

OMNIDATA and the Computerization of Scientific Data

For much of its history, the National Bureau of Standards has been a leader in providing the most reliable available technical data to scientific and engineering users. The Bureau has operated large-scale data evaluation programs in a wide variety of disciplines and has distributed the results through various publication channels. With the advent of modern computers, it was natural for NBS data experts to explore how computers could be used both for internal data management activities and for the public dissemination of NBS data collections. During the 1970s, the Office of Standard Reference Data (OSRD) and the various data centers that it coordinated embarked on a number of projects aimed at utilizing the growing power of digital computers to improve the efficiency and effectiveness of the SRD program. NBS quickly became recognized as a pioneer in this area.

At that time, there was a tendency for each data group to develop computer programs to handle the particular type of scientific data of concern to it. Since the SRD program covered such diverse classes of data, there was an incentive for OSRD to investigate the possibility of developing general purpose programs capable of managing a broad range of data. This effort led to the creation of the database management program OMNIDATA, described in the book *OMNIDATA, An Interactive System for Data Retrieval, Statistical and Graphical Analysis and Data-Base Management: A User's Manual* [1], authored by Joseph Hilsenrath and Bettijoyce Breen. This represents the first major publication of the National Bureau of Standards in the area of computerized scientific data management.

In the early 1970s when the OMNIDATA project was begun, data retrieval, data analysis, and data file maintenance were largely done in the batch mode, with programmers submitting punched cards for processing on a mainframe computer and, after an indeterminate period of time, receiving a printout of results. Also, computer programs designed to run routinely were almost always inflexible, and the writing of ad hoc programs to answer specific questions often entailed expense and delay out of proportion to the urgency of the problem that motivated the question in the first place. When early commercial general-purpose data management programs began to appear, primarily intended for business applications, they had adequate and roughly comparable search and retrieval capability, file defini-

tion features, and report generators. None, however, had enough data analysis and data manipulation facilities to handle the numerical and alphanumeric data files required for scientific use, nor did they address unique characteristics of scientific information, such as Greek and mathematical symbols, uncertainties, and varying number of significant figures, all of which added to the complexity to scientific data.

OMNIDATA was designed to overcome many of these shortcomings. Programs were written to analyze and store data. Rudimentary graphics packages created visualizations of data. The 45 unique modules and the supervisory program of the OMNIDATA system could be used not only by the computer professional, but also by the novice with little or no knowledge of computers. It could be run in demand mode, interactively from a computer terminal on a time-shared computer system. In addition to handling such administrative files as personnel, training, inventory, and travel, the system was robust enough to handle diverse scientific data files, including crystal structure data, thermochemical data, chemical kinetics data, and data on physical properties. Finally, OMNIDATA was designed and written to be modular, a hallmark of truly efficient computer programming and systems design, and a visionary precursor to what is today called component-based software.

Most of the individual modules provided tools for data retrieval, analysis, and reporting. Specifically, the system had facilities for searching and reporting; plotting and graphical analysis; arithmetic operations in general, and statistical analysis in particular; file partitioning and subsequent sequential analysis on subfiles; keyword indexing of bibliographic files; flagging, coding, and decoding of data items; analysis of questionnaires and surveys; and a large variety of data management and validation routines of use to both the end user of the data and the database administrator. An important feature of OMNIDATA was its ability to convert data files from other formats for use in the OMNIDATA system and to generate data arrays that could be accessible to other programs written in languages including FORTRAN, COBOL, and X BASIC. Also, the OMNIDATA system interfaced with the OMNITAB II program [2], used extensively at NBS and elsewhere, to provide a repertory of well-tested and highly accurate statistical routines.

The OMNIDATA system allowed for flexibility in file size. With it, one could profitably automate files from those with few records of a small number of data elements to elaborate databases of many records with numerous data fields. In contrast to other management information systems of the era, which limited the number of data items that could be searched or manipulated, each module in OMNIDATA was capable of operating on every data item, and the user could search on any field, even down to the character level. Its modularity, flexibility, and data analysis capabilities made OMNIDATA unique. It was a forerunner of the database management systems marketed today and the systems so essential to the functioning of the World Wide Web.

The use of OMNIDATA quickly spread to other parts of NBS, beyond the Standard Reference Data Program, so much so that there were continuing problems in providing support and extensions for a diverse group of users. The legacy of OMNIDATA was not the software itself, but the realization that scientific data projects could successfully convert to computerized operation—and that they would have to do so in order to continue as viable projects.

The lessons of OMNIDATA were critical to the push to computerize both the internal data operations and the dissemination of databases, which began in earnest in the 1980s. The effort was led by Bettijoyce Breen and John Rumble, Jr., who joined the Office of Standard Reference Data in 1980. During this decade every NBS/NIST data activity created databases of references containing the data of importance to its area of responsibility. Many developed specialized data entry programs that captured not only this bibliographic information, but also the numeric tabular and graphical data contained therein. In addition, the data centers developed suites of software that supported data evaluation through the use of discipline specific analysis, statistical procedures, and correlation techniques. At the beginning of this effort, most of the software incorporated the ideas found in OMNIDATA, with significant extensions to take care of specialized requirements.

Many of the data handling software packages developed at NBS were used by outside organizations. The NIST Crystal Data Center developed AIDS 80, a powerful package that evaluated and managed crystallographic data [3]. The NIST Alloy Phase Diagram Data Center created a suite of graphical digitization and database management tools that supported the international Alloy Phase Diagram Program run jointly by the American Society of Metals and NBS [4]. A similar set of graphics software was developed for handling ceramics phase diagrams under the NBS-American Ceramic Society Phase Diagram for Ceramists Program [5].

The computerized dissemination of NBS collections of evaluated data proceeded likewise. In the late 1970s, NBS worked with the Environmental Protection Agency and the National Institutes of Health to create and operate the Chemical Information System (CIS), which was the first online system to provide scientific numerical data. The CIS featured a powerful substructure and nomenclature search system that allowed users not only to search for data on a specific substance, but also to identify classes of chemical compounds with particular structural features. The CIS integrated databases built by many groups, including NBS thermochemistry and crystallographic data centers.

At the same time, the Standard Reference Data program began offering magnetic tapes of formatted data files suitable for outside users to load onto their own mainframe computers. It was the responsibility of the users to build their own search software and to manage the data. The PC revolution of the 1980s changed all that, and OSRD quickly began offering files similar to those on magnetic tapes on floppy disks. It soon realized, however, that users wanted self-contained packages that were easy to install and that included built-in user interfaces. By 1985 two systems were under advanced development. Steve Stein, then of the Center for Chemical Physics, was building an MS-DOS mass spectral data system, and Chuck Wagner of Surfex and John Rumble, Jr., of NBS were building an x-ray photoelectron spectroscopy (XPS) database on an Apple platform. The mass spectral database was released in 1987 and became an immediate success. Today it is incorporated into virtually every mass spectrometer sold. The next two years saw many other NBS databases released including the XPS database.

In 1982, NBS scientific database management efforts addressed another area, that of materials data. In November of that year, NBS cosponsored with FIZ Karlsruhe (a German government technical information center) and CODATA (the ICSU Committee on Data for Science and Technology) a workshop on computerized materials data. This Fairfield Glade meeting, named after its location in Tennessee, spawned an international activity to build an online materials data system. The proceedings from that workshop, *Computerized Materials Data Systems* [6], edited by Jack Westbrook and John Rumble, Jr., were widely circulated worldwide and became the “bible” for planning such an online system. The seminal meeting addressed the full range of topics related to online materials data systems, including their scope, user interfaces, system development, legal and economic issues, and barriers to be overcome. The Fairfield Glade Workshop concluded that it was not only possible to build such a system, but that one should be built as soon as possible.

Over the next five years, over 15 additional workshops were held covering the full range of engineering materials disciplines, potential user industries, and many other aspects. A prototype system was built with funding from NBS, the Department of Energy, and the Department of the Army. The Metals Properties Council, a trade association of the metals industry, worked with NBS and other groups to establish the non-profit Materials Property Data Network (MPDN). The MPDN, which soon became part of the STN dial-up network of the Chemical Abstracts Service, was a successful online data system in the pre-Internet era.

One offshoot of this effort was the establishment of ASTM Committee E49 on Computerization of Materials Properties Data. This committee, which was chaired by John Rumble from 1988 to 1993, rapidly became the international focal point for standards relating to all aspects of the computerization of materials data, and these standards are in wide use today.

The blossoming of NBS/NIST computerized data dissemination continued unabated until about 1994, when two concurrent changes hit the computer world. These were the release of the Microsoft Windows operating system and the Internet explosion. In one short period, the NIST PC databases and older online systems such as the MPD Network became obsolete as users demanded Windows versions of existing MS DOS data products. In addition, the Internet, and especially the World Wide Web, revolutionized online data delivery. Where previous online systems required years of development, the Web required only months of work. At the present time, NIST operates 15 Web-based data systems that receive thousands of users every day.

Today's scientific database environment is a long way from the mainframe world of 1976 when Hilsenrath and Breen issued their first manual for OMNIDATA, and even of 1986, when the first MPDN prototype came online. These changes have made the creation of databases considerably easier and have brought easy access for users worldwide to the full range of NIST evaluated data. The ideas in *OMNIDATA* and *Computerized Materials Data* were germinal in articulating the core principles of scientific database management and online dissemination. The distance traveled since then is a tribute to Hilsenrath, Breen, and Rumble and their foresight in leading NBS to the forefront of computerizing scientific data.

Joseph Hilsenrath began his career at NBS in 1948, first as a scientist/mathematician specializing in employee training, then as an experimental researcher in

high pressure physics, after which he became Chief of the Equation of State Section of the Heat Division. As the first computers became available at NBS, he became interested in the preparation of tables of thermodynamic data and in putting computer know-how in the hands of Everyman. His OMNITAB statistical packages received wide use in NBS and other Government agencies. He moved to the Office of Standard Reference Data in 1967, where he led the development of computerized typesetting methods that saw use throughout the Government. He retired in 1974, but for many years remained active in NBS reference data activities.

Bettijoyce Breen joined NBS in 1969 as a computer programmer for the Office of Standard Reference Data after receiving her B.S. in chemistry from William and Mary College. After completion of the work on OMNIDATA, she supervised the introduction of the next generation of computerized typesetting technology into NBS and later became head of the OSRD Data Systems Development Group. This group helped the various data centers in NBS and elsewhere to automate their operations and oversaw the preparation of database packages for public dissemination. She received an M.S. in Chemical Information from American University in 1975 and became active in the Chemical Information Division of the American Chemical Society, serving as both Treasurer and Chair of the Division. In 1990, as Bettijoyce Breen Lide, she joined the NIST Advanced Technology Program (ATP), where she set up the first ATP management information system. As an ATP program manager, she established the program on Information Infrastructure for Healthcare and later headed the Information Technology and Applications Office. In 1999 she received the George Uriano Award of NIST for fostering industrial-government interactions.

John Rumble, Jr. joined NBS in 1980 as the OSRD program manager for materials data. In addition, he played a strong role in computerizing NBS/NIST data activities in all disciplines. Rumble became Chief of the Office of Standard Reference Data in 1994. He has been active in many activities related to standards for scientific and technical information and was elected President of CODATA in 1998. He is also a Fellow of the ASM International and the American Society for Testing and Materials.

Prepared by John R. Rumble, Jr.

Bibliography

- [1] Joseph Hilsenrath and Bettijoyce Breen, *OMNIDATA, An Interactive System for Data Retrieval, Statistical and Graphical Analysis and Data-Base Management: A User's Manual*, NBS Handbook 125, National Bureau of Standards, Washington, DC (1978).
- [2] Joseph Hilsenrath, Guy G. Ziegler, Carla G. Messina, Philip J. Walsh, and Robert J. Herbold, *OMNITAB: A Computer Program for Statistical and Numerical Analysis*, NBS Handbook 101, National Bureau of Standards, Washington, DC (1966); David Hogben, Sally T. Peavy, and Ruth N. Varner, *OMNITAB II: User's Reference Manual*, NBS Technical Note 552, National Bureau of Standards, Washington, DC (1971).
- [3] A. D. Mighell, C. R. Hubbard, and J. K. Stalick, *NBS*AIDS80: A FORTRAN Program for Crystallographic Data Evaluation*, NBS Technical Note 1141, National Bureau of Standards, Washington, DC (1981).
- [4] Thaddeus B. Massalski (ed. in chief), *Binary Alloy Phase Diagrams, Vols. 1-3, Second Edition*, ASM International, Materials Park, OH (1990).
- [5] Richard M. Spriggs and Stephen W. Freiman ACerS/NIST Phase Equilibria Update, *Am. Ceram. Soc. Bull.* **75** (1), 83-85 (1996).
- [6] J. H. Westbrook and J. R. Rumble, Jr., *Computerized Materials Data*, National Bureau of Standards, Gaithersburg MD (1983).