most recent results at the time of the review (1965) differed by almost 25 percent. To Kieffer's data, by the way, lie between them. Agreement for other rare gases is generally worse.

The authors of this review first set out to analyze the basic experimental methods used. For each type of apparatus they developed a set of specific criteria that had to be satisfied before one could state that a meaningful measurement had been made. When they compared the papers with the criteria, they found that only 10 percent of the papers in the collection contained even the most rudimentary evidence concerning the essential questions.

Some papers failed to mention any precautions to prevent (or measure) secondary emission from the electron-beam collector, thus nullifying the measurement of the electron current. Others failed to show that the path length of the electrons in the ionizing region was defined. Many failed to demonstrate a meaningful measurement of the target gas density.

Of the 30 odd measurements of the helium cross section mentioned before, only six of the papers could be evaluated in terms of the criteria established. The conclusion of the paper is: "The rare gas cross sections are generally regarded as well known; the data present (here) indicate that this opinion is not well founded unless one considers 20-25 percent as a small uncertainty."

The figure on page 49 shows an even more startling fact. None of the papers regarded by the authors as demonstrably free from all of the identified possible sources of uncontrolled systematic error, and only 24 of them escaped rejection on the more serious grounds that the experimental method was demonstrably unreliable, or that there was positive evidence in the original paper that the authors had a systematic error they either ignored or corrected for improperly.

The conclusion that appears to be valid, at least for this sample, is that in any decade from the 1910's to the 1950's, a constant fraction of the authors—only 10 to 14 percent—performed and wrote up their work in such a way that it could be said on subsequent inspection to contain useful information.

Now, let us look at the study of electron impact excitation of atoms. The story repeats itself, but here the situation is much worse.

For almost all authors have ignored the fact that the radiation being measured may be polarized, especially with excitation near threshold, producing an anisotropic angular distribution.

To make matters worse, the optical apparatus employed more often than not is polarized to some degree. These polarizations introduce complicated and usually unknown energy dependencies. On top of that, the procedures for relating the radiometric calibration of the photon detector to the measurement standards were often deficient.

From his experience so far, Kieffer told his audience at the International Conference on the Physics of Electronic and Atomic Collisions in London last July, "The literature does not represent a careful and reasonably objective record of the work being reported. In fact, as any critical compiler will attest, it is almost impossible to judge the reliability of the data presented from the reports in the literature."

Now, none of what I have said proves that the rejected data were wrong. One of the hardest things to convince experimental students of is that it is possible for a sloppy experiment to get the "right" answer, and a very careful one may have an insidious error that escaped correction.

And yet, a right answer in the literature may still contain no useful information, if there is not a logical and scientifically consistent way to evaluate the likelihood that it is right. It's like rejoicing in the accuracy of your broken clock; it's right, too, twice a day.

If experience shows that experi-
ments are unreliable, theoretical progress will certainly be hampered as well. Many theorists will simply assert the experiments are in error when they conflict with a theory in which they have subjective confidence.

A careful examination of the field of atomic and molecular physics shows that theory and experiment are converging on one another slowly at best. The two developments proceed rather independently until the rare occasion when a critical and convincing test of theory can be made in an experiment that is demonstrably reliable.

With science in this condition— as it often is when the many-body problem is involved—we find theories agreeing with experiments after the fact, rather than before. As Professor Otto Oldenberg used to say, “All the phenomena of gas discharges have been explained, but few have ever been predicted.”

If standard reference data can be produced, their true test will be whether theorists are willing to abandon their work as invalid when it conflicts with such data.

My conclusion, therefore, is that in atomic collision data the problem is, how can we increase the useful information content of the literature at its source?

Clearly, if the useful information in the original literature can be increased by changing the habits of scientists and the way they write up their work, then the pressure on the information switching system will be greatly reduced, the problem of the critical compiler will be eased, the quality of the data available to the users will be increased, and the cost of producing the data in the first place can be cut dramatically.

For instance, it seems likely that about 1,500 papers will be written in 1967-1971 containing reports of cross sections for electron collisions. These will cost somebody about $30,000 apiece, or about $44 million. Many of the papers will surely have value, even if they contain no reference data that can be critically evaluated; their primary purpose may have been altogether different, so the cost of the data quoted is relatively small.

But, to the extent that the work has the measurement of such cross sections as its primary purpose, a substantial part of the 44 million dollars might be saved. How? Simply by not doing the work at all unless it is written up in such a way that it can be evaluated, and therefore become useful.

What can be done to improve the situation?

First, take heart from the fact that when critical compilers concentrate on the generation of logical criteria for judging whether certain measurement methods have been correctly applied, they are providing the tools with which the problem can be solved.

For, if subsequent investigators will make the necessary tests to confirm that the specified items are understood and correctly measured, their papers too will become capable of evaluation. And, if the information center scientists do a good job, their reviews will inevitably be read and referenced to the exclusion of much of the original literature, as Goudsmit said.

Undesirable as this anonymity may be, it will give the authors an incentive to make sure their written papers are an adequate record of what they did.

I can cite an example of a review that developed such a criterion, and it changed the quality of the work in the field almost overnight. Positive ion and electron recombination has classically been measured by examining the time dependence of the reduction of electron density in the afterglow of a discharge in a microwave cavity. If the electron density decayed inversely as the time, the conclusion was drawn that dissociative recombination was responsible, and competing effects of diffusion followed by wall recombination were small. Recombination coefficients determined this way were in very poor agreement.

Then, in 1960, E. P. Gray and Donald E. Kerr (3) published a criterion that related the range of electron-density decay over which the inverse time dependence must be demonstrated to the accuracy that could be relied on for the recombination coefficient. Thereafter, when data were presented at meetings, any speaker who did not indicate the extent to which the Gray-Kerr criterion was satisfied was sure to draw questions.

I hope to see more such evidence of a revival of sound scholarship in American science at the expense of cream skimming and of large numbers of hastily written papers.

Finally, we must look for alternative ways to preserve the detailed evidence that allows for a proper evaluation of an experimental result. Primary journals, like the Physical Review, are probably not capable of handling this amount of material, except when the property measured settles (or raises) an important problem of physics.

Perhaps, some day, investigators who measure such properties can be convinced to deposit such evidence, along with their results, directly with the appropriate information center for evaluation and dissemination, without going through the traditional archival literature at all.

References
1. S. A. Goudsmit, Is the Literature Worth Retrieving?, Physica Today 19, 9 (Sept., 1966)

WHERE WILL ALL OF THIS END?