In a number of the satellite systems, we have also included microfilming of the documents after indexing. The film is used in either fiche or cartridge form.

The operational flow is outlined in Figure 10.

SUMMARY

We have described a flexible approach for the input and retrieval of information from medium-sized document collections. This system is based on the use of the Jonker J-400 Termatrex drill. IBM cards are used to drive the drill, as well as generate printed term and document lists. This system can be implemented directly from the visual collation cards described earlier, and makes for a logical extension of retrieval techniques for growing systems.

In addition, the system makes it possible to control individual literature collections with a minimum of assistance from the Information Center. Two such applications are described, as is the simplified index work sheet on which they are based.

Conversion of the Termatrex-based systems to computer manipulation for large document files is in process and will be described shortly.

LITERATURE CITED

(1) Starker, Lee N., and J. A. Cordero, "A Multi-Level Retrieval System. I. A Simple Optical Coincidence Card System," J. CHEM. DOC. 8, 81–5 (1968).
(2) Jonker Corp., Gaithersburg, Md.

The Reliability of Property Data, Or, Whose Guess Shall We Use?*

LEE J. KIEFFER
Joint Institute for Laboratory Astrophysics, National Bureau of Standards, Boulder, Colo.

Received April 14, 1969

Increased reliability of property data can be achieved by critical evaluation of the techniques used in making measurements. The necessity for and the implementation of such evaluations are explored.

The two titles I have chosen for this article, when juxtaposed, appear to be contradictory. The message I wish to convey is that the level of reliability claimed for most data compilations is based on the experimenter's estimate of his systematic and random errors, and that in almost all cases his estimate of the systematic errors is a guess. I hope to convince you that this is true, and then explore a possible way of overcoming this obstacle to reliability.

The Joint Institute for Laboratory Astrophysics Information Analysis Center is concerned with data on low energy atomic collisions—cross sections, rate coefficients, etc. What distinguishes an Information Analysis Center from an Information Center is that the former concentrates on critical evaluation of the information or data it deals with. Data are selected on the basis of quality judgments which have been clearly outlined and defended. In our own center's program we are trying to push this evaluation procedure one step further.

The word reliability will be used here in a strict sense. A short definition of reliable data would be data which are presented with error bars which were chosen so that the probability of the "true" value lying outside of these limits of error is extremely small.

The ultimate user of data is not interested in why the data may be unreliable. The penalty he pays for using an incorrect value is not alleviated by the fact that the error was only a clerical one. Preserving the integrity of the data identified and assuring the user that it was obtained using current "best" measurement techniques are absolutely necessary to achieve reliability. There is another step possible, and that is assuring the user that these current "best" measurement techniques are reliable—i.e., they can measure what they claim to measure to the accuracy claimed. The last criterion is the most difficult one to satisfy in this long chain.

This problem was noted a few years ago by W. J. Youden, at that time a consultant to the National Bureau of Standards, in a paper on "Systematic Errors in Physical Constants," published in *Physics Today*.[1] "When two laboratories make independent determinations, each may attach to its 'best' value a $\pm$ sign followed by an estimate X of the error. This estimate of the error is often based upon a series of observations made under carefully controlled conditions. Experimenters soon discovered that if laboratories A and B reported values C_A and C_B for the same constant, the difference Δ between C_A and C_B was almost always a large multiple of their estimated error. Obviously, these calculated errors had no more to do with the real errors than the neatness of the laboratory or the promptness with which the investigator answered his mail."

Determining the reliability of a measurement is not a simple problem. It is clear that progress has been made in making measurements more reliably, but such progress is not easy. The kind of effort which has been lavished on fundamental standards and on constants measurements is the kind of effort necessary to achieve reliability in the quantitative measurement of physical properties.

Every experiment which purports to measure quantitatively a property of a well-defined physical system is based on a theory. This theory usually consists of a number of mathematical expressions which explain how to connect the dial readings of the instruments to the value attached to the property. For completeness, include in this scheme whatever theory is necessary to connect the instrument readings with the fundamental standards

* Presented before the Chemical Documentation Sessions, 4th Middle Atlantic Regional Meeting, ACS, Washington, D. C., February 12-15, 1969.

or secondary standards. In addition, there are a number of stated or unstated assumptions about systematic effects which in principle may affect the measured value, but are assumed to be negligible. I would define a reliable experiment in this way: A reliable experiment is one for which all known sources of systematic error have been delimited and their magnitudes properly assessed. This includes effects noted in the formal theory for the experiment and in the stated or unstated assumptions. Data resulting from such an experiment can be called reliable data. This is the best that can be done. It is, of course, a provisional judgment which can be overturned. If, on the other hand, the above definition does not hold true, then the theory for the experiment is defective; the data and associated error bars are a guess.

Recently I received an inquiry from a university professor about some ionization cross section data. He had found several numbers in the literature for a certain cross section and he asked my opinion. I reviewed the measurements and noted that in every instance the techniques used in making the measurements were defective—i.e., they had known sources of systematic error which had not been taken into account in the analysis. I had to reply that any opinion about the data was therefore a guess.

The dilemma facing not only the critical compiler but also the typical scientist in attempting to evaluate the material he finds in the scientific literature is clearly illustrated by the case of two recently published measurements of the cross section for ionization of atomic helium by electron impact. Techniques which had been unused for about 35 years were updated by two different laboratories. Several years of effort were expended in each laboratory, using large staffs of Ph.D's and modern equipment. The results reported by the two laboratories differed by about 25% and, of course, *their quoted errors were almost* a factor of ten smaller. Careful evaluation in our Center, including discussions with the authors, indicated that both experiments were defective.

Many of you will conclude at this point that the situation must be pretty bad in atomic collision physics, but of course in your own area things are in much better shape.

R. S. Marvin of the Rheology Section, National Bureau of Standards, sent me a copy of a talk he gave recently.[2] It rather nicely illustrates, I think, that this phenomenon is universal, but that the magnitude of the discrepancy between independent measurements varies.

"For several years we have been pursuing a project in our laboratory at National Bureau of Standards in which we are attempting to measure the viscosity of a liquid to 0.1%, using two absolute measurements which involve quite different kinds of flow. So far, we have been unable to obtain agreement between two such measurements to better than about 0.4%, despite the fact that both the reproducibility and our best estimates of the possible systematic errors in either measurement are less than 0.1%.

"We continue to measure and report viscosities to 0.1%, based on the value for water which has been generally accepted for several years now. This is justified on the grounds that the values so reported are almost surely valid to 0.1% as relative values, and for many purposes relative values are perfectly adequate. But if anyone really needs to know the accuracy of such values, I think they cannot count on anything better than $\pm 0.2\%$ on the basis of results up to this time."

It is unfortunate, but true, that the vast majority of data compilations made in the past have suffered from the fact that they were attempts to evaluate data from defective experiments critically, and therefore they had to rely on superficial criteria for evaluation. I don't believe anyone should be shocked by the notion that there are many properties we cannot measure reliably, and that we have been guessing in assigning values to such properties.

Our measurement techniques and our knowledge of what constitutes a valid measurement are advancing at such a rapid rate that our concept of how to compile, to evaluate, and to disseminate data must change. An archival record must be created in which enough published evidence exists for the user to determine whether a new measurement can be made which would improve the accuracy of the data, and, if not, what the accuracy is, based in the clearly stated theory of the experiment. Science and technology have, in the past, taken a rather relaxed attitude toward the problem of reliability. They have assumed that as knowledge and consequently measurement techniques advanced, new, better measurements could easily be made, and a lot of effort should not be expended on learning what is wrong with poor techniques. The evidence I have presented indicates that this assumption is no longer valid.

What we are trying to do in our information analysis center is to couple strongly the evaluation of the data presented with the evaluation of the measurement techniques used. We have sponsored, and will continue to sponsor critical reviews of measurement techniques with our own staff or outside consultants as authors. These reviews are published in the open literature. We hope to create in this way the documentation I mentioned previously. We also are developing objective, and we hope relevant, criteria by which we can judge the reliability of the measurement techniques. We are therefore concentrating on evaluating measurement techniques critically, looking for weak points in the theory of the experiments. Any systematic error whose magnitude is unknown, and hence cannot be compensated for, leads to an unreliable measurement.

I hope I have brought out the dilemma of high precision and low accuracy which faces us when we try to compile reliable data. Of course we cannot, in principle, guarantee that there is a solution to this problem. What we would like to achieve is an "air-tight" case. Unfortunately, as we all know, many an innocent man has been hanged.

LITERATURE CITED

(1) Youden, W. J., "Systematic Errors in Physical Constants," *Phys. Today* 14, 32–43 (1961).

(2) Paper presented at the Fifth International Congress on Rheology, Kyoto, Japan, October 7-11d1968.

This work was supported in part by the National Bureau of Standards through the National Standard Reference Data Program, and in part by the Advanced Research Projects Agency, monitored by the Army Research Office–Durham, under Contract No. DA-31-124-ARO-D-139.