Socioeconomic Aspects of Materials Data: Serving the User

Anthony J. Barrett†

ESDU/CODATA, 6 Deanway, Chalfont St. Giles, Buckinghamshire HP8 4JH, U.K.

Received June 4, 1992

Numeric and factual data on materials are used, together with physical laws and other data, directly in the decision-making processes of engineering design, production, performance assessment, and scientific research. These processes are conducted by human experts or by computers running under their supervision. The people involved spend most of their time doing things other than retrieving information or data. They have needs and make demands on the systems that they use that can easily be overlooked by those who provide, manage, and operate the systems; so relationships between data users and data managers can be rather indiferrent. At an earlier stage, the way in which data generators and data system managers are associated can have a important influence upon the quality of the data included in database and expert systems. Relationships between data users, data managers, and data generators are governed by a social climate that has changed little in the last 30 or so years. Systems that are now starting to provide the essential links between these communities will affect this climate and that in the wider society which they serve. Numeric database systems in general, and those on materials data in particular, have features that are quite different in several important respects to those familiar in conventional text storage and retrieval practices. This paper looks at these differences, at some of the needs of numeric data users, and at the operational, convenience, added value, quality, and benefits features that they expect. Economic consequence of serving these needs and some of the wider social implications are explored.

DATA IN THE HUMAN ENVIROMENT

Characteristics of Numeric Data. "Data are scientific or technical measurements, values calculated therefrom, observations, or facts that can be represented by numbers, tables, graphs . . . and which are used as a basis for reasoning or further calculation".[1] This widely accepted definition of the term 'data' raises important issues.

Scientific and technical data are actual measurements, usually numeric in form; they are not, within the context of this paper at least, references to where data might possibly be found. Numeric data are quite clearly distinguished from other types of record that are sometimes called "data", as for example, "textual data". Further, as the definition implies, data are applied as the actual basis of some form of technical adjudication by human beings or as the basis of some further calculation. This calculation will be made by a human or by a computational system such as would be the case in computer-aided engineering. In all cases, a human is closely involved, either applying or supervising the application of data. The next step after numeric data have been found is their immediate use in a design for an engineering artifact, in the critical analysis of an existing artifact or product, in a tool to be used in research, as the basis of a plan for the management of a resource, or for any of a wide range of other practical purposes.

Errors in Handling Data. "Engineering products are only as good as the data upon which they are based",[2] may be a slight overstatement—but they are certainly no better! The same might be said of all possible consequences of the use of data; so there is a preoccupation with the quality of every aspect of numeric data. Take by way of example the accuracy level to which numeric data must be stored and represented compared with that which is adequate for text. Some years ago I asked the chief executive of a sister company, which provided an on-line, full-text, legal information service, what was the accuracy achieved by operatives keying text into his "database". He told me that, working individually, they achieved an average accuracy of about 99.5%. That did not sound too bad until I looked at it another way. An accuracy of 99.5% is the same as an average error rate of 1 incorrect character in every 200.

A text sample of about 400 characters, with an accuracy of 99.5%, is illustrated in Figure 1. One thing to note is that the errors are quite obvious in the work 'machine' in the second line and the word 'novel' in the fifth. Further, even if the errors go unnoticed, very little harm is likely to follow as a result of using this faculty text.

Now contrast what we have just seen with a sample of numeric data of similar presentational accuracy.

Numeric data, reproduced with an accuracy of about 99.5%, are illustrated in Figure 2. The table was produced from data presented graphically,[3] and one digit in the approximately 200 numeric characters that the sample contains was misreported. Even knowing that there is an error in these numeric values, can we spot it? Perhaps it is at $\theta° = 45$, $a/b = 3.5$—the value looks out of place in a row of values that otherwise falls continuously? Or is it at $\theta° = 30$, $a/b = 1.75$—this looks like a repetition of the previous value in the row? In fact it is the first value in the body of the table that is incorrect; instead of 23.20 it should read 13.20. This error can not be found by simple inspection such as that which led us immediately to the discovery of the errors in the sample of text (Figure 1). Nor would it be discovered by using techniques that will sometimes reveal errors in well-conditioned data, for example, by taking differences, or differences between differences, among every pair of numeric values in the table. It is even unlikely to be revealed by making a careful graphical plot across all the data tabulated. The graphs from which these data were taken are discontinuous, each with several cusps and minima; these could not be correctly reproduced from the data tabulated in Figure 2. A faulty value of this magnitude, which remained undiscovered, might lead to the design of a structure that could collapse when it encountered a load little more than half that for which it had been designed! So, if proof testing or other safeguards failed to reveal the error, incalculable

† Chairman, CODATA Task Group on Materials Database Management.

0095-2338/93/1633-0022$04.00/0

```
ABSTRACT: Computerized materials property databases and networks
are at the leading edge of the introduction of machime readable,
numeric data systems in technology.  Managers of such systems are
faced with pioneering the acceptance both of a new product,
namely, high quality data, and of a nivel means for its delivery.
They are well able to catalog the features of their systems, but
some of them .......
```

Figure 1. Sample of text with an error rate of 1 in 200.

Buckling Stress Coefficients, K, for Swept Parallelogram Shaped Panels

a = length of panel, b = width perpendicular to a, θ^{o} = angle of sweep
t = plate thickness, f_b= critical compressive stress, E = plate modulus

K = buckling coefficient defined by f_b= KE{t/b}^{2}

a/b / θ^{o}	1.50	1.75	2.00	2.25	2.50	2.75	3.00	3.25	3.50	3.75	4.00
60	23.20	9.72	8.55	8.22	7.17	6.59	6.49	6.01	5.60	5.43	5.25
45	7.52	6.59	5.99	5.71	5.34	5.02	4.80	4.67	4.72	4.52	4.39
30	5.62	5.62	5.00	4.65	4.59	4.42	4.25	4.19	4.17	4.10	4.08
15	4.63	4.30	4.18	4.15	4.24	4.20	4.17	4.09	4.15	4.09	4.00
0	4.01	3.76	3.60	3.61	3.71	3.62	3.60	3.60	3.68	3.60	3.61

Figure 2. Sample of numeric data with an error rate of 1 in 200 in their presentation (for illustrative use only).

consequences in terms of human peril, financial loss, and corporate exposure might follow.

Figure 2 also illustrates the considerable amount of metadata that has to accompany numeric data in order to make them intelligible. In fact, more than that which is shown in Figure 2 would have to be provided in order to use the values of *K* that are tabulated.

In both of the examples above, the quality of the data might have been improved by double keyboarding. This would reduce the error rate and improve the appearance of a textual database. It would not of itself be an adequate procedure for numeric data; other measures have to be taken to underwrite an acceptable level of accuracy in numeric data recording and retrieval. Two principal classes of numeric data relating to materials describe the properties of materials in the one case and the mechanics of materials in the other. Those used in the illustration of Figure 2 are of the latter class. They describe a phenomenon, buckling, that is one of many in which any type of material might be involved. Phenomenological data, which are extensively encountered in engineering and technology, sometimes describe near chaotic circumstances where small changes in initial conditions have quite disproportionate consequences, and these may even be the reverse of those that might be expected. This gives additional significance to the quality features that numeric data must reflect. Some of these will be described later.

Consequences of Quality in Different Systems. The above examples illustrate some fundamental differences in the general requirements of systems that store and manipulate textual data and those intended to store and manipulate numeric data. They also illustrate some of the consequences of a lack of quality. In the case of textural databases, the absence of essential qualities may result in unnecessary cost, frustration, and, at worst, a breakdown in communication; the ultimate decision-making process is effectively insulated from the consequences of low quality and error. By contrast, in the case of numeric databases, that which is retrieved is by definition the final step in approaching a decision process that will directly and immediately affect a scientific judgment or pragmatic venture.

At some point in the future, computerized materials databases will interface even more directly with computerized design and manufacturing systems than at the present. Human intervention in the design process will be replaced by human supervision and eventually by computerized human supervision or, in other words, some form of expert system. Quality parameters will have to be particularly well-established when the data are no longer routinely available for inspection. All this is of crucial importance to the user of numeric data and most particularly to the user of materials property data. This is because materials data are dependent upon so many variables. These relate to the specification of the material itself, the form into which the material has been manipulated, the physical loadings and environment to which it has been exposed, and that to which it was exposed when the data were recorded, and many other such parameters. Each variable has to be precisely specified and accurately measured and recorded.

ASPECTS OF NUMERIC DATA QUALITY

Numeric data are meaningless in the absence of some appraisal, however vague, of their quality. We have seen how the accuracy to which numeric data are presented can affect their worth in application. The accuracy with which data are represented is but one quality, and there are many more qualities that need to be indicated. A guide for the formatting and the use of material property and chemical property data and database quality indicators has been proposed as a new standard.[4] It recommends a range of quality indicators that relate to

Data Quality:
- Source—handbook, government, producer, etc.
- Statistical basis—in terms of confidence levels for different guaranteed values.
- Material status—in production, experimental, or obsolete.
- Evaluation status—extent to which basic scientific integrity of data has been established and by whom.
- Validation status—extent to which applicability of data to specified operations has been established and by whom.
- Certification status—extent to which data values have been recognized by a warranting authority (such as FAA, Department of Defense, etc.).

Database Quality:
- Support status of database—if and how supported.
- Completeness of information—material form, condition, and processing.
- Test procedures—if standard test procedures, or not, or if derived from service experience.

The proposed standard includes informative guidelines on the use of the above indicators. The following additional criteria, relating to the operation of the database itself, should not be overlooked.

Data Operation:
- Accuracy and precision—with which data are stored and retrieved.
- User-friendliness.
- Availability—unrestricted, proprietary, government classified, etc.
- Distribution medium—on-line, floppy disc, CDROM, etc.
- Payment arrangements—on-line charge, subscription, etc.
- Access source—details of supplier.

A similar list could be prepared for phenomenological data on the mechanics of materials.

These quality indicators should not be thought of only as expressions of the technical excellence of the data and databases to which they relate. Technical excellence of itself could not justify the extremely high costs that are entailed in order to assess and control these qualities. There have to be strong social and economic justifications for such expense and effort.

SOCIAL IMPACTS OF NUMERIC DATA QUALITY

Response within Science and Technology. Striving for and attaining data quality has impacts within the scientific and technological community on those who use data and apply them and in the wider society that they all serve. At one time the quality of data was of concern to those who generated them and to few others. It was Weinberg,[5] who first drew attention to a division of labor that was taking place between those who discoverd facts and those who sifted, absorbed, and correlated them. Some of us, who started our professional careers around the end of World War II, found ourselves involved in this division of labor and in a social layering of science that Weinberg also predicted. As a young man, I was concerned with substantiating the quality of data in the engineering sciences. Sometimes I had to be 'put in my place' by some of my scientific elders and betters who viewed my data evaluation activities as little more than an impudent reflection upon their personal integrity! Within the scientific community itself the concern over quality had, and to some extent still has, social undertones in the relationship between the various groups of people who are involved in handling data, although the social impacts of data quality now extend beyond the boundaries that were foreseen by Weinberg. Today there is an expanding appreciation of the advantages to be gained from basing decisions on sound facts, and there is a wider awareness of the possible consequences of those decisions. Qualities beyond those established during the generation and the evaluation of data are now also of importance.

Response from Industry and the Consumer. Data such as numeric data on materials flow from their generation, in laboratory tests and model analyses, through the processes of evaluation and validation to their application along with empirical experience by users. Typically, these users are engineers and technologists in industry, and products and services are the result of their work. The reactions of those in industry who fund and trade in these products and services, those who labor to produce and operate them, and those who consume them are transmitted primarily through the market and through influence on legislation and on standarization.[6]

Data Users and Others in Industry. Gone are the days when engineering designers could cover the shortcomings of their knowledge by the profligate use of relatively cheap materials from secure sources of supply. Even so, there are some in industry who still view much of the care taken in establishing the quality of data as being misplaced; certainly many of the data that they actually use are not of the quality assured by skilled evaluation and validation. The attractions of these data are sometimes only that they were quickly available or that they cost little or nothing to obtain! But corporate exposure to the possibility of failure in the marketplace, expensive product recalls, and the price of major liability suits are the real cost of such data.

Labor organizations, particularly those concerned with transport, mining, and handling of hazardous materials, retain highly competent technical advisers who are well aware of the implications of data quality; they bring their influence to bear on legislation and in standardization bodies. Well-publicized examples concern the alleged effects on health of working in close proximity to asbestos and other substances and in proximity to electronic equipment. Labor union representatives, employers, and legislators now find themselves becoming more involved in the debate over the quality and interpretation of the relevant data than was once the case.

Consumer Reaction. The quality of the data used as the basis of comparing acceptable levels of hazard, inconvenience, and other social costs with the benefits offered is a matter of concern to an increasing number of consumer pressure groups monitoring the products and services of engineering, medicine, health, and other fields. The disquiet of consumers over inadequate or hazardous products and services is ultimately exorcized through market forces, although reaction via legislation and standardization is becoming more common. Such reaction has been highlighted by changes in product liability law. In product liability suits, the quality of the data used in the design of the product may be an issue. Community associations concerned with such matters as pollution, the location of hazardous plant, or waste disposal now possess or can retain expertise that can effectively challenge the quality and interpretation of scientific data.

Social Aspects of System Quality. The care with which numeric data are 'packaged' and made available influences the user's perception of their quality and the way in which they may be applied and relied upon. Smartly tabulated and graphed data in a well-presented hardcopy report have always been trusted more readily than those that are proffered in a slovenly or inconsiderate manner. Such familiar touchstones of quality are lost when data are held in computerized systems. This can induce fear and mistrust that will be aggravated if the system software is unreliable or is not 'user-friendly'. Further, the higher probability of error in retrieving data from systems that do not possess this quality affects the integrity of decisions based upon such data and, consequently, the exposure of the decision maker to loss of competitive position or to product liability litigation.

There is another aspect of system quality and user-friendliness that illuminates further the difference between textual and numeric 'databases'. The former are frequently used by information scientists and similar professionals who make frequent searches of the literature. They become familiar with the command languages and structure of the database and with the most efficient search techniques. Numeric databases, by contrast, are most likely to be used relatively infrequently by a wide range of engineering and scientific professionals seeking numeric data that they will apply themselves. They place a high value on systems that are easy to use and reliable and that include value-added software for, among other purposes, terminology translations, unit conversions, graphical representations of data, and comparison of competing materials. Numeric database systems that fill these requirements have to embrace human/computer interactions across a series interfaces that is more extensive and complex than in the case of a textural database. The successful interfacing of eye and hand with VDU and keyboard has to be extended to standards, nomenclature and units, system and software documentation, pointing and graphics devices, and the actual physical system that corresponds to the data being retrieved and manipulated.[7]

ECONOMIC ASPECTS, BENEFITS, AND THE MARKET

Motivations. It is difficult to demonstrate, in incontrovertible terms, the relationship between the cost of quality numeric data and the benefits that such data will bestow. Cost–benefits issues have been addressed many times in the past. For example, estimates have been made of the potential national costs or of the costs of company exposure to losses that might result from a failure to apply established engineering knowledge and data. With the same goal, but on a slightly different tack, many case histories have been collected relating to actual losses that have been traced back to a failure to apply known information. Invariably the sums of money involved are spectacular.[8,9] Even so, these dramatic results have done little to induce a widespread thirst for quality data or a preparedness to pay for them or to pay for the development of systems to disseminate them. The limited impact made by loss estimates and case histories has been further commented upon in refs 8 and 10.

The picture is not all gloomy. There are trends that are helping to promote the creation and use of high-quality data. More realistic prices for 'traditional' materials and interest in an expanding range of new materials are encouraging the more careful selection and critical use of materials both for established and for new applications. This in turn leads to a more intelligent appreciation of the value of reliable materials data. Other factors having a positive influence on the achievement of higher standards of quality in numeric data have already been noted at several places throughout this paper. These factors included societal mechanisms, the demands of labor and the consumer, management preoccupation with corporate exposure to product liability litigation, the response of the market, and more. My experience suggests that changes in attitudes induced, for example, by growing competition from newly industrialised nations, will achieve more toward the enhancement of the quality of the data actually used in the working world than any amount of governmental or institutionalized propaganda. Customary mechanisms, such as the production and promulgation of formal standards, have a part to play, but they will be of limited impact until the interaction between the quality of data and the fortunes of those who use them are more generally recognized; such standards may in any case reflect a quality level some way below that to which the market leaders would wish to aspire.

Interestingly it is the "younger" or more recently industrialized nations that may influence matters most. Many of them have a well-founded appreciation of the value of high-quality numeric data. The resolve of professionals in these countries to obtain and to apply up-to-date and comprehensive resources of materials data of proven quality is very evident[15] and has relatively strong national support. The more established industrial nations, which are generally in the best position to obtain and manipulate these resources, may do themselves a service by encouraging wider international collaboration in their exploitation and use. Dr Peter Bridenbaugh, Aloca's Executive Vice President, Science and Technology, and Director of the Alcoa Technical Center, in a discussion of what is needed to shorten the time between concept and commercialization of industrial products[16] includes the need to build electronically accessible databases stocked with materials properties. He concludes his summary with: "And finally we must eliminate parochialism from our thought processes. No single company, industry or even nation possesses the requisite intellectual or financial resources to succeed alone."

Spreading the Costs. Inconsistencies in sets of materials data, the lack of adequately defined metadata, and many other weaknesses in materials data compilations have long been tolerated in passive storage media such as handbooks and other paper systems operated by human data users. Attempts to improve matters have had only limited success. These shortcomings cannot be tolerated at all when the storage medium is part of a costly system which, when used, for example, in integrated computer-aided engineering, will dispense with human intervention almost completely. Because of the time and costs involved in the electronic capture, storage, retrieval, manipulation, and communication of materials data, there are strong incentives to establish the quality of those data, to be efficient in their storage and retrieval, and to share and distribute them over a wide market. Opportunities to spread costs and to increase revenue by exploiting domestic and international data interchanges make further demands on the quality of numeric data and database systems. Materials have to be uniquely identified, data have to be recorded consistently, terminology must be defined coherently—often in several languages—formats for the exchange of data must be standardized, and much more.

Marketing. At the same time that we are finding a need to promote high-quality data, particularly in many small and medium-sized businesses, the computerization of numeric data is gathering pace. The materials database manager is thus faced with introducing potential clients to—what are to them—the novel ideas and advantages of refined data at the same time that they are being introduced to an unfamiliar means of delivery. Many marketing experts see an attempt to introduce a new product at the same time as a new method of delivery as a sure recipe for disaster!

To a large extent, high-quality numeric data and the systems being devised for their storage and delivery are interdependent. Marketing them has to pay close attention to identifying benefits in sound practical terms and to the building of good relationships between the groups who produce them, those who operate them, and those who use them.[11,12] The identification of key benefits and of the true production costs of materials database systems is being addressed as part of the program of the CODATA Task Group on Materials Database Management,[13,14] by qualifying them in a structured way. It is hoped that the approach that is being used for the qualification of benefits in terms which are familiar to the user and that relating them to a structure of system functions and features as seen by the database manager will help to build bridges of communication between these two communities.

CONCLUSION

Technical numeric data and databases are increasingly underwriting the levels of safety, convenience, and tangible benefits that society in general expects to enjoy. Within the scientific and technical communities that are involved with these data, social rearrangements have been taking place, socioeconomic barriers are being recongized,[11] and user expectations are being sought and responded to.[12] Social tensions that arose in the aftermath of the division of labor between data generators and evaluators have started to subside.[6]

The quality of the benefits that society expects the scientific, technological, and industrial communities to provide can be no better than the quality of the data and other tools that they use. Quality, in this context, implies not only scientific accuracy and integrity but also many other features that those who use numeric data need and they they increasingly expect to be provided. Materials are fundamental to almost every aspect of the life of any nation. Although those who provide numeric data and database systems may never be completely market driven, they most certainly need to be highly market responsive and to take into account society's perception of what constitutes its own well-being locally, nationally, and globally.

REFERENCES AND NOTES

(1) Westbrook, J. H.; Grattidge, W. A Glossary of Terms Relating to Data, Data Capture, Data Manipulation and Databases. *CODATA Bulletin*; Hemisphere Publishing Corp.: New York, 1991; Vol. 23, Nos. 1–2.

(2) ESDU. Editorial Masthead, ESDU News, Engineering Sciences Data Unit Ltd., London, 1970–1982.

(3) ESDU. *Buckling Stress Coefficients for Parallelogram Shaped Panels in Continuous Flat Sheet under Uniform Compression*; Engineering Sciences Data Item 02.01.47; ESDU International plc.: London, 1957; with amendments, 1984.

(4) ASTM. *Standard Guide for Formatting and Use of Material and Chemical Property Data and Database Quality Indicators*; Standard E1484-92; American Society for Testing and Materials: Philadelphia, Sep 1992.

(5) Weinberg, A. M. *Reflections on Big Science*: Pergamon Press: London, 1967.

(6) Barrett, A. J. Social Implications of Data Quality. Proceedings of the 11th International CODATA Conference, Karlsruhe, Germany, 1988.

(7) Barrett, A. J. User-friendliness in Numeric Data Systems. Unpublished Communication to the International Council for Scientific and Technical Information, Numeric Data Group, New York, May 1987.

(8) Barrett, A. J. The Costs of Not Having Refined Information. In *The Value of Information as an Integral Part of Aerospace R&D Programmes*; Advisory Group for Aerospace Research and Development (AGARD) Proceedings, CP 385; AGARD: Paris, 1985; pp 5.1–5.9.

(9) Brown, W. F.; et al. *Materials Properties Data Management—Approaches to a Critical National Need*; National Materials Advisory Board, National Academy Press: Washington, DC, 1983.

(10) Barrett, A. J. Data and Systems Requirements in Computer Analysis of Engineering Components. Society for the Advancement of Material and Process Engineering (SAMPE) Symposium Proccedings; SAMPE: Corvina, CA, 1984; Vol. 29, pp 519–531.

(11) Rumble, John R., Jr. Socioeconomic Barriers in Computerizing Materials Data. *Computerization and Networking of Materials Databases*; Glazman, J. S., Rumble, J. R., Jr., Eds; ASTM STP 1017; American Society for Testing and Materials: Philadelphia, 1989.

(12) Kaufman, J. G. Increasing Data System Responsiveness to End-User Expectations. *Computerization and Networking of Materials Databases*; Kaufman, J. G., Glazman, J. S., Eds.; ASTM STP 1106; American Society for Testing and Materials: Philadelphia, 1991; Vol. 2.

(13) Barrett, A. J. The Benefits and Economic Consequences of Materials Property Databases. *Computerization and Networking of Materials Databases*; Kaufman, J. G., Glazman, J. S., Eds.; ASTM STP 1106; American Society for Testing and Materials: Philadelphia, 1991; Vol. 2.

(14) Jackson, W. G. The Database as an Industrial Product: Production Cost Analysis. Unpublished Communication TGMDM 910509 to the CODATA Task Group on Materials Database Management, Cambridge, Sep 1991.

(15) Nishijima, S., Monma, Y., Eds. *Proceedings of the Association for Science Cooperation in Asia Seminar on Materials Database Technology*; National Research Institute for Metals: Tokyo, 1992.

(16) Bridenbaugh, P. R. *Framing the Commercialization Debate*: *U.S. Political, Industrial and Social Priorities*; Forum on Accelerating the Commercialization of Materials for Competitive Products; National Materials Advisory Board, National Research Council: Washington, DC, Nov 1991.