Critical Data for Critical Needs

As a national laboratory dedicated to the advancement of measurement science and engineering, NBS/NIST recognized early in its history that the assessment of the quality of measurement results is a fundamental component of its mission. The 1981 paper by David Lide, *Critical Data for Critical Needs* [1], offers both a status report and a prognosis of major changes expected from the revolution in information technology that was beginning at that time.

The compilation and critical evaluation of material property data as an important NBS function can be dated to the Bureau's involvement in preparation of the *International Critical Tables* [2] in the 1920s. NBS played a leadership role in the project to produce this seven-volume series, whose contributors included hundreds of scientists from all parts of the world. The Editor-in-Chief was E. W. Washburn, Chief of the NBS Chemistry Division, and the Editorial Board included NBS Director G. K. Burgess. Many Bureau staff members contributed to the work, which is still used and cited 80 years later. Over the ensuing decades, NBS started new specialized data activities in areas such as phase equilibrium for ceramics (1930s), chemical thermodynamics (also 1930s), and atomic spectroscopy (late 1940s). However, the rapid build-up of government-supported science and engineering after World War II brought increased demands for a systematic program, fully integrated into the NBS research agenda.

NBS leaders such as Allen Astin, Edward Brady, and Lewis Branscomb worked with Congress and President Johnson's administration to establish the world's first formal government-endorsed data evaluation program. This effort culminated in the enactment of the Standard Reference Data Act of 1968 (PL 90-298, now designated as 15 U.S.C. 290). This Act of Congress established the National Standard Reference Data System (NSRDS), a coordinated and comprehensive program with an objective to "ensure that reliable reference data are easily accessible by scientists, engineers, and the general public." NBS was given the responsibility for coordination of the NSRDS, but other government agencies and private organizations were expected to participate.

In response, NBS set up a series of formal data evaluation centers that covered a wide range of physical, chemical, and materials disciplines. Joint projects were started with other agencies, professional societies, trade associations, and foreign laboratories. Several publication channels were established, in particular the *Journal of Physical and Chemical Reference Data,* which was published in partnership with the American Institute of Physics and the American Chemical Society. This journal served not only as an outlet for compilations done at NBS, but also attracted many data reviews and compilations from outside authors. Other scientific societies were engaged to publish compilations on specialized topics. As an example, four volumes of the *Crystal Data Determinative Tables,* prepared at NBS, were published through the Joint Committee for Powder Diffraction Standards [3].

The start of the 1980s saw NBS data activities on the brink of a new era. With data evaluation procedures for different types of data now well established, taking advantage of the computer revolution became the next obvious step. At the same time, major private sector organizations realized the advantage of working with NBS to improve the quality and accessibility of data of importance in their world of interest. As this new era began, the paper *Critical Data for Critical Needs* was published in *Science.* The genesis of this paper dates to the 1980 International CODATA Conference in Kyoto, Japan, where the keynote speaker was Phillip Abelson, then Editor of *Science.* Abelson was intrigued by the talks he heard on new technology for automated retrieval and dissemination of scientific data, as well as the evidence of growing international collaboration, and he offered *Science* as a vehicle for publicizing these developments. At once the paper documents the progress towards better quality data as well as the opportunity presented by computerization and by cooperative ventures. The paper sub-headings indicate its coverage: classes of data; the growing need for good data; quality control of data bases; the electronic revolution in data dissemination; and the need for cooperation.

The discussion could not be clearer, nor the vision better defined. What had been accomplished to date was impressive, but the future was equally promising.

The paper begins by creating a context for NBS data work, namely that of the "information explosion" of the last 25 years. In reaction to this exponential increase in scientific and technical information, numerous organizations had begun creating electronic collections of literature citations, abstracts, and even documents themselves in an effort to facilitate identification and retrieval of relevant information. However, the bulk of that work was concerned with helping users locate the scientific literature that reported research results. If users were to take advantage of the power of computers, the entire range of scientific content including *the data resulting from research,* which are usually in numerical or graphical format, would have to be available in electronic form. The remainder of the paper covered all major aspects of scientific data storage, retrieval, and dissemination.

After defining what is meant by the term "data"—repeatable measurements, observations and statistical results—Lide discussed the importance of quality in data activities. As science and engineering were confronting increasingly complex problems, from depletion of the ozone layer to shortage of critical materials, important decisions required data of the best quality to be available. With the advent of computer modeling as a major technology, data quality was even more important because of the potential for modeling results whose dependency on input data would be difficult to trace. The National Standard Reference Data System, as established by NBS, was a concerted attempt to involve expert scientists in the effort to assess the quality of reported measurements. As already mentioned, NSRDS maintained a series of continuing data centers with expertise in a well-defined discipline. Lide described several of these data centers and their current approach to evaluation of laboratory data on well-defined substances and materials. He also described parallel activities of the World Data Centers, which had been established by the International Council of Scientific Unions to perform a similar data quality assessment task for observational data in fields such as geophysics, oceanography, and atmospheric physics.

The paper then turns to the challenges of the electronic revolution and, somewhat surprisingly, defines the two paradigms for computer delivery of data that are still used today: installation on one's own computer and access via networking to remote data collections. At the time the paper was written, the first recognizable personal computers were just beginning to appear. Yet the essential features of scientific databases stored on PCs are all addressed, namely local control, heavy use, inclusion of search software, and the facility to transfer data to computational software and other applications. The description of online data services is equally prescient, even though it was based on a model of subscription services in which the user connects directly to a remote computer rather than through the World Wide Web of the year 2000. What is particularly interesting is the recognition that users routinely require a multitude of data resources to solve real-life problems, so that online data resources need to be integrated together to provide maximum impact and benefit.

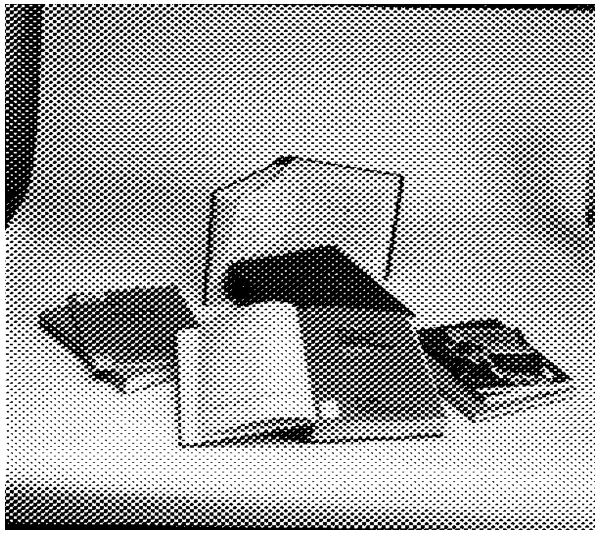

Fig. 1. A sample of current Standard Reference Data products.

The last major section of the paper describes a time of rapidly evolving cooperation to handle the flood of data and to take advantage of the electronic revolution. The cost of data evaluation is high, the volume of data too great, and the change in technology too rapid for any group to go it alone without ample resources. Further, in many cases, the need for consistency among collections of recommended data can be a prime motivator to work together rather than in competition. Lide mentions CODATA, the Committee on Data for Science and Technology of the International Council of Scientific Unions, as a major factor in bringing together data experts on an international basis.

The paper concludes with a prediction of three trends: the need for reliable data becoming more pressing; computer-based data dissemination methods, especially online systems, growing in use; and coordination in the development of computer-based systems being essential.

Unlike the majority of papers in this Centennial volume, this paper does not present original research results. Instead, it describes one of NBS's largest and best-known programs, the Standard Reference Data Program. More importantly, the paper describes quite accurately what challenges the NBS data programs were facing in the last three decades of the NBS/NIST first century. At the time of publication, NBS was beginning a series of intensive and large-scale cooperative programs that changed the availability and quality of data in many areas of interest to industry. The names of the programs suffice for description:

- American Society for Metals–NBS Alloy Phase Diagram Program
- American Ceramics Society–NBS Phase Diagrams for Ceramists Program
- National Association of Corrosion Engineers–NBS Corrosion Data Program
- Design Institute for Physical Property Data (DIPPR), in cooperation with the American Institute of Chemical Engineers
- NIST fluids property data programs, in cooperation with the Gas Producers Association, the Compressed Gas Association, and the Supercritical Fluids Extraction Consortium.

The trend continues today with NIST recently forming the Research Collaboratory for Structural Biology with Rutgers University and the University of California at San Diego to operate the Protein Data Bank.

In the years following the publication of Lide's paper, NBS/NIST expanded its data activities even further to include engineering data and even data used to "calibrate" statistical and other kinds of software. At the same time, the proliferation of personal computers, as envisioned by Lide, transformed forever the dissemination of NIST/NBS data. By 1995, over 70 PC databases were available for sale from NIST. Parallel efforts on traditional online services and the Internet/World Wide Web brought even increased availability. As of May 2000, NIST was already providing online access to its evaluated data via 15 web-based systems.

The basic components of the NIST data activities in 2000 remain essentially the same as defined in Lide's paper. Quality is the defining feature. The hallmark of NIST data work remains the evaluation by recognized experts. Easy availability is equally important so that the return on the taxpayer investment in data is maximized by the widest possible dissemination of data. Finally, cooperation ensures that limited resources are not wasted on duplicative efforts and that users are not confronted by competing claims for best quality.

Data evaluation remains an important component of the overall NIST measurement portfolio, providing a snapshot at a given time of the quality of measurement technology in different fields. NIST is the world's leader in the evaluation of physical science and other data, a tribute to the foresight and vision of leaders such as David Lide, Edward Brady, and many others. The paper *Critical Data for Critical Needs* remains a classic in defining the scope and importance of data in the advancement of science and technology and of the impact computers have had on scientific data work.

David R. Lide was hired by NBS in 1954 to set up a microwave spectroscopy laboratory in the Thermodynamics Section of the Heat and Power Division. In the early 1960s he led the integration of NBS research programs in infrared, microwave, and ultraviolet spectroscopy into a single Molecular Spectroscopy Section, which he headed until he became Director of the Office of Standard Reference Data in 1969. He was active in various national and international organizations, including stints as Secretary General and later President of CODATA and President of the Physical Chemistry Division of the International Union of Pure and Applied Chemistry. He received Department of Commerce Silver and Gold Medals and the Samuel Wesley Stratton Award of NBS for his research in spectroscopy, as well as the Herman Skolnik Award and the Patterson-Crane Award of the American Chemical Society for contributions to chemical information. After leaving NIST in 1988, he became Editor-in-Chief of the *CRC Handbook of Chemistry and Physics* and has published several other books and electronic databases.

Prepared by John R. Rumble, Jr.

Bibliography

[1] David R. Lide, Jr., Critical Data for Critical Needs, *Science* **212**, 1343-1349 (1981).

[2] Edward W. Washburn (editor-in-chief), *International Critical Tables of Numerical Data, Physics, Chemistry, and Technology,* Vols. 1-7, published for the National Research Council by the McGraw-Hill Book Company, New York (1926-1933).

[3] J. D. H. Donnay and Helen M. Ondik (eds.), *Crystal Data Determinative Tables, Vol. 1-4,* Third Edition, National Bureau of Standards and Joint Committee on Powder Diffraction Standards, Swarthmore, PA (1972-1978).