

55 CODATA / NEWSLETTER

FEBRUARY 1991

Terms Relating to Data, Data Capture, Data Manipulation and Databases—A Glossary

HIGHLIGHTS

Asian-Oceanic Data Sources	4
CODATA Books	3
CODATA Calendar	3
CODATA Glossary	1,2
F. D. Rossini	4
Materials Database Management Newsletter	5,6

CODATA's charter gives special attention to data, data capture, data manipulation, and related concepts and practices. Because its scope and activities are both international and interdisciplinary, it is particularly important that CODATA have access to a broad glossary of relevant terms. With the advent of *CODATA Bulletin* 23 (1,2) in March 1991, such a *Glossary* is available to the international data community. It has been edited by Drs. J. H. Westbrook and W. Grattidge and:

- provides a convenient reference tool for individuals or organizations building databases for broad use or devising standards relevant to their construction;
- educates users and builders of databases in the subtleties and nuances of terminology in the field; and
- contributes to the standardization process in the field of terminology.

There is, perhaps, no more basic building block in the development of an area of science and technology than its terminology. If we are to understand what is meant by the terms used for the specific concepts in a field, or if we are to communicate technical ideas and facts clearly to others, we must have a terminological resource of defined terms which are precise, clear and unambiguous. We must contend with neologisms constantly being coined (hypertext, kriging, metadata), the use of ordinary words in senses special to this subject (field, string, window), acronyms (ASCII, SI, WYSIWYG) and jargon (debug, dump, hashing), all of these examples having become working terms in the data field.

The use of computers in the field of technical data has added new urgency to the terminology issue. Today, computers are used extensively to collect, validate, store, manipulate, search, and retrieve data in all areas of science and technology. Computer-based information storage (both in databases and in expert or knowledge-based systems) has become a major tool in the practice of

(cont'd. to page 2)

The Committee on Data for Science and Technology (CODATA) was established in 1966 by the International Council of Scientific Unions.

Working on an interdisciplinary basis, CODATA seeks to improve the quality, reliability, processing, management, and accessibility of data of importance to science and technology.

Glossary of Data Related Terms

(cont'd. from page 1)

data collection, organization and use. In considering the terminology associated with such data practices, we must cope with: a) three different origins of terminology (the data source, the database producer, and the user); and b) the inflexibility of the computer in responding to search commands unless it is primed with a memory store of a rich variety of synonyms and thesaurus relationships. Thus, while the computer provides a magnificent capability for the storage and manipulation of data—until these problems of terminological diversity are recognized and addressed—the computer will fail to provide complete and accurate retrieval for a broad range of users.

The need for such a *Glossary* in CODATA was first identified at the La Gaillarde planning meeting in June 1983 and was formally proposed for implementation at the 14th General Assembly in Jerusalem in 1984. Dr. J. H. Westbrook was asked to undertake the leadership of the activity. Input was solicited from all entities in the CODATA organization, and recent CODATA conference proceedings were perused. A preliminary version of a much briefer form of the present *Glossary* was developed as an appendix to the proceedings of the 1985 Schluchsee workshop, "Materials Data Systems for Engineering" and reprinted along with some additional terms in *CODATA Bulletin #69, Guide to Material Property Database Management* (Nov. 1988).

In the meantime, standards bodies such as ISO, ANSI, IEEE, and ASTM have issued terminology standards in fields such as computers, information processing, software engineering, etc., which include many terms pertinent to the data field. In building the present *Glossary*, therefore, where terms with definitions appropriate to the data field have been located in other glossaries or terminology standards, they have been adopted with attribution to their source. Sometimes, useful definitions (explicit or implicit) were located in referenced literature sources and have been incorporated with attribution. Surprisingly, for a large fraction of the total list of over 1200 terms, no satisfactory definitions were found in terminological standards or other glossaries on hand, and had to be crafted by the editors. Suggestions for terms for inclusion and reactions to proposed definitions came from many members of the CODATA community as well as others.

The scope chosen for the present *Glossary* involves subjective, non-exhaustive, occasionally redundant choices, and certain terms will be considered superfluous by some readers. As mentioned above, the *Glossary* concentrates on terms pertinent to data, data capture, and data manipulation in the broadest sense, whether these refer to paper or electronic data records. However, data terms specific to individual technical disciplines have been deliberately omitted. A few computer and mathematical terms have been included but only those necessary to clarify the meanings of other definitions. Another broad exclusion is those terms within the subject field, but only just now being created and applied to novel concepts and processes as evidenced by current research papers and conference proceedings. While some of these will undoubtedly become part

of the standard vocabulary of our field, consensus has not yet been reached on usage and definitions. Other exclusions are: company proprietary terms and trade names; multi-word terms whose meaning can be inferred from the definitions of their components; and terms whose meaning for the data community could be inferred directly from standard English. A few of the most frequently encountered acronyms appear, but many more are possible. Only terms and definitions in English have been included; identification of corresponding terms in other languages and translation of definitions must await broader consensus on the present *Glossary*.

The intended audience for the *Glossary* includes:

- authors and readers of CODATA and related publications
- designers and builders of databases, data compilations and data analysis techniques
- developers using computer techniques for textual analysis and knowledge-based systems in the data field
- writers of terminological standards whose scope overlaps the data field

For each of these groups, the *Glossary* is hoped to be an informative reference, a convenient tool, and an effective agent for the promotion of clarity, consistency, and understanding in speaking and writing about the field of data.

The present *Glossary* is the first comprehensive effort for CODATA and is still subject to much improvement and enhancement.

If the work is found to be useful, it is hoped that readers and users will both submit new terms for inclusion and critique the present definitions, so that subsequent editions (perhaps every 3-5 years) can be more extensive, useful and accurate. Suggestions for revision or expansion should be in the form of a proposed change in text together with a rationale therefor and other supporting comments or authorities. A machine-readable version in straight ASCII is available on diskette at the CODATA Secretariat for possible future distribution.

Should acceptance and utility of the *Glossary* become established, preparation of editions with equivalent terms and definitions for languages other than English might well be undertaken by one or another of CODATA's national committees.

This 240-page issue is available from Hemisphere Publishing Corporation (79 Madison Avenue, Suite 1110, New York, NY 10016-7892) or from Taylor and Francis, Ltd. (4 John Street, London, U.K. WCIN 2ET) as *CODATA Bulletin* No. 23 (1,2), 1991, entitled "A Glossary of Terms Relating to Data, Data Capture, Data Manipulation, and Databases" for \$80.00 or as part of an annual subscription to the quarterly *CODATA Bulletin* (ISSN 0366-757X).

CODATA Publications

New Perspectives in Scientific Complex Data Management, A Workshop of the CODATA Task Group on Computer Graphics and Artificial Intelligence, Paris, 21-24 September 1989, *CODATA Bulletin* 22-4. (a)

International Register of Materials Database Managers, prepared by the CODATA Task Group on Materials Database Management, June 1990, *CODATA Special Report No. 13*. (b)

A Glossary of Terms Relating to Data, Data Capture, Data Manipulation, and Databases, edited by J. H. Westbrook and W. Grattidge, *CODATA Bulletin* 23-1,2. (c)

Database Developments in Asian-Oceanic Countries, a CODATA Task Group Symposium held in Seoul, Korea, 31 January - 2 February, 1991, edited by E. F. Westrum, Jr. and Yaru Hu, *CODATA Bulletin* 23-3. (d)

Books and Databases

Composite Index for CRC Handbooks. 3rd edition. Three hard-bound volumes, the entire Index on CD-ROM. (e)

The McGraw-Hill Handbook of Essential Engineering Information and Data, edited by E. N. Ganic and T. G. Hicks. (g)

Compilation of ASTM Standard Definitions. 7th edition. (h)

High-Temperature Aqueous Solutions: Thermodynamic Properties, Roberto J. Fernández-Prini. (i)

CRC Handbook of Thermodynamics, edited by Palmer. (j)

NIST Structures and Properties Database and Estimation Program. (k)

Heat Capacities and Entropies of Organic Compounds in the Condensed Phase, Vol. II, Eugene S. Domalski and Elizabeth D. Hearing. (l)

CRC Handbook of Chemistry and Physics, 71st edition, 1990-1991, editor-in-chief David R. Lide. (m)

(a) Hemisphere Publishing Corporation, New York, 1990, xii + 137 pp., \$33. ISSN 0366 757X.

(b) CODATA Secretariat, 1990, 15 pp.

(c) Hemisphere Publishing Corporation, New York, 1991, x + 240 pp., \$80. ISSN 0366 757X.

(d) Hemisphere Publishing Corporation, New York, 1991, xxx + 139 pp, \$40. ISSN 0366 757X.

(e) Provides a single comprehensive index to topics from the fields of biomedical science, biology, chemistry, engineering, computer science, physics, and mathematics. CRC Press, Inc., 2000 Corporate Blvd., N.W., Boca Raton, Florida 33431, U.S.A. 3 volumes, 3,000 pages, 8 1/2 x 11. Catalog no. 284K. Regularly \$1,195.00, Special pre-publication price \$995.00. ISBN 0 8493 0284 6.

(f) CRC Press, Inc., 2000 Corporate Blvd., N.W., Boca Raton, Florida 33431, U.S.A. 2528 pages, 7 x 10, 1989. Catalog no. 470BF. Regularly \$97.50. Until December 31, 1991, price is \$24.95.

(g) McGraw-Hill Publishing Company, 11 West 19th Street, New York, NY 10011, U.S.A. 212-337-5945 or 212-337-5951. November 1990. 1,072 pages, illustrated, 6 x 9. \$89.50. ISBN 0 07 022764 0.

(h) ASTM, 1916 Race Street, Philadelphia, PA 19103, U.S.A. 215-299-5400. TWX: 710-670-1037. ASTM European Office: 68a Wilbury Way, Hitchin, Herts SG4 0TP England. 0462-31525. TWX: 825684 ATP G. 1990. 560 pages, soft cover. \$75.00. ISBN 0 8031 1238 6. PCN: 03-001090-42.

(i) Presents a unified treatment of aqueous systems, discusses near-critical behavior and high-temperature solution chemistry, and treats electrolytes, nonelectrolytes, sub- and

supercritical systems. December 1990. CRC Press, Inc., 2000 Corporate Blvd., N.W., Boca Raton, FL 33431, U.S.A. Ca. 184 pp., 96 illus., 7 x 10. Catalog no. 5760DFX. Approx. U.S.\$99.50/Outside U.S. \$117.00. ISBN 0 8493 5760 8.

(j) 1987. CRC Press, Inc., 2000 Corporate Blvd., N.W., Boca Raton, FL 33431, U.S.A. Catalog no. 3271ZZ. Approx. U.S.\$179.00/Outside U.S. \$210.00.

(k) NIST Standard Reference Database 25. Available from Standard Reference Data, National Institute of Standards and Technology, Building 221/Room A320, Gaithersburg, MD 20899. \$2400.

(l) Reprint No. 393 of *Journal of Physical and Chemical Reference Data*, 19 (4), 881-1047, 1990, published by ACS and AIP for NIST.

(m) CRC Press, Boca Raton/Ann Arbor/Boston, 1990. ISBN 0 8493 0471 7.

CODATA Calendar

1991

March

25-27 CODATA Executive Committee. Paris, France

July

---- CODATA Materials Database Management Task Group. Trondheim, Norway

1992

April

---- Materials Database Experiences. Moscow, U.S.S.R.

August

17-20 International CODATA Conference. Beijing, P.R. China

21-22 CODATA General Assembly, Beijing, P.R. China

NIST Structures and Properties Database and Estimation Program

This database contains thermochemical data for nearly 5,000 compounds from three widely-accepted Standard Reference Databases—NIST Positive Ion Energetics Database, NIST Chemical Kinetics Database, and NIST JANAF Thermochemical Tables. It also features a complete implementation of Benson's Group Additivity estimation method for gas phase heats of formation, entropies, and heat capacities. Properties are estimated solely from structures that you draw with an integrated, easy-to-use drawing module. No knowledge of estimation methods is required. Some special features are automatic perception of rings and long range interactions, determination of symmetry number correction, computation of equilibrium constants for user-created chemical reactions, and inclusion of 50 new group values and 80 new ring corrections derived from recent literature data. Further information may be found on this page in footnote k.

Professor Frederick D. Rossini, CODATA's first president (1965 to 1969) died October 12, 1990, of pneumonia in Juno Beach, FL, at the age of 91. He was an internationally known authority on petroleum thermochemistry.

A native of Monongahela, PA, Rossini held undergraduate and master's degrees from Carnegie Tech, and a doctorate in chemistry from the University of California at Berkeley.

From 1928 to 1950 he was a scientist at the National Bureau of Standards, Washington, D.C., rising to the position of Chief of the Section of Thermochemistry and Hydrocarbons. After a ten-year association with the Carnegie Institute of Technology, Pittsburgh, PA, where he was Silliman Professor and Head of the Department of Chemistry, as well as Director of the Chemical and Petroleum Research Laboratory, he went to Notre Dame University in 1960, and in 1965 he was the second scientist to receive Notre Dame's Laetare Medal, the University's highest honor. In 1967 he became Notre Dame's first Vice-President for Research and Sponsored Programs after a term as Dean of Notre Dame's College of Science.

He left Notre Dame in 1971 as professor emeritus and joined the faculty of Rice University, remaining there, first as a Professor of Chemistry and then as professor emeritus in residence, until 1978.

He was a member of the National Academy of Sciences and the American Academy of Arts and Sciences and a fellow of the American Association for the Advancement of Science, the American Physical Society, and the American Institute of Chemists.

Professor Rossini, with the cooperation of Dr. Guy Waddington of the U. S. Office of Critical Tables and of Dr. Harrison

Brown, Foreign Secretary of the (U. S.) Academy of Sciences and a member of the Executive Committee of ICSU, set in motion the convening of a Working Group in late 1964. In January 1966 the General Assembly of ICSU meeting in Bombay, India, approved the establishment of CODATA (The Committee on Data for Science and Technology). At the first meeting of the new Committee in Paris in June 1965, Professor Rossini was elected President of CODATA. At that time, only six member countries and ten Scientific Unions were involved in CODATA.

His many contributions to CODATA's early history have been noted in *CODATA Newsletter* No. 38 (1986), page 2 *et seq.*

During his career, Rossini served as President of Sigma Xi, the Washington Academy of Sciences, the Permanent Council of the World Petroleum Congresses, and the Associated Midwest Universities. In addition, he was a member of the American Chemical Society, the American Institute of Chemical Engineers, the American Petroleum Institute, the American Society for Engineering Education, the Chemical Society, the Geochemical Society, Phi Kappa Theta, Tau Beta Pi, the Philosophical Society of Washington, Phi Lambda Upsilon, the Cosmos Club in Washington, D.C., and the Franklin Institute.

He was given the National Medal of Science in 1977. Among other honors he received were honorary doctorates from six schools; the Hillebrand Award of the Chemical Society of Washington, D.C.; the John Price Wetherill Medal of the Franklin Institute; and the Priestly Medal of the American Chemical Society. The Rossini Lectureship of the International Union of Pure and Applied Chemistry honors his many thermodynamic contributions.

TG on Data Sources in Asian-Oceanic Countries Meets

The Fourth Meeting of the CODATA Task Group on the Survey of Data Sources in Asian-Oceanic Countries was held at Olympic Youth Hostel, Seoul, Korea, on 31 January and 1 February, 1991. There were twenty-nine participants from China, Japan, and Korea as well as from India, Indonesia, the Philippines, Taiwan, and Thailand. Regretfully, there were no participants from Australia, Malaysia, and Pakistan.

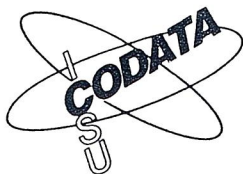
Prof. Mu Shik Jhon, Chairman of the Organizing Committee, presented his welcoming address. He expressed his gladness for having such an important meeting in Seoul, and he thanked Dr. Young-Kyu Yang and his staff for the preparation for this Meeting and the three Korean organizations which supplied financial support.

Prof. Mitsuo Tasumi, Chairman of the Task Group, then reviewed the expansion of membership and the publications from the Task Group in the past year. Seven sessions of oral

presentations by 25 speakers (with pre-distributed documentation) followed.

The updating and enlarging of the *Directory of Data Sources*, the most important expected fruit from the Task Group, were confirmed to be on-stream. More than 600 entries from at least eight countries are anticipated. These will be available in early 1992 as a *CODATA Bulletin*. Australia, Malaysia, and Pakistan will be asked whether they intend to join the survey, and, if so, the status of their endeavor.

Although the final decision on the venue of its next Meeting was postponed, it will most likely be held in Beijing (without symposium) during the International CODATA Conference period in 1992. Discussion of future—post 1991—activities recognized the salutary aspects of the Task Group's present activities and stimulation of database development in the Orient.



Task Group on Materials Database Management

Materials Database Newsletter

February 1991, Number 13

The seventh meeting of the CODATA Task Group on Materials Database Management took place on the 14th and 15th July 1990 in Columbus, Ohio, USA, immediately preceding the 12th International CODATA Conference. Progress was made on a number of projects, including: analysis of the costs and benefits of materials databases; development of a standardized description of materials database systems; and production of a Register of Materials Database Managers (now available as *CODATA Special Report* Number 13). The plans for a symposium/workshop on International Exchange of Materials Data, to be held in Moscow in April 1992, were also reviewed.

DATABASES

The **Mechanical Properties Database for Materials** being developed by the Shanghai Research Institute for Materials is now in operation. The database contains data for 200 materials, including steels, nonferrous metals and plastics. The emphasis is on fatigue and high-temperature test data, and the system supports graphics functions. The next stage will involve the development of an expert system. CONTACT: Ms. Bao Mu Lan, Shanghai Research Institute for Materials, 99 Handan Road, Shanghai, China.

STANDARDS

VAMAS Technical Working Area 10 "Materials Databanks" has published a report on **Significance of Data Evaluation Models in Materials Databases**. Based on the results of a round robin comparison of data evaluation methods for creep and fatigue data of typical engineering steels and alloys, the report concludes with a number of recommendations: worldwide information exchange and discussions on methods and models for materials data evaluation should be encouraged; an inventory of commonly used data evaluation methods should be developed; the technology of numeric data exchange should be studied in collaboration with computer engineering societies on the basis of ISO/IEEE standards; and international activities to establish a neutral, machine-independent, standardized format for materials data exchange should be increased. Further information and copies of the report are available from: Dr. S. Nishijima, National Research Institute for Metals, 2-3-12 Nakameguro, Meguro-ku, Tokyo 153, Japan.

The National Association of Corrosion Engineers (NACE) has published Report No. RP0690-90 **Standard Format for Collection and Compilation of Data for Computerized Material Corrosion Resistance Database Input**. The standard defines data categories and specific data elements that are considered desirable or essential to accommodate search strategies and reliable data comparisons in computerized corrosion databases. The document is available for \$10 (NACE members), \$12 (non-members) from: NACE Customer Service Department, P. O. Box 218340, Houston, TX 77218, USA.

CALENDAR

12-14 June 1991: Weybridge, Surrey, UK

PROPERTIES OF ENGINEERING MATERIALS: METROLOGY AND STANDARDS. CONTACT: B. E. Larcombe, MISI, Building 1, National Physical Laboratory, Teddington, Middx. UK, TW11 0LW.

19-22 March 1991: Petten, NETHERLANDS

CEC/CODATA Joint Workshop on **MATERIALS DATA FOR COMPUTER AIDED ENGINEERING**. CONTACT: H. Kröckel, CEC JRC, Institute for Advanced Materials, P. O. Box 2, 1755 ZG Petten, The Netherlands.

7-9 May 1991: Atlantic City, NJ, USA

Meeting of ASTM Committee E49 on **COMPUTERIZATION OF MATERIALS PROPERTY DATA**

CALENDAR
(cont'd.)

13-14 June 1991: Cleveland, OH, USA

STANDARDIZING TERMINOLOGY FOR BETTER COMMUNICATION: PRACTICE, APPLIED THEORY, RESULTS. CONTACT: ASTM, 1916 Race Street, Philadelphia, PA 19103-1187, USA.

9-11 September 1991: Cambridge, UK

Third International Symposium on **COMPUTERIZATION OF MATERIALS PROPERTY DATA.** CONTACT: Dr. T. Barry, Division of Materials Metrology, National Physical Laboratory, Teddington, Middx., TW11 0LW, UK.

3-5 November 1991: San Diego, CA, USA

Meeting of ASTM Committee E49 on **COMPUTERIZATION OF MATERIALS PROPERTY DATA.**

April 1992: Moscow, USSR

Symposium/Workshop/Exhibition on **MATERIALS DATABASE EXPERIENCE - OPPORTUNITIES FOR INTERNATIONAL EXCHANGE.** CONTACT: CODATA Secretariat, 51 Boulevard de Montmorency, 75016 Paris, France.

EDITOR: W. G. Jackson. EDITORIAL OFFICE: The Institute of Metals, 1 Carlton House Terrace, London, UK, SW1Y 5DB. Tel: +44 71 839 4071; Tlx: 8814813; Telefax: +44 71 839 2289. There are no restrictions on the reproduction and distribution of the contents of this Newsletter.

Editor : Edgar F. Westrum, Jr.
Department of Chemistry, University of Michigan,
Ann Arbor, MI 48109
Telephone: (313) 764-7357 / Telex: 8102236056
FAX: 1-313-747-4865

Published four times per year (February, May, August, November)

Associate Editor: Phyllis Glaeser
CODATA Secretariat, 51 Blvd. de Montmorency,
75016 Paris, France
Telephone: 33 1 45250496 / Telex: 630553
FAX: +33 1 42889431 / Cables: ICSU Paris 016

Assistance in dissemination provided by National Committees

INTERNATIONAL COUNCIL OF SCIENTIFIC UNIONS
COMMITTEE ON DATA FOR SCIENCE AND TECHNOLOGY



CODATA / NEWSLETTER

CODATA, 51 bd. de Montmorency, 75016 Paris