ASTM Manual on

The BUILDING of MATERIALS DATABASES

CRYSTAL H. NEWTON

EDITOR

Manual on the *Building of Materials Databases*

Crystal H. Newton, Editor

ASTM Manual Series: MNL 19
ASTM Publication Code Number (PCN)
28-019093-63

1916 Race Street ■ Philadelphia, PA 19103

Library of Congress Cataloging-in-Publication Data

Manual on the building of materials databases / Crystal H. Newton, editor.
(ASTM manual series ; MNL 19)
"ASTM Publication Code Number (PCN) 28-019093-63."
Includes bibliographical references and index.
ISBN 0-8031-2052-4
1. Materials--Data bases. I. Newton, Crystal H. II. Series.
TA404.25.M36 1993
025.06'62011--dc20 93-36460
CIP

Copyright © 1993 AMERICAN SOCIETY FOR TESTING AND MATERIALS, Philadelphia, PA. All rights reserved. This material may not be reproduced or copied, in whole or in part, in any printed, mechanical, electronic, film, or other distribution and storage media, without the written consent of the publisher.

Photocopy Rights

Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by the AMERICAN SOCIETY FOR TESTING AND MATERIALS for users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that the base fee of $2.50 per copy, plus $0.50 per page is paid directly to CCC, 27 Congress St., Salem, MA 01970; (508) 744-3350. For those organizations that have been granted a photocopy license by CCC, a separate system of payment has been arranged. The fee code for users of the Transactional Reporting Service is 0-8031-2052-4 93 $2.50 + .50.

NOTE: This manual does not purport to address (all of) the safety problems associated with its use. It is the responsibility of the user of this manual to establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.

Printed in Ann Arbor, MI
Nov. 1993

Foreword

THIS MANUAL WAS prepared to address a need perceived by ASTM Committee E-49 on Computerization of Material and Chemical Property Data for guidance in using standards for assistance in developing material property databases, but is not to be considered a standard. This manual, and the standards it discusses, often cannot provide final answers as these are dependent on the database application. What this manual does provide is guidance to help database design teams address the questions for particular materials database applications. In addition, the manual may serve as a focal point for the developing technology and standardization in the material property database community.

This publication was sponsored by ASTM Committee E-49. Several members of ASTM Committee E-49 contributed to the development of the manual concept and outline; the efforts of John R. Rumble, Jr., Bert J. Moniz, Keith W. Reynard, and Jack H. Westbrook are acknowledged. The reviewers, who played an essential role in the development of the manual, also deserve recognition.

Crystal H. Newton
Editor

Contents

Overview

THIS MANUAL FOCUSES on the building of material property databases and the standards that are available to assist in the process. The building of databases has been discussed in general terms in many references. What is important to consider here are the steps in the database building process that are different for material property databases. What are the key decision points? Where can you find resources for help at those key decision points? Most importantly, how can standards help with the process of building a materials database? This manual, and the standards it discusses, often cannot provide final answers as these are dependent on the database application. What this manual does provide is guidance to help database design teams address the questions for particular materials database applications.

Chapter 1 provides an introduction to the development of material property databases. The value of material property databases is discussed. Key concepts that are used throughout the manual are introduced. The standards organizations involved in materials property databases are discussed. This manual focuses on the use of standards developed by or in cooperation with ASTM Committee E-49 on the Computerization of Material and Chemical Property Data. ASTM Committee E-49 is at the forefront in developing standards in this area. The final section of this chapter introduces the steps involved in the design of a materials property database. The steps highlight the use of the ASTM E-49 standards and the other chapters in the manual.

Chapter 2 discusses the functions of the personnel involved in building a database and considerations regarding the system architecture particularly applicable to materials databases. Chapter 3 addresses the different types of material property data and database applications, which influence the system architecture. The data dictionary can be developed with the help of ASTM standard guides. ASTM Committee E-49 has divided materials data into two areas: the identification of the material and the recording of test results. Chapter 4 discusses the nomenclature and standards for identification of engineering materials, and Chapter 5 discusses nomenclature and standards for recording test results and material properties.

Chapter 6 contains information on evaluating data and database quality. Again, depending on the type of data, the application area, and the use of the database, data quality may be indicated as part of each record in the database, once for each record, or as a general indicator of the quality of an entire database. Chapter 7 discusses the operation and maintenance of databases for computers ranging from PCs to mainframes. Chapter 8 considers the transfer of data between databases. The last chapter, Chapter 9, includes example data records from a composite material database, developed with the assistance of the ASTM E-49 standards.

Crystal H. Newton, Editor
Materials Sciences Corporation
Fort Washington, PA 19034

1

Introduction to the Building of Materials Databases

Crystal H. Newton[1]

FOCUS

This chapter provides an introduction to the development of material property databases. The value of material property databases is discussed. Key concepts that are used throughout the manual are introduced. The standards organizations involved in materials property databases are discussed. This manual focuses on the use of standards developed by or in cooperation with ASTM Committee E-49 on the Computerization of Material and Chemical Property Data. ASTM Committee E-49 is at the forefront in developing standards in this area. The final section of this chapter introduces the steps involved in the design of a materials property database. The steps highlight the use of the ASTM E-49 standards and the other chapters in the manual.

VALUE OF MATERIAL PROPERTY DATABASES

What is a Database?

The term database is commonly used in two ways: Traditionally, the word database has been used to describe any collection of information. More recently the term is used to describe a computerized collection of related information which can be used without knowing the details of the storage structure, namely, a computerized database. The latter definition will be used in this manual without requiring the use of the modifying word, computerized. Note that the more traditional definition is still used by many engineers and scientists.

Databases can be compared to two other computerized collections of information, the spreadsheet and the expert system. A spreadsheet may contain data, but the structure of the data storage, for example, cell location, must be known to access the data. An expert system is predominantly a collection of rules while a database is predominantly a collection of facts or properties. There is not a completely clear distinction between the two since some manipulation of data by rules is often implemented in materials databases, and an expert system often contains data (facts).

The Value of Materials Databases

The value of materials databases is considerable but, unfortunately, difficult to quantify. The financial benefit of easy access to high-quality data during the design process is an intangible figure, difficult to sell to managers who have to make a decision based on the bottom line. One difficulty is to isolate the contribution of good data readily available to a project or, conversely, the cost of having poor data. The cost of developing a new material from concept through certification can be quite high. Intuitively, if the steps involved in documenting the properties of the new material do not need to be repeated, the benefits of accessibility to the original data can be substantial. Structured lists of benefits used in order to provide a basis for the demonstration of the economic consequences of the use of materials databases have been developed [*1*,*2*]. Various socioeconomic barriers to the development of material property databases have been discussed [*3*]. Of particular note is the separation of database features from the associated benefits as shown in Table 1.1.

To be useful to a wide range of users, materials information in a database should contain all the information necessary to regenerate the data. It is difficult to establish the benefits of a materials database, in part, because it is not simply only making the data currently available on paper in a user-friendly computerized form. What is extremely important in the development of a database, is ensuring that the *metadata*, the documentation that identifies the material, the test method, and pertinent variables, are included in addition to the material properties data.

This manual concentrates on the ways that standards, particularly the standard guides developed at ASTM, can be used to develop effective materials properties databases. The guides provide recommendations for the metadata that need to be included in the planning of the database structure. In addition, information regarding types of data, the evaluation of the data, operation of the database, and planning for the exchange of data between databases is included.

KEY CONCEPTS

Important to any discussion is a clear understanding of the concepts involved. Several concepts are important in the development of materials databases. Some of these concepts are defined in the ASTM guides discussed in subsequent

---

[1] Project engineer, Materials Sciences Corporation, 500 Office Center Drive, Suite 250, Fort Washington, PA 19034.

Copyright© 1993 by ASTM International www.astm.org

TABLE 1.1—Functions, features, and benefits of material databases as developed by CODATA (part shown for illustration only) [1].

Feature	Benefit
2.1 FUNCTION: ORGANIZATION AND STRUCTURING OF DATA AND INFORMATION	
2.1.1 Database provides comprehensive coverage and the immediate availability of the full range of data to an unlimited number of users. If the database did not exist, the data would be dispersed in the literature and would only be accessible with difficulty.	Access to a wide range of data, more quickly, and at costs more widely shared than by other means.
2.1.2 Assuming that the database is maintained by a team active in the scientific field represented by the data, the database will be qualitatively and quantitatively reliable. The need for checking, verification, and comparison by the user is minimized.	Access to data requiring little if any work to establish its quality.
2.1.3 A database is a coherent source of reference and a working example for anyone setting up a new database (or an extension) based upon test programs. So test programs can be more rationally and economically defined on the basis of awareness of existing data.	Optimization of test programs, requiring fewer tests, less screening, and less effort on data comparison.
2.1.4 A database organizes a mass of data.	Search locations are minimized.
2.1.5 Data are collected in one location.	Search time for relevant data is minimized.
2.1.6 Commentaries on the data can be included.	Reliability indications and cautions can be obtained.
2.1.7 Databases can easily be updated.	The consequences of errors and inadequacies of out-of-date data may be avoided.
2.1.8 When databases are organized and structured to acceptable standards, different databases may be interfaced to exchange data.	The task and costs of providing quality data may be shared.
2.1.9 Databases facilitate fast retrieval and comparison of data.	Saving of time in engineering applications of data.
(... and so forth)	(... and so forth)

chapters. Discussion of these concepts and highlights of their importance are included here for reference during use of the manual.

Data Terminology

Data and Metadata

In the materials fields, information is often divided into (1) data that represent properties, experimental measurements, and so on, and (2) metadata. CODATA has developed the following definition for data [4]:

> Data—The set of scientific or technical data measurements, observations, or facts that can be represented by numbers, tables, graphs, models, text, or symbols and which are used as a basis for reasoning or calculation. (Sometimes called information bits or databits). Note "data" is a plural form; "datum" is the singular.

ASTM Committee E-49 defines metadata in accordance with ASTM Terminology Relating to Building and Accessing Materials and Chemical Databases (E 1443):

> Metadata—Information that describes other data. Metadata are used to identify, define, and describe the characteristics of data.

The division between properties and experimental measurements and metadata depends on the application area and the purpose of the database. As an example, consider yield strength for cast iron. The yield strength is often considered as part of the material identification and may be considered as metadata. It is, however, an experimental measurement. For applications where a number of measurements of the yield strength are made or yield strength is considered to be a dependent variable based on other parameters, the yield strength should be treated as experimental data.

The decision of what information can be treated as metadata will affect the grouping of the information in the database. A database may be organized so that the material identification and test method information are included once while the experimental results are repeated for each specimen. Some of the guides for recording test data discussed in Chapter 4 point out sets of fields which might be repeated. This again is a decision for the database design team. The relative amounts of metadata and experimental information should be considered. A decision then needs to be made based on the trade-off of the storage space saved by repeating groups of fields balanced against the increased complexity of the database programming.

Schema

Schemas are views of the database architecture. Four different schemas are commonly considered [5]:

1. The physical schema views the data as the bytes stored in blocks on disks and possibly tapes.
2. The internal schema is the view of the data as logical files.
3. The conceptual schema is the global view of the data in the database, stored in the data dictionary by most systems as a list of files, records, fields, relationships, and constraints.
4. The external schema is the view the user has of the data. There may be several different external schemas for the same database.

The discussions in this manual refer primarily to the conceptual schema. The physical schema depends on the database management system being used and is beyond the scope of this manual. The internal schema is often considered to be the same as the conceptual schema, such as in the American National Standards Institute (ANSI) standards. One or more external schemas should be developed by the designer based on the users' needs. Particularly, if the users can be divided into groups based on needs, an external schema for each group of users should be considered.

Data Dictionary

The data dictionary stores the conceptual schema, definition of data elements, and additional information on the database.

Recording Format

The ASTM E-49 guides provide assistance in the development of standard recording formats. These formats include essential and recommended fields, category sets, value sets, and units for specific purposes.

Data Element

ASTM Guide for the Development of Standard Data Records for Computerization of Material Property Data (E 1313) provides the following definition:

> Data element—An individual piece of information used to describe a material or to record test results, for example, a variable name, test parameter, etc., synonymous with data item.

Field

A field is the fundamental location for storing a data element, defined in ASTM E 1443 as:

> Field—An elementary unit of a record that may contain a data item, a data aggregate, a pointer, or a link.

Fields are established for a record based on the data elements that the database is required to store.

Essential Field

ASTM Committee E-49 has defined an essential field as [*4*]:

> Essential field—A field in a record that must be completed in order to make the record meaningful in accordance with the pertinent guidelines or standard. Note: fields are considered essential if they are required to make a comparison of property data from different sources meaningful. A comparison of data from different sources may still be possible if essential information is omitted, but the value of the comparison may be greatly reduced.

A field that is identified as an essential field needs to exist in the database, according to E-49 recommendations. Also, the judgment is made that the datum should be available for any data set. One of the questions that may be asked in judging the quality of a database (see Chapter 6) is the inclusion of all fields considered to be essential for the application. Guidelines that recommend essential fields thus also become recommendations for essential data. A different judgment on quality is made for each record based on having all the essential fields filled. This carries further to the experimental procedure where a reference to the ASTM E-49 guides implicitly requires that all data to fill the essential fields must be recorded.

The fields that should be considered essential for a database for a given function and application area need to be determined by the database design team. The ASTM standard guides provide guidance but not requirements in this area. The connotation of essential fields may vary from guide to guide. For example, compare the data recording guides for metals and composites. The composites document identifies many more fields as essential when compared to the metals document. The effects of testing method and material parameters are not fully understood for composite materials; consequently, many variables need to be documented to maintain the usefulness of the data when these effects are determined.

In developing a database, the interpretation of "essential" depends on the type of material, the industry involved, and the database application. One consideration is how much data is intended to be covered by the database being designed. If certain fields should be considered essential for 90% of the data with additional fields necessary for the remaining 10%, a database design team should consider which type of data are needed for the particular database application the team is addressing.

Value Sets and Category Sets

Most of the guides for material identification and recording test results include value sets and category sets. ASTM E-49 defines these two terms as [*4*]:

> Value set—An open listing of representative acceptable strings that could be included in a particular field of a record. Discussion: a closed listing of such a string is called a domain or category set.
>
> Category set—A closed listing of the possible or acceptable strings that could be included in a particular field of a computerized record.

Most of the sets of acceptable strings in the guides are value sets (incomplete sets). Some fields for character strings have no set of acceptable values. The development of category sets is impractical for some types of material information. As an example, consider the field, material identification. How many different materials exist? A value set for this field could list thousands of acceptable strings and still not be complete. Again the database design team needs to establish value sets and category sets that are as comprehensive as possible and use standardized strings when available.

Allowed Value

In designing a material property database, the concept of allowed values should be considered. An allowed value is defined as in ASTM E 1313:

> Allowed value—A member of a defined set of permitted values; for example, a category set, a value set, etc.
>
> Discussion—For quantitative parameters, the set is a theoretically or experimentally based range of possible numeric values; for qualitative parameters, the set shall consist of a finite number and enumerated list of standard words or a well-defined system of codes.

TABLE 1.2—Fields in the term record structure [25].

IDENTITY_BLOCK	DEFINITION_BLOCK	THESAURUS BLOCK
• TERM	• DEFINITION	• USED_FOR
• TYPE	• DEFINED_BY	• BROADER_TERM
TERM_NUMBER	• DEFINING_DOCUMENT	RELATED_TERM
CROSS_REFERENCE_INDEX_CODE	OTHER_RELATED_DOC	NON_SYNONYMOUS_TERM
USED_IN	LABEL	STANDARD_TERM
PERTINENT_TO_MATERIAL_CLASS	ABBREVIATION	BROAD_APPLICATION_AREA
	SYMBOL	UNIT_CLASS
	MNEMONIC	STANDARD_UNITS
	• RQD_VAR	VALIO_UNIT

RECORD_MOD_BLOCK	POSSIBLE_ALLOWED_VALUES_BLOCK
MODIFIER_TIME	DATA_VALUE_TYPE
MODIFIER_DATE	ALLOWED_VALUE
MODIFIER	
MOD_DESCRIPTION	

• Essential field

The concept of allowed values can be used to establish types of fields and ranges for checking input data.

Additional Terminology

Additional terminology is provided in ASTM E 1443 and in each of the guides.

Thesaurus

The need for a thesaurus, common to all types of databases, should be emphasized for materials databases. Many terms that are used for field names have a number of synonyms. Westbrook and Grattidge provide an example of the synonyms for modulus of elasticity, as shown in Table 1 of Ref *6*. In addition to field names, many synonyms exist for data in category or value sets. ASTM Subcommittee E49.03 on Terminology has developed a practice for a term record structure for use in developing data dictionaries and thesauri (ASTM Practice for Structuring Terminological Records Related to Computerized Test Reporting and Materials Designation Formats [E 1314]). The fields in the practice are shown in Table 1.2. The need for terminology standardization and harmonization is discussed below.

Units of Measurement

Implicitly or explicitly associated with almost every technical data value is the unit of measurement. In constructing a database, the design team needs to be aware of the users' assumptions regarding these units. The user is going to be most comfortable using a database that reports the data in the units most commonly used in the application area. The degree to which units need to be stored as part of the data set (identified as part of table and graph titles or headings), or assumed, needs to be considered by the database design team. If more than one system of measurement is commonly used, units conversion and storage of data in original units need to be addressed as well. As discussed in Chapter 2, the importance of identifying accuracy of data increases when unit conversion is implemented.

STANDARDS TO AID THE DATABASE DEVELOPMENT PROCESS

The use of standards in the development of a materials database provides guidelines for selecting and defining data elements, creating the data dictionary and database schema, and developing the database functional requirements. Many materials databases have been developed with incomplete data and with inadequate capabilities. Materials are being used in ever widening and increasingly advanced applications. The materials area, in fact, is considered one of the most critical areas for new technology. In order to use new materials and to use existing materials in new ways, data that accurately reflect the materials' capabilities are vital. Guidelines for the data to be included in a database and guidelines for the database, such that it adequately manipulates the required data, help database designers meet these needs.

Additional reasons for standardization in the database area include the development of databases that are used internationally and the fact that the rate of exchange of information is rapidly increasing. The amount of communication of technical data is increasing as is the ability to access databases remote from the engineer. Examples of projects to enable ready access of a number of databases are described in Refs *7* and *8*. As industries operate with increasing involvement internationally, the need for standards for databases and exchange of data is increasing.

Standards Organizations

ASTM Committee E-49 on Computerization of Material and Chemical Property Data

ASTM is a U. S. national consensus organization "formed for the development of standards on characteristics and performance of materials, products, systems, and services; and the promotion of related knowledge" [9]. The society depends on the development and adoption of standards, including test methods, definitions, recommended practices, classifications, and specifications, through a voluntary consensus process. Essential to this process is consideration of minority opinions.

ASTM Committee E-49 on Computerization of Material Property Data was organized in 1985 [*10,11*]. Chemical data was added to the scope of the committee in 1991. The committee's scope is currently being revised to "The promotion of knowledge and development of standard classifications, guides, specifications, practices, and terminology for building and accessing computerized material and chemical databases, and exchanging information among those databases and computer software applications and systems using the data therein" [*12*]. The committee has its activities divided between two sections: materials and chemicals data. Active subcommittees within the Materials Section, shown in Fig. 1.1, include the following:

- E49.01 on Materials Designations
- E49.02 on Data Recording Formats
- E49.03 on Terminology
- E49.04 on Data Exchange
- E49.05 on Data and Database Quality

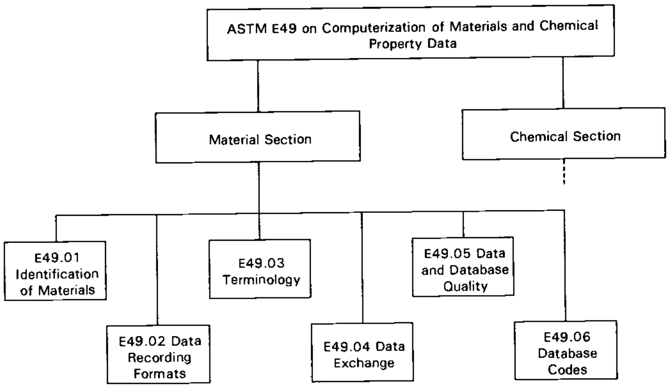


FIG. 1.1—Organizational chart for ASTM Committee E-49 Material Section.

- E49.06 on Database Codes

These represent areas of activity for the computerization and exchange of material property data. The development of formats for material properties have been divided into two principal areas: (1) the identification of the material and (2) the recording of test data and properties. The two areas are reflected in ASTM Subcommittees E49.01 and E49.02 and in Chapters 4 and 5 of this manual. Standards applicable to both types of formats and databases in general are developed in the other subcommittees and discussed in the remaining chapters. Information on the activities of ASTM E-49 can be obtained from the ASTM E-49 staff manager, ASTM, 1916 Race Street, Philadelphia, PA, 19103, telephone (215) 299-5513, and facsimile (215) 299-2630.

The approach that has been taken by ASTM E-49 has been to establish standards related to data collection and the contents of databases. The committee feels that it is inappropriate to standardize database design and user interfaces. These areas are presently under considerable creative development, and it is too early for standards. Guidance for the requirements in this area is being developed. It is assumed that the database designers will consider the kinds of questions listed later in this chapter on p. 7.

A number of formats relevant to the identification of materials and the recording of test data in computerized material property databases have been developed. These formats are listed in Table 1.3 with other standards developed by ASTM Committee E-49. These standards can be obtained from ASTM. Additional standards are under development, also noted in Table 1.3. The relatively large number of standard formats that are under the jurisdiction of other committees and organizations is unusual for an ASTM committee. However, ASTM Committee E-49 actively supports this coordination and has been designated a resource committee by ASTM, to be consulted by other committees that are developing recommendations for the computerization of any type of material or chemical information. The standards developed by ASTM E-49 are discussed in this chapter and most of the other chapters in this manual.

International Standards Organization (ISO) STEP Materials Team

The international standardization of data exchange related to products is occurring under ISO Technical Committee

TABLE 1.3—Standards developed by or in coordination with ASTM Committee E-49.

Designation	Title
	IDENTIFICATION OF MATERIALS
E 1308	Guidelines for the Identification of Polymers (Excluding Thermoset Elastomers) in Computerized Material Property Databases
E 1309	Guidelines for the Identification of Composite Materials in Computerized Material Property Databases
E 1338	Guidelines for the Identification of Metals and Alloys in Computerized Material Property Databases
E 1339	Guidelines for the Identification of Aluminum Alloys and Parts in Material Property Databases
E 1471	Guidelines for the Identification of Fibers, Fillers, and Core Materials in Computerized Material Property Databases
AWS A9.1	Describing Arc Welds in Computerized Material Property Databases (under American Welding Society jurisdiction)
*	Guidelines for the Identification of Ceramics in Computerized Material Property Databases
*	Guidelines for the Identification of Coatings of Engineering Materials in Computerized Material Property Databases
*	Guidelines for the Identification of Copper and Copper Alloys in Computerized Material Property Databases
*	Guidelines for the Identification of Steel Alloys in Computerized Material Property Databases
	DATA RECORDING
E 1313	Guide for the Development of Standard Data Records for Computerization of Material Property Data [being revised as Guide for Recommended Formats for Data Records Used in the Computerization of Mechanical Test Data for Metals] X1. Bearing Test Data Based on ASTM Method E 238 X2. Plane-Strain Fracture Toughness Test Data Based on ASTM Method E 399 X3. Tensile Test Data Based on ASTM Test Method E 8 X4. Compression Test Data Based on ASTM Test Method E 9 X5. Notched Bar Impact Data Based on ASTM Test Method E 23
E 1443	Guide for the Development of Standard Data Records for Computerization of Mechanical Test Data for High Modulus Fiber-Reinforced Composite Materials (under ASTM D-30 jurisdiction)
	DATA RECORDING
E 1454	Guide for Data Fields for Computerized Transfer of Digital Ultrasonic Testing Data (under ASTM E-7 jurisdiction)
E 1475	Guide for Data Fields for Computerized Transfer of Radiologic Testing Data (under ASTM E-7 jurisdiction)
G 107	Formats for Collection and Compilation of Corrosion Data for Metals for Computerized Database Input (under ASTM G-1 jurisdiction)
	Standard Data Records for Computerization of Power Frequency Magnetic Testing (under ASTM A-6 jurisdiction)
AWS A9.2	Recording Arc Weld Material Property and Inspection Data in Computerized Databases (under American Welding Society jurisdiction)
*	Guide for the Development of Formats for Recording Data Generated by Standard Tests
*	Tensile Test Data for Plastics According to ASTM D 638 and D 638M

TABLE 1.3—Standards developed by or in coordination with ASTM Committee E-49 — Continued.

Designation	Title
*	Guide for Recommended Data Format of Sliding Wear Test Data Suitable for Databases (under ASTM G-2 jurisdiction)
*	Fatigue Crack Growth Rate Test Data (E 647)
*	Strain Controlled Fatigue Testing per ASTM E 606
	TERMINOLOGY
E 1314	Practice for Structuring Terminological Records Relating to Computerized Test Reporting and Materials Designation Formats
E 1443	Definitions of Terms Relating to Building and Accessing Materials Databases
	DATA AND DATABASE QUALITY
E 1407	Guide for Materials Database Management
E 1484	Guide for Formatting and Use of Material and Chemical Property Data and Database Quality Indicators
E 1485	Guide for the Development of Material and Chemical Property Database Descriptions
	DATABASE CODES
*	Practice for Coding of Materials Property Data

*Draft

184/Subcommittee 4 on Industrial Data and Global Manufacturing Programming Languages. A materials model is under development which addresses material identification and properties. Application protocols, which are higher level models, will be written for specific materials applications. For information on the materials activities of ISO TC 184/SC4, contact the STEP Secretariat, NIST, A247 Metrology Building, Gaithersburg, MD 20899.

Terminology

The need for standardized terminology related to material databases must be emphasized. A VAMAS report [*6*] notes "Much effort is expended on terminology in information technology. To the users it sometimes seems to be a tedious and nit-picking exercise. To those who suffer financially as a result of a misunderstanding, all effort is worthwhile."

An example of the diversity in terminology is shown in Table 1.4 from Westbrook and Grattidge [*13*]. It should be noted that many more terms than those shown in the table are possible. A variety of symbols and mnemonics are used in place of a property name in databases. The directionality of material and direction of the applied load are often combined with the property name to compound the number of possible terms. An example of the same term used for a completely different concept is the use of modulus in viscoelasticity where it is defined as the efficacy of returning input energy by an elastomer. An additional complexity for the modulus terms is the method of calculating the value. The method of calculating the modulus and location on the stress-strain curve on which that calculation is based has been ignored in many databases to date. However, for many materials, linear stress-strain curves are the exception. In these cases, the method of calculation must be identified as part of the term or by other means.

TABLE 1.4—Example of the diverse synonyms for a given term [*13*].

Modulus of Elasticity	
Elastic Modulus	Tensile Modulus
Young's Modulus	Coefficient of Elasticity
Modulus of Elasticity	Modulus in Tension
Stretch Modulus	Deflection Resistance
Monotonic Modulus	Extensional Modulus
Static Modulus	E

International harmonization of terminology is even more complex. First of all, terminology that has a common ground in spoken language within one country may be rather different in another country in what is, at least nominally, the same language. Standards for terminology used in the materials field are the common reference vocabulary of the *European Materials Databanks Demonstrator Programme, Common Reference Vocabulary* [*14*], the ISO bibliography of all ISO standards on terminology [*15*], and the ASTM compilation of terms from terminology standards [*16*]. None of these have as yet addressed the real problem of harmonized terminology.

PLANNING A MATERIAL PROPERTY DATABASE

Much of the information available in texts on building databases is applicable to building a materials database. Recommended procedures for building a database are described in a number of texts [*5,17,18*]. Of particular note is Ref *5*, which refers specifically to databases in science and engineering. The information included here is not intended to replace those basic texts, but to add information relative to material property databases and the standard guides available to provide assistance. The steps involved in the database design process are considered in Refs *17* and *19* among others. Each of these approaches is described in slightly different terms, but they are all similar. Here the database design process can be considered to include the following steps for the planning of the database:

1. Identify the application.
2. Select the project team.
3. Plan user involvement.
4. Define the database functional and performance requirements.
5. Select hardware and software.
6. Identify the data elements.
7. Build the data element dictionary.
8. Group data elements.
9. Identify the retrieval characteristics of each group.
10. Identify the relationships between groups.
11. Develop the database schema.

These steps result in the development of the conceptual and internal schemas. Each of these steps will be considered briefly with considerations specific to material property databases and the use of standards to assist in material property database development.

Identify the Application

In designing a database for material properties, it is important to consider the intended application for the database. A database intended to provide material properties for use in design may be very different from a database that replaces lab notebooks in the testing lab. Other possible applications include quality control, tracking life-history of a material/product, and material research. The type of the material and the philosophy of the organization building the database will affect the amount of influence application has on the database design.

Consider a basic property of a common material: the yield strength of 6061-T6 aluminum alloy of 280 MPa (40 ksi). For some applications, these data—alloy, heat treatment, and yield strength—may be adequate. For many applications, it might also be beneficial to know additional information, such as the material form and dimensions during heat treatment, the test method, the type of test coupon, the location where the test was performed, the environmental conditions, and preconditioning.

A quality control database may require a relatively small set of data fields, such as date, lot number, measurement, and value. The quality control application is often a well-defined environment where the same test method, specimen types, and conditions are consistently used. The other extreme is a research database where every parameter may be varied at some point. The product lines and the number of candidate materials described in the database should be considered. The quality control database is generally limited to materials that the company uses or produces, while a design or research database will also include candidate materials. The material class or classes also interact with the company philosophy and role. A company that supplies raw materials would support a database that makes any application data available to their customers. A company that produces an end-product will often restrict access to data, noting that the data are part of their market advantage.

These are examples of considerations necessary in the design of a database. The standard guides discussed, particularly in Chapter 3 and 4, consider a general approach to data. These are intended to be guidelines that can then be applied to the design of each database. Questions that should be asked by the database design team include:

- What materials will be included in the database?
- How much information is needed for each material?
- Will the database be used for more than one purpose?
- What information is required for each application for each material?
- What assumptions are likely to be violated in even a small number of cases?
- What information will help users get the most use from the database?

The range of data that the database can handle must also be considered. Should the database have fields for 90% of the data and metadata with the remaining data and metadata included in the comment or footnote fields or should a database contain sufficient fields to cover all information? These questions must also be considered by the design team for each database.

Another possibility that should be considered in a database design is the long-term use of the database. The short-term specific goal may be more limited in scope than the eventual use of the database. As much as possible, the long-term scope and application of the database should be considered at this stage. If data are expensive to obtain, as often is the case, it may be cost effective to develop a database with increased documentation requirements to get the most use from the database. If the eventual use of the database is not clear, a more flexible approach to the database design may be desirable.

Select the Project Team

The development of a database involves four different job titles. Depending on the situation, a team member may be involved with one or more of the responsibilities. The *project leader* is responsible for organization of the project, ensuring that both materials and software communities are represented, and selling the database concept within or outside the organization, or both. The *software engineer* is responsible for programming. The materials community is represented by the *data provider* and the *user*. Depending on the application of the database, the data providers and the users may be members of the same group. Frequent and thorough communication between the database programmers and the materials community throughout the project is essential to the development of a useful database. The users must communicate their needs to the database programmers. The database programmers must develop the database concept. Both groups must understand and review the concept in terms of the requirements of the users, the data providers, and the programmers. The satisfaction of the users with the database can be greatly improved if both the database programmers and the materials community discuss tradeoffs in task efficiency. The user will perform some tasks many times and will not be tolerant of slow speeds in these tasks. Other tasks will be performed less often and may be programmed for slower speeds in exchange for higher speeds on primary tasks or greater database capacity. If this task prioritization is based on assumptions by either community about the needs of the other without thorough communication, chances for high levels of user approval are small.

The project leader is responsible for seeing that thorough communication occurs. If the data providers and users have not been involved with databases extensively, they will need to be educated in how to develop the database requirements from their perspective. An engineer or materials scientist will often consider the data values as they commonly appear without considering implicit assumptions regarding units, quality, and accuracy. Test methods need to be considered for possible effects on the data values. The information in Chapters 4 and 5 on identifying materials and reporting test procedures and results can be used to highlight some of this information. Additional information on the project team is included in Chapter 2 and Refs *20* and *21*.

Plan User Involvement

The users must not only be identified as part of the project team. The plan should include their involvement at each stage of the design process. User involvement is critical to the development of a database. The fundamental criterion for evaluating a database can be stated as follows: "Is the database used?" Does it meet the users' needs? Is the database sufficiently easy to use that users consult the database rather than some other reference? The probability of achieving satisfactory answers to these questions increases with the amount of user involvement in the entire database design process.

Define the Database Functional and Performance Requirements

The functional and performance requirements must be defined based on the application and the users' needs. Functional requirements are defined in terms of queries and responses. Requests for information that the database needs to satisfy are identified in the functional requirements. The data which are required to respond to the queries are also part of the functional requirements. For example, in material selection, a single material that satisfies a set of design criteria often does not exist. If several materials meet or exceed the design criteria, all of these materials should be identified to the user. If no materials meet all of the design criteria, the database should be able to identify those that are closest. Both of these possibilities should be considered in developing the functional requirements.

The performance requirements for a materials properties database identify needs for numbers of materials, conditions (for example, temperature, environment), and properties. An additional performance requirement should be the acceptable response time for one or more of the typical queries. The functional and performance requirements are the concrete information used to specify a database that will meet the application and users' needs. ASTM Guide for Materials Database Management (E 1407) provides guidelines for additional requirements for the database such as a thesaurus and data transfer capabilities. These topics will be discussed further in Chapters 2, 6, and 8.

Select Hardware and Software

The selection of hardware and software for the database will be based on the application, the interests of the users, the functional and performance requirements, and the skills of the project team. Hardware selection for the materials property database is based on users' needs, hardware available to the users, and user's preference. Two recommendations for software should be considered. The first is use existing software, shells, database programs, tools, and so on, whenever possible. That is, *do not reinvent the wheel.* The opposing recommendation is do not use software that is inadequate for the application. The cost of software development is a significant portion of the database design, but inadequate software can make the entire project unattractive to the user community and therefore render the database unusable or unused.

Identify the Data Elements

The data elements to be included in the database need to be identified. Factors to consider in identifying the data elements include the application, the users' needs, and the best use of the data. The impact of the application and the users' needs has been discussed in general terms.

TABLE 1.5—Tester identification fields for different applications.

Fields	a. E 1313-E8	b. In-House	c. External
Engineer			●
Technician		●	
Facility		●	●
Street Address			●
City			●
State			●
Zip/Postal Code			●
Country			●
Phone		●	●
FAX			●
E-mail		●	●

To provide a more specific example, the influence of the database application on the design of a database may be considered for a small subset of information which may apply to all types of materials properties databases. Consider the identification of the person or facility performing the test. Many of the ASTM E-49 guides for the recording of test results list recommended fields for identifying the test. These fields are currently being reconciled among the different guides. The fields in Table 1.5 may be included. For a database that includes only standard, automated test results, many of these fields would not be required. ASTM E 1313, Appendix X3, does not recommend any of these fields as shown in Table 1.5, column a. For tests where the possibility of the influence of the operator must be considered, more information is required. For an in-house database, this may be the technician (Table 1.5, column b). For a database that is used to accumulate data from a number of different facilities and organizations, the engineer responsible for the test results with affiliation, postal address, phone number, facsimile number, and e-mail address may all be stored in a record (Table 1.5, column c). In a well-designed database, this information would probably not all be stored with each data record, but the information would be accessible in another part of the database. If the database stores validated data, the facility, engineer, or technician involved in performing the test may not be relevant. In this case, information identifying the validating organization and process may be of greater interest.

The best use of the data should also be considered. Current and potential future uses of the data and database should be evaluated. If the addition of a few fields or data elements would increase the current usage of the database or would reasonably be expected to increase the lifetime over which the data, or database, or both, are useful, these data elements should certainly be added. If more extensive modifications to the data element list are required, the increased cost of development and the increased software and hardware demands should be compared to the benefits of the modification. An example of this consideration is the development of a database within one department of a company. If additional fields would allow the database to also be used by another department in the company, now or in the future, that possibility should be considered now. In other words, like many other projects, the advantages and cost savings of planning ahead also apply to databases.

The majority of the guides developed by ASTM Committee E-49 are most useful at this stage in the development of the database. These guides are based on a division of the information into material identification and recording of the test procedure and results. The use of these guides requires at least one guide representing each of the types of information. For an example, consider a database for properties of metals and alloys. The fields and related data elements to identify each metal or alloy are recommended in ASTM Guidelines for the Identification of Metals and Alloys in Computerized Material Property Databases (E 1338), and the fields and related data elements to record the testing information and results are recommended in ASTM E 1313. An example with greater complexity is a database for properties of fiber-reinforced composite materials and their constituents. For material identification, ASTM Guide for the Identification of Composite Materials in Computerized Material Property Databases (E 1309), ASTM Guide for Identification of Fibers, Fillers, and Core Materials in Computerized Material Property Databases (E 1471), ASTM Guidelines for Identification of Metals and Alloys in Computerized Material Property Databases (E 1338), ASTM Guide for the Identification of Polymers Thermoset Elastomers Excluded in Computerized Material Property Databases (E 1308), and ceramics (currently under development) may all be required. Additional information regarding weaves and prepregs may also be needed. (Guides for preimpregnated material and fiber assemblies are in the planning stage in Subcommittee E49.01.) For recording test data of composite materials, ASTM Guide for Development of Standard Data Records for Computerization of Mechanical Test Data for High-Modulus Fiber-Reinforced Composite Materials (E 1434) is available. Any of the other test recording guides may be relevant depending on what properties and which materials will be included in the database. The use of the guides is illustrated in Fig. 1.3, which can be compared to Fig. 1.2, the metals and alloys example.

In considering the standard guides, the database design team may encounter some difficulty in deciding which guide to use. For example, the criteria that discriminate between a polymer and a polymer-matrix composite can be somewhat unclear. There is a gray area where materials can be considered one or the other depending on the application and the user community. There is a whole area of reinforced plastics that are described by the composite material definition, but the user community treats them as plastics. An example material that may be treated as a plastic or composite material is liquid crystal polymer (LCP). Depending on the use of the data, while LCP is not by definition a composite, it may be treated as one in a database in order to include directionality effects, which are often neglected in polymer databases. ASTM E 1308 includes an appendix that includes guidelines for discriminating between the two types of materials. This information was not included in ASTM E 1309 because the criteria are not the same as those for discriminating between metals and metal-matrix composites and ceramics and ceramic-matrix composites. (The subcommittee plans to add the tables to ASTM E 1434 when all three have been developed.) Database designers should note that the critical criterion for deciding between two different guides is as follows: Are the fields provided in one guide and not in the other needed for the particular database application?

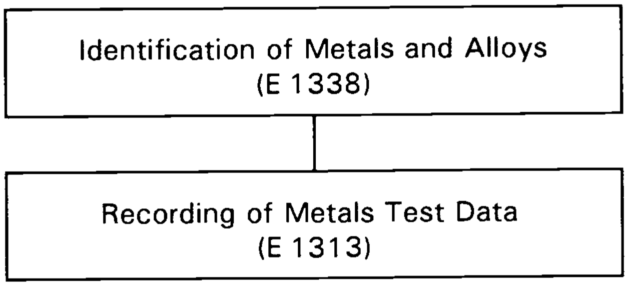


FIG. 1.2—Guides used in the development of a database for metals and alloys.

For the identification of materials, for which ASTM guides are not available, the best approach is to select the guide that seems to come closest to the situation and use that guide as a model. Similarly, the test recording guides can be used as models for situations where an appropriate guide does not exist.

Build the Data Element Dictionary

The construction of the data element dictionary is the next step in the development of the database. All information required for each data element is accumulated. For each data element, several types of information may be stored in the database or established as part of the database design. Among these are the data element name, any equivalent or synonymous names, value representation, allowed values, units, quality indicator, and any other information required to understand the data value. The data element name and any equivalent or synonymous names are discussed in the section on a thesaurus and also in Chapter 2. The value representation may be a single value, a range of values, an av-

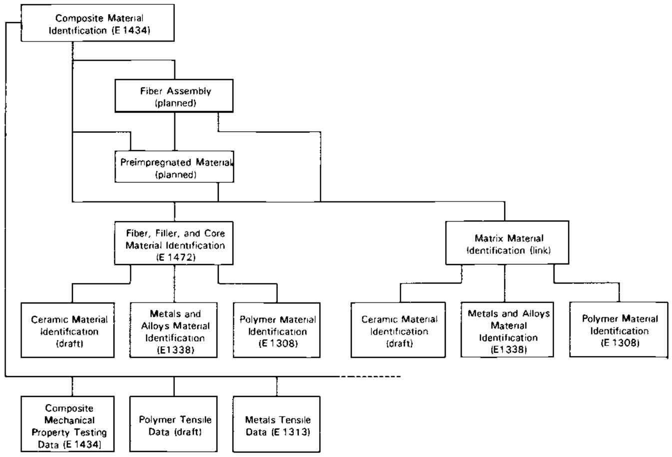


FIG. 1.3—Guides that may be used in the development of a database for composite materials and their constituents.

TABLE 1.6—Standard document identification.

Field	Value	Field	Value
ASTM, ISO, or other applicable standard method number	ASTM E 8		
Date of applicable standard	1990	Standard test method	ASTM E 8 90

erage, and average with error ranges, an average with a standard deviation or coefficient of variation, a typical value, a maximum or minimum permitted value, and so on. A quality indicator may be used to show that the value is based on a single measurement or several measurements, and to show the level of evaluation and certification applied to the original measurement(s). The allowed values need to be considered. For quantitative information, the allowed values are those values within a theoretically or experimentally determined permissible range. For qualitative information, the allowed values may be established in a value set or category set. Establishing allowed values provides the basis for data checking during data input.

The units that are used for a particular element should be considered. A set of units should be selected for the database. These are the standard units, most commonly the International System of Units (SI). The database designer, however, also needs to be aware of the units that the typical user is accustomed. For example, the user may be accustomed to seeing measurements in the inch-pound measurement system, or to certain quantities measured in centimeters or Angstroms (both now deprecated units in SI (ASTM Practice for Use of the International System of Units [SI] [the Modernized Metric System] [E 380]). If this is the case, it may be feasible to store the data in standard units, and present the data in the customary units or both standard and customary units.

Finally, additional information may be necessary to fully understand the value. For example, the elastic modulus of composite materials may be calculated several different ways. It may be necessary for a database to indicate the method of calculation as a *chord* modulus between *1000* and *6000* microstrain or an *initial tangent* modulus evaluated using a certain *curve-fitting method* with parameters *x*, *y*, and *z*. Separate fields may be required for each of these values.

The representation of several pieces of information in one field might be considered. For example, the data recording guides, discussed in Chapter 5, generally have separate fields for the standard test method number and the date of approval. These two fields may be concatenated into one field, providing that data entry checking is sophisticated enough to check for the three different pieces of information (Table 1.6). On the other hand, three fields may be used: (1) the standards producing organization, (2) the standard method number, and (3) the date of approval. For standard test methods where more than one method is included in the same document, more information may be included, for example, ASTM Test Method for Compressive Properties of Unidirectional or Crossply Fiber-Resin Composites (D 3410-B) indicates the compression test of a composite material using conical wedge grips.

The use of generic constructs or blocks of fields where one or more fields define the remaining fields may be suitable for the database design. An example of the use of a generic construct is included in ASTM E 1471 where different types of dimensions are used to characterize fibers, fillers, and core materials. Table 1.7 shows the generic construct used and the value sets appropriate for two of the fields. The generic construct eliminates the need for a separate length field, width field, inside diameter field, and so on. In other words, these fields are the equivalent of the 60 fields, the 12-dimension parameter fields combined with the 5-dimension distribution parameter fields (including the sample size). This is particularly efficient when the number of fields that are used for any one record are small. These fields, as well as other fields in the ASTM guides, may need to be repeated.

Group Data Elements

Most database programs are more efficient when the data elements are grouped in some logical fashion. These groupings may depend on the identification and test results of individual specimens as compared to groups of specimens, fields that may be repeated for a single data set, or data elements. Some indication of element grouping is apparent in the data formats. An example of individual specimens and groups of specimens is ASTM E 1434. Many of the test methods to which E 1434 applies, require the reporting of statistical parameters for groups of specimens, such as the average, standard deviation, and coefficient of variation. ASTM E 1434 recognizes that the designers of many databases would like to include the results for individual specimens, and the guide provides fields for both types of approaches. In the case where data are reported for individual specimens, the relevant group of fields is repeated for each specimen.

Another example of repeating fields is shown in Table 1.6. Both of the example dimension parameters for a fiber may apply to the same fiber. Provision should be made in a database containing fiber information to repeat fields 13 through 17 in Table 1.6 as necessary. Other logical groupings of data elements may be data elements related to specimen conditioning, specimen preparation, composition, experimental procedure, and so forth. Such groupings may be based on the data input but should certainly consider the groupings natural to the database user who is accessing the information.

Identify the Retrieval Characteristics of Each Group

For any database structure, the retrieval characteristics of the group should be considered. The groups discussed so far have not really depended on the type of database structure that is used. That is, the structure of the database could be relational, object-oriented, or hierarchical. The implementation of the groupings within the database depend on the structure of the database. At this point, it is easier to continue the discussion assuming a relational database structure. Similar considerations apply to other database structures.

TABLE 1.7—Dimension information for the identification of fibers, fillers, and core materials.

Field No.[a]	Field Name	Value Sets or Units	Example: Filler (particulate)	Example: Fiber	Example: Fiber
13	dimension parameter	see Value Set 1	median size	diameter	filament count
14	dimension value	floating point	2.3 μm	0.145 mm	12 000
15	dimension distribution parameter	see Value Set 2	standard deviation	NA	NA
16	dimension distribution parameter value	floating point	1.2 μm	NA	NA
17	dimension distribution sample size	integer	5	NA	NA

Value Set 1		Value Set 2
Length	cell size	standard deviation
Width	percent open cell	range (+−)
Inside diameter	denier	coefficient of variation, %
Outside diameter	filament count	other (specify)
Thickness	fiber yield	...
Wall thickness	other (specify)	...

[a] Field numbers are provided for referencing Table 6 in E 1471; no other meaning should be attributed to them.

In a relational database, the groupings would be files or tables. As in other types of databases, the field or fields used to identify and access a particular record in a table or file should be essential and required to be filled. Those fields that are identified as essential fields in the guides may be considered as candidate keys preferentially over other fields.

Identify the Relationships Between Groups

Relationships between the groups will depend on the database structure selected and the groupings. If the application for the database is archival of experiments, the designer may have decided to include the material identification information for each specimen. In such a case, there may only be one group (of file), containing both material identification and test recording information. In many cases, it may be more efficient to store material identification in one group and the individual specimen results in another. How records in one group are associated with records in another group depends on the database structure. It is very important that these relationships be considered during the database design process.

Develop the Database Schema

The database schema is the design of the database structure. It should include all of the information accumulated and decisions made in the steps prior to this one. Three basic schemas have been defined in ANSI standards: (1) a conceptual schema, (2) an internal schema, and (3) an external schema (discussed in Refs *17* and *25*).

Steps 6 through 11 have developed the conceptual schema of the database. Normalization of the schema should be considered at this time. Normalization is the elimination of redundancies among the database fields. Further discussion of the development of the database schemas is beyond the scope of this chapter but is covered in most database texts including Refs *5*, *17*, and *18*. The reader with a background in database design and management should understand these procedures. A reader with a materials background should consult the database expert on the project team or one of the database texts referenced in this chapter. Additional information on the design and implementation of a materials property database is included in the other chapters of this manual, particularly Chapters 2, 6, and 7.

CONCLUSIONS

This chapter has considered the value of materials property databases, key concepts in their design and implementation, standards that are available to assist in the design and implementation of materials property databases, and when those standards can be used in the design process. The three most important recommendations from this chapter are as follows:

- Assemble a database team representing materials expert(s), database expert(s), and users.
- Use the ASTM E-49 formats as a starting point for the development of your data dictionary and database schema.
- Tailor your database schema to your application.

The next four chapters of this manual discuss the database infrastructure, types of data and applications, and the standards for material identification and recording test results. Chapters 6 through 8 consider data evaluation, the management of a materials property database, and data transfer. The final chapter provides an example of a materials property database implemented using EXPRESS.

REFERENCES

[*1*] Barrett, A. J., "The Benefits and Economic Consequences of Materials Property Databases," *Computerization and Networking of Materials Databases: Second Volume, ASTM STP 1106*, J. G. Kaufman and J. S. Glazman, Eds., American Society for Testing and Materials, Philadelphia, 1991, pp. 17–25.

[*2*] Barrett, "The Provision of Materials Property Data via Computerized Systems: the Role of CODATA," *Materials Information for the European Communities*, N. Swindells, N. Waterman, and

H. Kröckel, Eds., Commission of the European Communities, Luxembourg, 1990, pp. 7–11.

[3] Rumble, J. R. Jr., "Socioeconomic Barriers in Computerizing Material Properties Data," *Computerization and Networking of Materials DataBases, ASTM STP 1017*, J. S. Glazman and J. R. Rumble, Jr., Eds., American Society for Testing and Materials, Philadelphia, 1989, pp. 216–226.

[4] *Guide to Material Property Database Management*, CODATA Bulletin No. 69, Nov. 1988.

[5] Rumble, Jr., J. R. and Smith, F. J., *Database Systems in Science and Engineering*, Adam Hilger, Philadelphia, 1990.

[6] Reynard, K. W., "VAMAS Activities on Materials Data Banks," *Computerization and Networking of Materials DataBases, ASTM STP 1017*, J. S. Glazman and J. R. Rumble, Jr., Eds., American Society for Testing and Materials, Philadelphia, 1989, pp. 43–52.

[7] Kaufman, J. G., "The National Materials Property Data Network, Inc.—A Cooperative Approach to Reliable Performance Data," *Computerization and Networking of Materials Databases, ASTM STP 1017*, J. S. Glazman and J. R. Rumble, Jr., Eds., American Society for Testing and Materials, Philadelphia, 1989, pp. 55–62.

[8] Swindells, N., "An Outline Description of the Demonstrator Programme," *Materials Information for the European Communities*, Commission of the European Communities, 1990, pp. 55–61.

[9] *Regulations Governing ASTM Technical Committees*, American Society for Testing and Materials, Philadelphia, July 1992.

[10] Kaufman, J. G., "Standards for Computerized Material Property Data—ASTM Committee E-49," *Computerization and Networking of Materials Databases, ASTM STP 1017*, American Society for Testing and Materials, Philadelphia, 1989, pp. 7–22.

[11] Rumble, J., "Standards for Materials Databases: ASTM Committee E-49," *Computerization and Networking of Materials Databases: Second Volume, ASTM STP 1106*, American Society for Testing and Materials, Philadelphia, 1991, pp. 73–83.

[12] Bylaws, ASTM Committee E-49 on Computerization of Material and Chemical Property Data, American Society for Testing and Materials, Philadelphia, Feb. 1992.

[13] Westbrook, J. H. and Grattidge, W., "The Role of Metadata in the Operation of a Materials Database," *Computerization and Networking of Materials Databases, Second Volume, ASTM STP 1106*, J. G. Kaufman and J. S. Glazman, Eds., American Society for Testing and Materials, Philadelphia, 1991, pp. 84–102.

[14] Gasiunas, A. A., Halliday, T. D., and Berman, P., *European Materials Databanks Demonstrator Programme, Common Reference Vocabulary*, Commission of the European Communities, Luxembourg, 1988.

[15] *ISO Bibliography 8, National Standard Vocabularies*, 3rd ed., International Organization for Standardization, Geneva, Switzerland, 1985.

[16] *Compilation of ASTM Standard Definitions*, 7th ed., American Society for Testing and Materials, Philadelphia, 1990.

[17] Date, C. J., *Introduction to Database Systems*, Addison-Wesley Publishing Company, Reading, MA, Vols. 1 and 2, 4th ed., 1987.

[18] Korth, H. F. and Silberschatz, A., *Database System Concepts*, 2nd ed., McGraw Hill, Inc., New York, 1991.

[19] *Compilation of ASTM Standard Definitions*, 7th ed., American Society for Testing and Materials, Philadelphia, 1990.

[20] Atre, S., *Data Base: Structured Techniques for Design, Performance, and Management*, 2nd ed., John Wiley & Sons, New York, 1988.

[21] Gorman, M. M., *Managing Database: Four Critical Factors*, QED Information Sciences, Inc., Wellesley, MA, 1984.

[22] Gorman, M. M., *Database Management Systems Understanding and Applying Database Technology*, QED Information Sciences, Inc., Wellesley, MA, 1991.

2

Program Infrastructure

Edwin F. Begley[1]

INTRODUCTION

What Will and Will Not Be Included in This Chapter

This chapter will focus on the methodology and basic system components needed for successful implementation of a materials database. The discussion will be presented in a "hints, tips, and everyday wisdom" manner so that the reader, hopefully, will come to recognize troublesome areas and, perhaps, sidestep or hurdle them when possible.

Some readers may wish to delve into the underlying theory of database systems, which will not be explored in this chapter. Excellent, extensive, and eloquent references are available, including Date [*1*], Martin [*2*], Ullman [*3*], and Weiderhold [*4*].

At times, the use of database jargon will be unavoidable. The last section of this paper is a glossary of terms that should help the reader.

Building a materials database is not achieved simply by applying some technical programming skills to an information problem. In fact, writing code is actually a small part of the total effort. Balancing the interactions of the people involved in the project is the true challenge, and so, the rest of this introduction will spotlight them: the users, the data providers, the software engineers, and the project managers. It is important to realize that it is not unusual for an individual to fall into more than one of these categories, but each will be treated separately in the subsequent discussion.

The Vision of the User Community

The user of a scientific database generally assumes a more active role in its design than the user of a business or commercial database application primarily because the technical nature of the stored information requires special attention. Scientists and engineers can best define how the numeric data will be used in their calculations, experimental designs, and product development, which influences the specifications for the database. The requirements for commercial applications, in contrast, are dictated most often by management needs, but the software systems are rarely used directly by managers. In this sense, the scientific user community's vision for the uses of the final production database carries special weight. Sometimes the intended uses differ significantly, and, as will be discussed in the section on the responsibility of the software engineer, the software must be written so that any conflicts that might arise are minimized.

The Contribution of the Data Provider

The data provider may often be the individual who also performs experiments and collects data. In addition to the responsibilities of designing the experiment properly (calibrating instrumentation, using appropriate measurement standards and sampling procedures), the provider must meticulously record all experimental details. If this information is not reported, the data loses their context and may become somewhat less valuable to the user. This problem can be particularly acute when computerized data acquisition systems are used, since procedural details may be embedded in software. And, even if data are accurately reported and measurement details are documented, the data provider is likely to encounter variations in the actual measured values. In such events, the data should be critically evaluated to establish a level of validity acceptable to the individuals who will use the data.

Data providers who are not involved with the generation of data, but rather its location, must possess knowledge of where public and private sources may be found. Locating information requires contacts with colleagues, identification of reprints, reports, directories of data sources, and referral centers, especially computerized data centers and information services.

The Responsibility of the Software Engineer

The primary responsibility of the software engineer, simply stated, is to construct a database system that satisfies the broad spectrum of user needs. This can be a difficult task particularly when a diversified user community exists and data arrive from disparate sources. Software must be written that allows each distinct group of users to view the data in a manner appropriate to its needs. This implies choosing the best structure for storing the data and building a flexible user interface that is properly designed only in close collaboration with the users themselves.

Furthermore, the software engineer must be prepared to

---

[1] Computer scientist, NIST, Building 223, Room A256, Gaithersburg, MD 20899.

Copyright© 1993 by ASTM International

manage data that are provided in nonuniform formats. This usually requires conversion software based upon an understanding of the rules for interpreting the data.

In essence, the software engineer must reconcile the different user requirements for the database and the contrasting formats of the supplied data while maintaining the project's budget and schedule.

The Support of the Project Manager

To provide quality support for the development of a materials database, the project manager must be caring, technically competent, and respected. A good project manager will realize immediately that people (users, data providers, and software engineers) are the primary participants in the construction of the database, not computers and software. The needs and demands of all these individuals must be carefully considered and given balanced judgment. Naturally, without a technical understanding of the issues involved and thorough attention to details, the project manager would be incapable of making pertinent choices and would rapidly lose the respect required for successful leadership.

The project manager carries the responsibility for the entire project and must ensure that the database is completed on time and within its budget. She must also ensure that the software meets the users needs, functions well and easily, and has been carefully tested and is as error free as possible. Additionally, she must exercise foresight to determine how the materials database might be integrated with other applications of interest to the user community.

Lastly, she must be flexible enough to know when to compromise and when to stand firm when making decisions regarding the difficult issues and unforeseen problems that inevitably arise.

THE DEMONSTRATION SYSTEM AND THE PROTOTYPE

What Are They and How Are They Different?

The purpose of building a demonstration system is to identify the goals, problem characteristics, resources, and participants pertinent to the database project. Implementation should occur as quickly as possible with the primary emphasis on problem scope and representation. While the demonstration system is not intended to be complete, it serves as a vehicle for understanding the problem, the information required to solve the problem, and the possible routes to solution of the problem.

The prototype sharpens the definition of the concepts, subproblems, and control features identified during the demonstration stage and transforms them into a working system.

Why Are They Important?

The most important reason for developing demonstration and prototype systems is to tie database functionality directly to user expectations by gaining timely feedback on how well the development process meets those needs *before* the project begins to drift. In the past, development methodologies tended to place too much emphasis on "up-front" analysis. Most of a project's funding, sometimes as much as 70%, would be spent on analyzing system requirements and writing functional specifications. Once the analysis was completed, coding would begin, and the user ultimately would be presented with a final product. This approach lacked a clear system of checks and balances, and the user all too often would arrive at the uncomfortable and difficult position of having to accept, without redress because of exhausted project resources, a system that did not adequately meet stated needs.

After enough failures of this kind, project managers and software engineers adopted a different approach to database development known as the iterative methodology. The key aspect of this methodology is sustained user participation in every stage of database implementation. Development proceeds as a series of well-defined tasks leading to demonstration and prototype systems, respectively. The user reviews the results of each task and must give approval before work on subsequent tasks begins. If the user disapproves, the task in question is revisited until it is satisfactorily completed.

Certainly, analysis is required but demonstration and prototype systems serve as excellent checkpoints for assessing how well the database design reflects the analysis. Superficially, the iterative methodology appears time consuming, but it more than adequately ensures that the database project is developed according to the needs of the user.

How Are They Built?

The demonstration system is built almost exclusively by the user, although the software engineer provides extensive encouragement and guidance. The goal is to get the user to lay out ideas on paper. For example, the user probably has some thoughts regarding how interaction with the software should look and proceed. The software engineer will assist the user with designing interface screens such as menus and the link between the menu's choices and the specific actions each choice initiates. Clearly, there is a certain level of functional specification occurring at this point, which is completely user driven. When the user and software engineer mutually agree that enough progress has been made, computer code will be written that simply gives a "look and feel" quality to the ideas on paper. Stubs, which are essentially placeholders for code not yet available, are written for the actions corresponding to menu choices, for instance. Then the demonstration package is presented to the user who essentially tries it out, fully aware that it is not yet a database system but rather a tool for tightening requirements and specifications. The user, of course, may ask the software engineer to make some adjustments and when completely satisfied will give approval for prototype development to begin. During the prototyping stage, fully functional software is written, that is, stubs are fleshed out, and the user can perform searches of the database and display retrieved information.

COMPUTING FACILITIES

The Hardware and Software Conundrum: Work With What's Available or Start From Scratch?

Before work on the database begins, a very difficult decision must be made regarding the choice of a development platform, that is, the combination of hardware and software necessary to implement the database system. There are intertwined constraints composed of budgets, deadlines, people, and politics, and it is unlikely that a clean, straightforward resolution of the hardware and software conundrum exists. The project manager, therefore, must be capable of making a firm, potentially unpopular, decision based on experience and common sense. A streamlined example of some questions to be addressed and some plausible conclusions is presented in Fig. 2.1.

SYSTEM ARCHITECTURE

The Data Dictionary

A data dictionary is essentially a guide for understanding the information in a database and has features similar to those found in a language dictionary. In the data dictionary, one can find a description, the origin, and the usage of each specific piece of data presented in the database. Unlike a language dictionary, a data dictionary will provide additional information describing the relationship of a given piece of data to all other pieces of data. It will also indicate who has responsibility for ensuring the quality of a particular piece of information and which format best fits the data, such as numeric, alphanumeric, date, or customized [5].

The data dictionary is the framework on which the database is built. A great deal of work has already been done by ASTM, which eases the development of dictionaries for material property databases. ASTM Subcommittee E49.01 on Identification of Materials has issued standard guides for the identification of metals and alloys (ASTM Guide for Identification of Metals and Alloys in Computerized Material Property Databases [E 1338]) and aluminum alloys and parts (ASTM Guide for Identification of Aluminum Alloys and Parts in Computerized Material Property Databases [E 1339]) in computerized material property databases. Subcommittee E49.02 has issued a guide (ASTM Guide for Development of Standard Data Records for Computerization of Material Property Data [E 1313]) for the development of standard data records for computerization of material property data. The guidelines contained in these standards help the developer define the informational content of the database. For example, Table 2.1 lists fields useful for the generic identification of metals and alloys. The discussion of this table in ASTM E 1338 includes definitions for each field.

Internal Consistency

Diverse user communities often see the same set of data in different ways, and problems may arise from misunderstandings associated with differing viewpoints. Building a data dictionary is a very effective way to extract whatever elements are common to all viewpoints and thereby create a consistent, logical meaning for particular pieces of information. This is especially important when data providers with different backgrounds and interests need to understand what data are required of them.

In addition to enabling a user to understand the contents of the database, the dictionary imposes internal consistency in the way data are entered and stored. An individual will not be able to add alphanumeric information into a field that

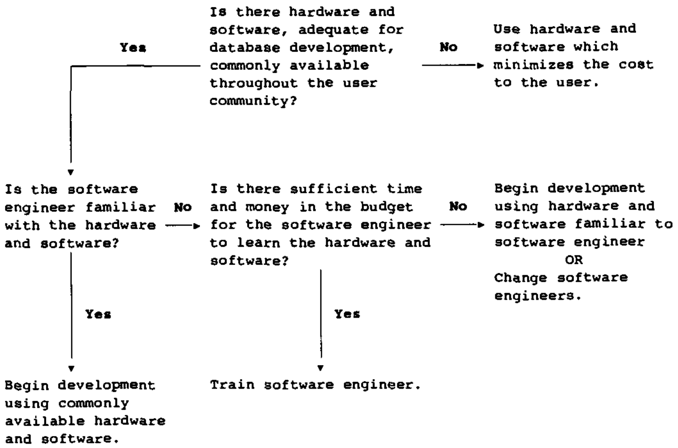


FIG. 2.1—Questions and possible actions regarding hardware and software choices.

TABLE 2.1—Sample format—generic identification of metals and alloys (ASTM 1338).

Field Number[a]	Field Name	Category Sets, Values, or Units
1	Material Class	Alphanumeric String
2	Family Name	Alphanumeric String
3	Family Subclass	Alphanumeric String
4	Application Group	Alphanumeric String
5	Product Group	
6	Specification Organization[b]	Alphanumeric String
7	Specification Number[b]	Alphanumeric String
8	Specification Version[b]	Alphanumeric String
9	Specification Designation[b]	Alphanumeric String
10	Unified Number System Number[b]	Alphanumeric String
11	Common Name[b]	Alphanumeric String
12	Compositional Detail (Key)[b]	Alphanumeric String
13	Elemental Symbol[b]	Alphanumeric String
14	Measured Weight Percent[b]	Floating Point
15	Minimum Weight Percent[b]	Floating Point
16	Maximum Weight Percent[b]	Floating Point
17	Manufacturer	Alphanumeric String
18	Country of Origin	Alphanumeric String
19	Manufacturer's Plant Location	Alphanumeric String
20	Production Date	YYMMDD
21	Manufacturer's Designation	Alphanumeric String
22	Lot Identification	Alphanumeric String
23	Additional Detail (Key)	Alphanumeric String
24	Primary Process Type[b]	Alphanumeric String
25	Primary Process Detail (Key)	Alphanumeric String
26	Secondary Process Type[b]	Alphanumeric String
27	Secondary Process Detail (Key)	Alphanumeric String
28	Part Identification Number[b]	Alphanumeric String
29	Geometric Shape[b]	Alphanumeric String
30	Thickness[b]	Floating Point mm (in.)
31	Width	Floating Point mm (in.)
32	Length	Floating Point mm (in.)
33	Fabrication History[b]	Alphanumeric String
34	Fabrication Details (Key)	Alphanumeric String
35	Service History[b]	Alphanumeric String
36	Service Details (Key)	Alphanumeric String
37	Supplementary Notes	Alphanumeric String

[a] Field numbers are for discussion purposes only.
[b] Essential field, if applicable.

has been defined as strictly numeric, for example, or will need to enter data in a specific format if required.

Interlinking Capability

Another important function of the data dictionary lies in its use for interlinking applications. An industrial site, for instance, might maintain a materials database for coordination of research results, for computer-aided design (CAD), and for materials selection and ranking. The CAD package might require property measurement values for a material to be used in a design component. The data dictionary should provide sufficient information about the content of the database for an interface to be built, which will identify the relevant property fields, retrieve the needed data, and pass them to the CAD package for subsequent processing. The same materials database might also be used by an expert system, which will select and rank candidate materials according to some preferred criteria for pre-design engineering analysis. It would be wasteful to maintain three separate, identical databases to satisfy these different needs. The interlinking capability of the data dictionary can greatly reduce costly, redundant, and very likely, inconsistent facilities for collecting and storing information, which is useful for multiple applications.

The Schema and Subschemas

A schema is a perspective, a way of seeing the information in a database. The three widely accepted schemas are called conceptual, physical, and external or subschema.

The conceptual schema is the complete, logical view of the entire database including the data dictionary along with the data existence requirements and constraints. This conceptual viewpoint is represented by the fields, record structures, files and file relationships used to build the database.

The physical schema is basically the viewpoint of the computer's operating system and includes descriptions of database file characteristics, that is, physical layout including field and record sizes, links to indexes, and links among files that comprise the database.

The external or subschema is the user's and often a program's view of the database. The subschema is so named because frequently the user or program is presented with only a subset of the full contents of the database. One user, for instance, may have use for certain information contained in the database and, therefore, will only be able to "see" that particular subset whereas another user with different interests may be presented with another subset of appropriate data.

Subsequent subsections and sections of this chapter will explore the conceptual and external schemas. Discussion of the physical schema would involve examining the low-level nuts-and-bolts of the database management system, which is beyond the scope of this chapter; but, the reader developing an understanding of the database program infrastructure will not suffer significantly from such an omission.

Fields, Records, and Files

A field is the smallest piece of information contained in the database. As an example, a field, named MANUFACTURER, can be defined to contain the name of the material's manufacturer, nothing more and, generally, nothing less.

A record is a collection of related fields. Looking back to Table 2.1, the combination of the 37 tabulated fields would be a record for the generic identification of a metal or alloy. All records pertaining to metals and alloys could be combined to form a file. And, finally, a database is a collection of related files. Building on the example provided by Table 2.1, one might have a file for the identification of metals and alloys, another file containing particular thermal property measurements for each metal and alloy, and perhaps others for additional property measurements or supplementary information. Linking all these files creates a database.

Data Existence and Constraints

Data existence and constraints refer to the checks and balances for ensuring the quality and validity of the data and, in turn, the quality of the database itself. Generally, the database developer is concerned primarily with field and record existence and constraint issues.

Existence refers to whether a field must have a value associated with it. It may not always be possible to provide a name for the MANUFACTURER in the identification file. If a name can be provided, the developer may wish to define some constraints on its type, size, syntax, and value. Perhaps the type should be an alphanumeric string rather than a number or date. Its size would most likely be restricted to less than 35 characters. Its syntax might enforce last name followed by a comma followed by first name, and its value may not include certain manufacturer names.

At the next level, the developer could decide to reject records if one field or a combination of fields did not meet some criteria. An identification record for the metal or alloy might be rejected if information for any of the essential fields (Table 2.1) was missing.

Data Entry

Building data dictionaries and database schemas sometimes appears straightforward compared to the problem of actually entering data into the database system. The next three sections will discuss useful approaches to streamlining the data entry process.

The Importance of Developing a Data Reporting Format

Often data exist on paper in formats unsuitable for direct entry into the database. In such cases, a data reporting form based on the fields included in the database schema and subschemas will assist data providers as well as those performing data entry. The form or worksheet establishes guidelines and promotes consistency and completeness for data reporting. This in turn eases the data entry process. Appendix A [*6*] is one example of a data reporting format and was designed for the Structural Ceramics Database [7] project at the National Institute of Standards and Technology using ASTM Committee E-49 guidelines for the characterization of the data source and for the specification of the material.

Conversion Software for Alternative Formats

A data reporting worksheet may, of course, be of limited value if the data had been previously recorded in machine-readable form. In such cases, customized data conversion programs may need to be written to rearrange the source data into the structure required by the database schema. To successfully prepare conversion software, it will be necessary to understand: (1) the logical and physical structure of the machine-readable data, (2) the rules relevant to interpretation of that data, and (3) the connection between the data items to be converted and the database fields and records. Ideally, information useful for achieving this understanding will be available, but sometimes it will be incomplete, and the only recourse will be careful analysis coupled with common sense.

What to Do When the Data Arrive on Storage Media Different From the Target Medium

Periodically, it will not be possible for the data provider to send machine-readable information on a storage medium convenient for conversion and data entry. Furthermore, equipment may not be available to those working on data entry to convert from the source medium to the target medium. So what can be done when the data arrives on 2400-ft reels of 9-track 800 bpi magnetic tape[2] and can only be converted and loaded if it is on 3.5-in. high density diskettes,[2] but equipment is not available to change the medium? Find the section entitled "Data" in your local Yellow Pages and look at the "Data Processing Services" subheading for businesses that specialize in data conversion. Also, cross-check the "Computers—Software & Services" section listed under "Computers." Many of these companies also offer a scanning service for nonmachine-readable formats.

The User Interface for the Storage, Retrieval, and Display of Data

One of the most important considerations in the computerization of materials property data is the interaction between the computer and the user. A database is typically a large collection of files and fields, which can present a daunting navigational challenge to finding desired information, particularly if the user must learn the syntactic and semantic subtleties of the language of the database management system. To eliminate this burden and to assist the user in locating information effectively, a "user-friendly" interface is coded which allows access to the database without requiring a knowledge of file structures, field attributes and characteristics, or languages.

User-friendly to a certain extent is a subjective evaluation. The interface designer, however, can influence this evaluation by gaining a knowlege of the user's subject domain including its terminology and designing with an emphasis on clarity, consistency, error handling, and help facilities. The interface must meet these requirements in each of its three major components: information storage, retrieval, and display.

TABLE 2.2—Techniques for reducing or increasing displayed information.

Reduce information by:
- Providing graphic rather than alphanumeric displays
- Formatting displays to correspond to the user's immediate requirements
- Providing less powerful commands
- Providing less complex interactions

Increase information by:
- Providing more powerful commands
- Providing more complex interactions

Visual Real Estate

The display area, or visual real estate, of a video monitor is finite and, consequently, valuable. The software engineer must take care to use the given area effectively when implementing the interface between the user and the materials database. The primary design focus should be on the quantity of information presented. There are a number of techniques available to the software engineer for reducing or increasing the amount of information displayed (Table 2.2) [*8*].

In addition to the quantity of information, attention must be given to the appearance of the interface, since it can easily enhance or detract from an individual's ability to use the database. In this respect, aesthetics is an important part of each screen's design and plays a large role in user acceptance. Some general guidelines are to keep all interface screens simple and uncluttered. Color combinations should be minimized to avoid distractions and to maintain consistency among all screens. Space should be allocated for the user to see the results of any actions taken. All possible actions from a screen should be presented to the user, preferably in the same location on all screens. Among these actions should be a method for getting help. Finally, some portion of the screen should be reserved for announcing the present status of the system [*9*]. It is disconcerting to most users, for example, when the display simply remains blank and impassive during a search of the database, particularly during long searches. It is much better to write a few lines of code which generate a message indicating that a search is in progress. A few more lines of code may even be used to tell how far the search has progressed.

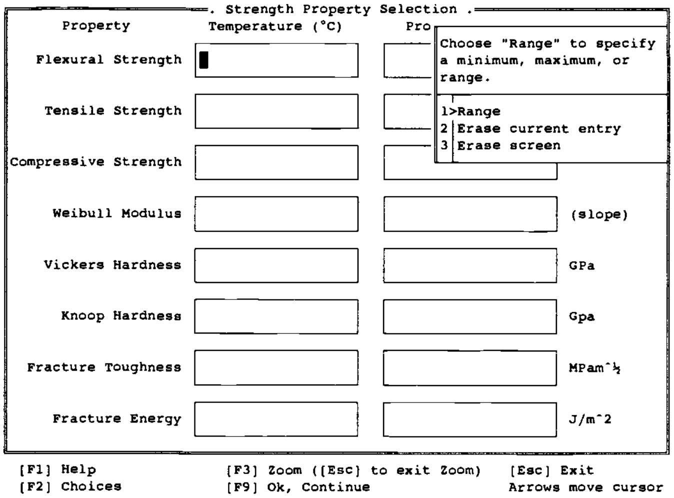
Of course, there are wide variations in hardware and software that might dictate the use of different styles of interfaces: line-by-line mode, full-screen mode, and graphical user interface, or GUI (pronounced goo-ee), but all of the design guidelines for the presentation of information can and should be observed.

Line-by-Line Mode

Many systems still exist that only write to the display one line at a time with limited cursor control. Generally, menus that list choices for the user's selection are used extensively under these circumstances. When composing menus, consider the order and number of the menu items. The user will most likely have some preference for ordering these items by relative importance or for grouping similar items. And, certainly, menus with too few or too many selections are to be avoided. Lastly, using the menu system becomes increasingly complex as the number of menus rises. Look for ways

---

[2] Commercial designation.



```
                    . Strength Property Selection .
      Property         Temperature (°C)       Pro
                                                 Choose "Range" to specify
  Flexural Strength                              a minimum, maximum, or
                                                 range.

   Tensile Strength                              1>Range
                                                 2 Erase current entry
                                                 3 Erase screen
Compressive Strength

     Weibull Modulus                                           (slope)

    Vickers Hardness                                           GPa

      Knoop Hardness                                           Gpa

  Fracture Toughness                                           MPam^½

     Fracture Energy                                           J/m^2

 [F1] Help              [F3] Zoom ([Esc] to exit Zoom)   [Esc] Exit
 [F2] Choices           [F9] Ok, Continue                Arrows move cursor
```

FIG. 2.2—An example of a full-screen fill-in-form. The user may move the cursor to the appropriate prompt (represented by rectangles) and specify values for the property and/or measurement temperature.

to reduce the number of menus without sacrificing clarity or functionality. If the number of menus cannot be reduced, use system status messages and screen titles to help the user navigate.

Full-Screen Mode

A full-screen, or block, mode interface offers more flexibility than the line-by-line design because memory buffers are used to store and manipulate the screen. One major benefit is flexible cursor control. The software engineer can sense the location of the cursor and also move it as desired. As a result, such constructs as fill-in-forms may become part of the interface. These forms greatly simplify data entry and query specification. Figure 2.2 is an example of a full-screen fill-in-form.

The Graphical User Interface

One of the most recent significant advances in computing is the development of the GUI. GUIs are immediately recognizable because of their visual orientation. They have windows, pointing devices, icons (small pictures linked to actions), buttons, menus, and dialog boxes. Output to the screen is in the form of text, pictures and icons with input taking place either directly or indirectly (through dialog boxes, for instance) through the same screen.

The GUI is definitely user-oriented but can present complex technical challenges to the software engineer. Even the simplest GUI includes long lists of features and options. Also, the learning curve is very steep. Tools, however, are gradually becoming available that help streamline the development of a graphical user interface.

Software Documentation

Some readers may wonder why software documentation is included in this section on system architecture. It is the author's humble opinion that integrating documentation with the other database components greatly facilitates understanding and maintenance of the software. Yet, for many software engineers, documentation is anathema. Often, they will claim the software is self-documenting, that is, that it is so clearly written that it explains itself. Generally, this statement is true only for the programmer herself and only for a finite period of time when the code is still fresh in her mind. Revisiting a section of code even one month from its writing quickly reveals the limits of one's memory. Another protestation is that documenting software wastes precious time. It does not. The author has had full meals of spaghetti code without documentation condiments and can claim unequivocally that such repasts are distasteful to the palate, difficult to swallow, nearly impossible to digest, and have kept him at the dinner table for uncomfortably long periods of time.

Good documentation is actually quite easy to write. It need not be verbose, just clear and concise. Explanations should include what a module or section of code does, what the variables mean and how they are used, and a discussion of any special conditions or features that exist. The documentation should be placed close to the code it refers to and distinguished from the code itself by being placed in a box.

After the database project has been completed, a section known as the "change list" should be created at the very beginning of the software. The change list includes the date a change was made to the software; the name of the individual who made the change and some information on how to contact her; a short, clear, and concise discussion of the nature of the change; and a list of the sections of code that were affected by the change.

SYSTEM FEATURES

The Basics

The normal capabilities of a database system, data storage, retrieval, and display, should be supplemented with a few basic features that ease and enhance the user's interaction with the software. These are data security; a help facility; a status line; data indexing; brief and full display capabilities; data downloading; and data facilities for accuracy, significant figures, and quality indicators.

Security

The database software and the information contained in the database must be protected from unintended and, within reason, malicious deletion or alteration. The former is relatively easy to accomplish because modern hardware and software provide sophisticated control features. The latter can be more difficult. There will always be unscrupulous individuals insufficiently clever to find meaningful ways to occupy their minds, and thwarting them requires good and regular backup procedures.

Help Facility and Glossary

The most basic and arguably the most important feature of a thoughtfully implemented interface is the help facility. If the user does not understand the meaning of a term, does not know what to do next in a particular circumstance, or has made an error, help needs to be available. It is very common to establish a function key, generally F1, for what is know as "context-sensitive" help. Context-sensitive simply refers to what the user is currently trying to do. For instance, the user may need to specify a value for a particular prompt but does not know what that prompt means. By writing software that "knows" on which prompt the cursor is located, that is, the context is known, the F1 key will supply help appropriate for that prompt. Also, general help may be given for the database system as a whole and is usually available by pressing key combinations which include F1, such as Shift+F1 or Ctrl+F1. Alternatively, systems with mouse support often have a button simply labeled "Help." After clicking on the button, a help facility is revealed.

Error Handling

Humans are error prone and, with respect to a database system, can make mistakes as end users or as software engineers. Errors are caused by any combination of misunderstandings, lack of appropriate information, lapses in logic, syntax errors, or inadvertent mistakes.

The incidence rate for software errors can be greatly reduced by proper software design and implementation. Software should be written in components called modules. Each module will perform a specific task and can be tested independently. As a result, it is more straightforward to identify and correct programming errors. An additional and important benefit of the modular design is that it simplifies revisions to the software, since only those modules requiring changes are affected. Software modularity will be revisited in the section on Getting the Software to Work Right.

Errors on the part of the user may also be significantly reduced by altering the software design. Menus, function and hot keys, buttons, icons, and mouse support are among the most well-known constructs created to simplify the user's interaction with software and thereby reduce frustration and increase efficiency. Menus, for example, can be very effective in the reduction of typographical errors and can also provide a good reference frame for the user of a complex database system. Function keys, hot keys, buttons, and icons with clearly defined actions and meanings relieve the user of the burden of learning a command language. The mouse streamlines the user interface by providing a point-and-shoot alternative to cursor and "Enter" key combinations when working with menus, icons, and buttons.

Despite the best efforts of the software engineers, errors are still likely to occur. It is important to recognize this fact and prepare for it by providing helpful error messages and graceful recoveries.

Too often, cryptic or seemingly nonsensical messages are displayed that only confuse and annoy the user. A common example is "Invalid entry, try again." What was wrong? What should be tried next? It takes just a few lines of code to run a value through an edit check, discover what was wrong, and inform the user clearly and concisely. Cutting corners on error trapping and reporting can undermine an otherwise well-written program.

```
Records in Database                      Index for Manufacturer Field in
                                         Database

Record Key: 1                            Record Key: XYZ Corporation
Material: silicon nitride                Pointer: 1
Manufacturer: XYZ Corporation            Pointer: 3
Designation: XYZ-100
Impurities: Fe                           Record Key: ABC, Ltd.
                                         Pointer: 2
Record Key: 2
Material: silicon carbide
Manufacturer: ABC, Ltd.
Designation: ABC-1
Impurities: Al, Ti

Record Key: 3
Material: silicon nitride
Manufacturer: XYZ Corporation
Designation: XYZ-250
Impurities: Al
```

FIG. 2.3—An example of an index for a field in a database.

Recovering from an error should not take the user by surprise. The best action would be to provide a helpful message and return the user to the point just prior to where the error occurred.

The Status Line

One or two lines on the video display should be reserved for messages that provide information about where the user is in the database system and what kind of work is being performed. Typically, they are located at the bottom of the screen and are displayed using video attributes different from their surroundings for easy identification. The status lines are particularly useful for short help and error messages. Many systems employ color, such as red for error message text, to draw the user's attention to particularly important information.

Indexing

Nearly everyone would agree that an index in a book is useful primarily because it expedites the location of information. Indexes for a materials database serve the same purpose, and all modern database management systems include facilities for constructing them.

Database indexes are usually composed of a key value and a pointer back to the record in the database that contains that key value. Figure 2.3 represents a scaled down example of what an index might look like for a manufacturer of a material found in a database of silicon nitrides and silicon carbides.

To the left of the figure are three sample database records. Each record has a unique identifier, the record key, to distinguish it from all other records. In the example, integers are used but any unique identifier is usually permissible. Often, if there are no unique candidates, fields (perhaps Manufacturer and Designation in this example) will be concatenated for use as the record key.

To the right are two Manufacturer index records. The key for the index record is the manufacturer's name and the pointer is the key of the database record that contains that name in its manufacturer field. As is clear from the fist index record, multiple pointers are possible because multiple database records contain that particular manufacturer's name.

This example touches very lightly on the myriad possibilities for customizing database and index records. Quite frequently, memory addresses are used instead of pointers because records can be accessed more efficiently if the database management system does not need to perform any calculations to determine the locations of records. This special topic and many others may be found in the references provided in the introduction to this chapter.

One needs to be judicious, however, and not create inappropriate indexes because they consume precious disk space and may not add any significant speed improvements for locating information. For example, it is generally not a good idea to index a field that has little variation in its content. In these circumstances, it is nearly as quick to sequentially examine the database records as it is to access the index, read a long list of pointers, and collect the relevant records. There is too much overhead associated with the use of an index of this type, which includes consumption of disk space devoted to the maintenance of that index. Also, it is not wise to index fields that will be searched infrequently, if at all. Disk space will simply be wasted on such indexes.

In summary, indexes are extremely valuable for finding information fast, but selecting appropriate database fields to index requires thought and insight into how the user actually wants to access information.

The Brief Display

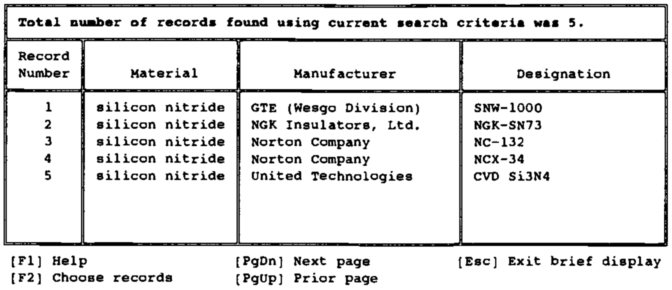
Searching a database of even moderate size often identifies many records satisfying the search criteria, but the user will probably be interested in only a select subset of those retrieved. As an aid for establishing the subset, a list of the material specifications along with several other fields found in the records is displayed. The user may choose those records of interest in the list for which complete information is subsequently provided. The list is called a brief display because it does not contain all the information in a record but rather enough to whet the user's appetite (Fig. 2.4). It should be noted that the fields to be included in the brief display are usually specified by the user during the demonstration stage of development.

The Full Display

Presentation of the complete record is called the full display (Fig. 2.5). The software engineer must exercise caution in designing the full display taking care to compartmentalize similar information and to highlight important fields such as cautions or special notes concerning the material. It is very possible that the full display of a record will require multiple screens, and the transition from screen to screen and record to record should be implemented clearly and consistently.

Downloading

Users will want to place information retrieved from the database in reports, papers, or other applications software, and a downloading capability simplifies this use of the database. Options should allow downloading to a file and, possibly, to output devices such as printers and plotters. File formats typically include ASCII but others, such as DIF (data interchange format), may be required depending upon user needs. Windows-based software will provide a clipboard or dynamic data exchange (DDE).



Total number of records found using current search criteria was 5.

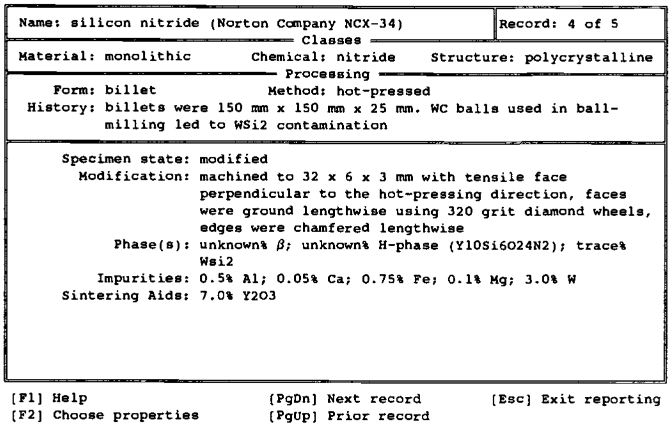
Record Number	Material	Manufacturer	Designation
1	silicon nitride	GTE (Wesgo Division)	SNW-1000
2	silicon nitride	NGK Insulators, Ltd.	NGK-SN73
3	silicon nitride	Norton Company	NC-132
4	silicon nitride	Norton Company	NCX-34
5	silicon nitride	United Technologies	CVD Si3N4

FIG. 2.4—An example of a brief display. The user presses [F2] to select records for full display.

Data Facilities for Accuracy, Significant Figures, and Indicating Quality

Numeric data for materials databases, as in all scientific databases, vary widely in precision, which is tied to the precision of the test method, and accuracy [*10*]. Measurements may be made to many significant figures or simply reported as a range of values, but, whenever possible, it is important to store and display the appropriate number of significant figures. Searching for a special real value with a certain number of significant figures, however, is a difficult if not impossible task. Searches involving such numbers are conducted more effectively if the user is given the opportunity to specify a range of values. It is customary to provide capabilities for selecting a minimum and/or maximum value along with relational operators, such as < (less than), <= (less than or equal to), > (greater than), and >= (greater than or equal to), which allow flexibility in searching numeric values. Figure 2.6 shows how this was implemented for one particular database.

Since materials databases include numeric data from a wide variety of sources, it is important to include an indication of the quality of the data. The quality indicator should be supported by a discussion or short commentary on the precision and resolution of the instrument used to generate



```
Name: silicon nitride (Norton Company NCX-34)        Record: 4 of 5
                              Classes
Material: monolithic     Chemical: nitride   Structure: polycrystalline
                            Processing
    Form: billet                 Method: hot-pressed
 History: billets were 150 mm x 150 mm x 25 mm. WC balls used in ball-
          milling led to WSi2 contamination

     Specimen state: modified
       Modification: machined to 32 x 6 x 3 mm with tensile face
                     perpendicular to the hot-pressing direction, faces
                     were ground lengthwise using 320 grit diamond wheels,
                     edges were chamfered lengthwise
           Phase(s): unknown% β; unknown% H-phase (Y10Si6O24N2); trace%
                     Wsi2
         Impurities: 0.5% Al; 0.05% Ca; 0.75% Fe; 0.1% Mg; 3.0% W
     Sintering Aids: 7.0% Y2O3
```

FIG. 2.5—An example of a full display. The user selected this record from the brief display and all information about the material is made available.

the data as well as the specific test method employed. Three recommended quality indicators arose from work done on a large materials database project [*11*]: limited use data, qualified data, and highly qualified data. Suggested standards for each category are summarized in Table 2.3 (see Chapter 6).

The Bells and Whistles

Additional system features that are often requested by users include a graphics capability, units conversion, statistical analysis, and a thesaurus. These features, while not strictly necessary, enable one to make more effective use of the database.

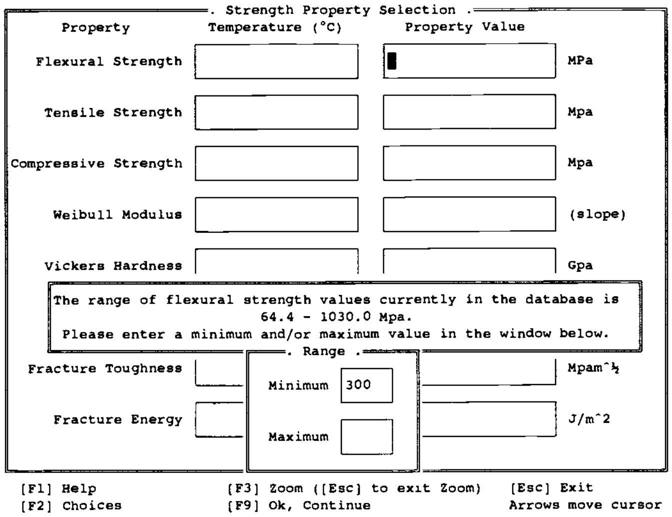
Graphics Facilities for Data Visualization

The human eye and brain more readily perceive periodicity and trends in numeric data when they are presented in graphical rather than tabular fashion. Adding a graphics facility to the database system, however, can be a formidable task depending upon user requirements. When sophisticated capabilities are needed, it is best to explore linking the database to one of the many excellent commercially available graphics packages. An important issue to consider when pursuing this alternative, however, is the licensing requirement of the package, particularly for personal computers. Royalty-free add-ons greatly simplify the distribution of the completed database system.

Another possibility is a library of graphics functions. These libraries are available for a wide variety of higher level languages such as C and FORTRAN. While they do not contain the user interface found in graphics packages, they do offer great flexibility and, in most cases, a very rich collection of graphics functions. Of course, if the project is flush with money, time, and talent, graphics functions may be written from scratch; but, the author's viewpoint is that there are very few reasons to do this. The best graphics libraries have taken many years to develop and have been carefully and extensively tested. The time saved by using these libraries is well spent on developing the most robust materials database system possible.

Units Conversion

Most numerical information has a preferred unit of measurement associated with it, but often there are many possible alternative units. Ideally, storing, searching, and displaying database records will occur using the units of the original measurement, but conversion software may be written to provide users with a choice of units based upon preferences. The software can be embedded in the database system, and most often simply entails a conversion factor or formula. The tradeoff that must be addressed involves when to use alternative units. If the database system is totally flexible, units conversion may occur anywhere, but there may be significant degradation in system response particularly during search and display. It is possible to avoid dynamically converting units by redundantly storing the database using a fixed set of alternative units. The drawback with this approach is the excessive use of disk space.



The range of flexural strength values currently in the database is 64.4 - 1030.0 Mpa.
Please enter a minimum and/or maximum value in the window below.

Minimum 300
Maximum

FIG. 2.6—An example of how range searching was implemented for the Structural Ceramics Database. The user places the cursor at the appropriate prompt, the software states the acceptable range of values, and the user can then request minimum and maximum values.

TABLE 2.3—Suggested data quality standards [*12*].

Limited Use Data
- Data are traceable to an individual, organization, or reference (both the data "Source" and "Digitizer" are identified)
- After independent review, an identifiable authority approved the digitized version for inclusion in the database
- Basis of the data is identified
    - a. experimental measurements
    - b. derived data—specify theoretical basis and data
    - c. estimated data
- Type of data is indicated
    - a. original point values
    - b. analyzed data
        - b1. standard fit—specify fit and data
        - b2. fit unknown

Qualified Data
- Number of measurements and data sets stated
- Nominal confidence limits estimated (that is, 0.90, 0.95, *n*)
- Traceable materials specification assures reproducibility
- Testing methods are specified and conform to a standard
- Data are traceable to a testing/data generating organization or individual

Highly Qualified Data
- High confidence limits determined (that is, 0.99, 0.95, *n*)
- Perform minimum number of individual measurements
    - a. from minimum number of sample lots
    - b. from multiple suppliers (if appropriate)
- Data determined for each variable (that is, form, processing condition, size, and so forth) that significantly affects property
- Independent testing performed (other than the producer and preferably by several testing labs)
- A second, independent evaluation (evaluator identified)
- All features explainable
- Producer(s) identified

Statistical Analysis

Much of the discussion on graphics facilities also applies to statistical analysis. There are many excellent packages available for analyzing data and, if requirements are extensive, it is advisable to consider interfacing the database system with the statistical package. This may be accomplished easily by using the downloading facility to prepare files that may be uploaded to a variety, perhaps the most common, of statistical packages or, with more effort, by writing a seamless interface to the statistical programs.

Minimal requirements, such as summary statistics for the number of values in a data set, minimum, maximum, mean, variance, and standard deviation, may be coded and integrated in a separate module within the database system.

Thesaurus

A thesaurus of terms can be an important part of the online help facility and a requirement for units conversion. It is usually a separate file in the database with records that can be searched and displayed. The thesaurus includes definitions as well as relationships among terms that support the use of the materials database system. For units conversion, a term in the thesaurus will be a unit, and the record for that unit will also contain information about alternate units and the necessary conversion factors and formulas.

Figure 2.7 contains examples of possible thesaurus records. Each record may be composed of the data elements as follows:

TERM	is the key to the thesaurus record.
TERM_TYPE	is the category into which the term fits, such as material, property, unit, or variable.
DESCRIPTION	contains text describing TERM.
BROADER_TERM	is a multiple occurring data element which identifies broader categories into which TERM fits.
STANDARD_TERM	is the standard nomenclature for TERM.
USED_FOR	is another multiple occurring data element which contains commonly used alternative names for TERM.
STANDARD_UNITS	holds the name of the standard measurement units for TERM.
STANDARD_CONVERSION_FACTOR	provides the conversion factor for STANDARD_UNIT which yields TERM.

```
TERM = A92024
TERM_TYPE = material
DESCRIPTION = This wrought aluminum alloy, containing copper, magnesium, and
manganese as hardeners, can be precipitation hardened by heat treatment to
strength levels among the highest of the commercially available aluminum
alloys. Because of it high strength, formability, machinability, and
availability in most product forms, it is widely used in aerospace structures.
Nevertheless, its use is limited somewhat by its relatively inferior corrosion
resistance and weldability as compared with some other aluminum alloys. Sheet,
strip, and plate are available in clad form which is recommended for
applications where corrosion may be a problem. The alloy has useful strength
at temperatures up to about 300°F. With further increases in temperature, its
strength decreases sharply.
BROADER_TERM = aluminum alloys - wrought, heat treatable
BROADER_TERM = aluminum based - wrought
BROADER_TERM = aluminum alloy
STANDARD_TERM = A92024
USER_FOR = 2024
USED_FOR = 24S (obsolete)
USED_FOR = SAE J454 (2024)
USED_FOR = AA 2024

TERM = bulk modulus
TERM_TYPE = property
DESCRIPTION = Ratio of mean normal stress to the change in volume per unit
volume. It is a measure of incompressibility of a material when subjected to
hydrostatic pressure.
BROADER_TERM = compression properties
STANDARD_TERM = bulk modulus
USED_FOR = bulk modulus of elasticity
USED_FOR = bulk modulus
STANDARD_UNITS = pascal

TERM = centimeter
TERM_TYPE = unit
DESCRIPTION = A unit of length equal to 1/100 of a meter or 0.3937 inch.
BROADER_TERM = size units
STANDARD_TERM = centimeter
USED_FOR = cm
STANDARD_UNITS = m
STANDARD_CONVERSION_FACTOR = 0.01

TERM = cold reduction
TERM_TYPE = variable
DESCRIPTION = Percent cold reduction of a material, usually during a step in
the fabrication process
STANDARD_TERM = cold reduction
USED_FOR = cold reduction
STANDARD_UNITS = %
```

FIG. 2.7—Examples of possible thesaurus records, which might be used with a materials property database [*13*].

GETTING THE SOFTWARE TO WORK RIGHT

Alpha Testing and Debugging

Alpha testing is the performance testing that is conducted by the developer or testing group prior to the public release of the software. For large projects, alpha testing is performed by a professional testing group as code is being developed. This reduces the development cycle and gets the product to market faster than waiting for the software to be completed and then iteratively tested and debugged. For smaller projects, alpha testing is often done by the individuals who wrote the software. While this is an undesirable circumstance because it is hard for the software engineers to be objective about their own code, it is, nonetheless, reality.

The first stage of alpha testing is to develop a plan that includes what will be tested, how testing will proceed, and who will do the testing. Commit the plan to paper or computer file so that nothing will be forgotten, and also remember that the plan must be flexible enough to accommodate changes.

Determining what to test can be particularly difficult when the software contains many options and features. The best approach is to tabulate the options and features along with the test cases that will be applied to each. Test cases should include normal cases, boundary conditions, and invalid conditions. Normal cases are the obvious things a user might do: pick a choice from a menu, enter valid values to a prompt, perform a search, produce a report. If the obvious cases do not work, testing should discontinue until the software has been debugged.

Boundary conditions test the limits of the software options and features. For example, what will the software do if the user attempts to download a record to a file for which insufficient disk space is available?

Invalid conditions are those that are just plain wrong. For instance, what happens when the user enters an invalid value for a prompt? Does the software crash or the computer hang?

Determining how to test the code involves choosing between black-box and white-box (or glass-box) testing. Black-box testing means that one is only concerned with whether the software does what it is supposed to do and not with how it actually does it. If written specifications have been prepared for the database, they serve as a statement of what the software should do and testing may proceed from the specifications.

White-box testing is essentially a code review. The tester reads through the code while applying tests for the normal cases, boundary conditions, and invalid cases. This manner of testing can be extremely valuable because careful reading of the code will unveil the precise location of errors and can suggest additional test cases. However, it is possible to become so entrenched in the code that the tester begins to view the logic of the solution in the same way as the developer and, therefore, the same errors remain undetected.

The software engineer should not have primary responsibility for testing the code unless no one else is available. The developer is not likely to be highly objective or likely to test conditions that were not explicitly coded if they were not part of the functional specifications. Choose an individual, preferably with prior testing experience, who will execute a test plan and accurately report the results.

The 1%, 10%, and 100% Loads

If the materials database is of moderate to very large size (roughly 1000 to greater than 1 000 000 records), consider testing at three levels of capacity: 1% of the data loaded, 10% loaded, and 100% loaded. Testing in this manner will more readily help expose performance problems, such as slow searches, sorts, and reports, since reference points will be developed in reasonable increments. The records for the 1% and 10% loads may be selected at random from the full data set.

GETTING THE SOFTWARE TO WORK BETTER

Beta Testing and Debugging

After alpha testing is complete, that is, the developers and testers feel that the database software is robust enough to be placed in end user hands, beta testing may begin. Select beta testers from the pool of people who might use the software and recognize from the outset that these will not be professional testers and should not be expected to rigorously review the software according to a pre-designed plan. They will use it in intended and unintended ways and will invariably find bugs. When the problems have been corrected, the database is ready for the finishing touches in preparation for production release.

Fine Tuning

Software engineers must carefully weigh the advantages and disadvantages of fine tuning software that, in most cases, works well and works fast. In a database application, attention certainly must be focused on potential bottlenecks, such as search, retrieval, sort, and display times, but fine tuning every module of code is not only unnecessary, it can create problems. The software can easily become difficult to read, maintain, and extend. Let common sense be your guide.

MOVING FROM PROTOTYPE TO PRODUCTION

The Professional Touch and Software Distribution

The emphasis to this point has been on the infrastructure of a materials database system. But, no matter how well constructed the database may be, if "little" things are neglected, the user may view the system in a somewhat less favorable light and, worst of all, simply not use it. This section will discuss several items that will improve the overall quality of the finished product.

User Documentation

It should be every software engineer's goal to create a system that can be used without any documentation. Nevertheless, one must be realistic and recognize that all users differ in their approaches to learning about a software application, in this case a materials database. Good user documentation, therefore, is an essential part of the total database package.

Part of the documentation, online help, has already been discussed. The printed user manual should include clear software installation instructions if needed, a guide with explanations of every feature in the database system, and a sample session that the user can run using the database, which will simultaneously provide an assurance that the software was installed correctly, and a tutorial on how to use the system.

The most important stylistic issue regarding the user manual is to minimize jargon and, where its use is unavoidable, provide examples with explanations.

Automated Installation

Installation of the materials database, particularly if it is to be placed on a personal computer or workstation, is essentially the software engineer's salutation to the user; and a crisp, professional installation process announces that the software was developed with care and with an emphasis on quality. First impressions should not be underestimated since they directly influence the user's overall perception of the software product.

The installation process should be straightforward and as accommodating as possible with clear instructions available for every step. It must be thoroughly tested and debugged. It is also an excellent idea to test the process on individuals who have little or no experience installing software, prior to public release of the database system. Weaknesses in the installation software and instructions will be exposed quickly.

Runtime Packages and License Agreements

Runtime packages and licensing agreements protect the rights of software engineers and distributors. Runtime packages are trimmed down versions of development packages. They will run applications that were built using a particular package but do not contain the full set of development tools.

For nondistributed systems, each user usually receives a runtime copy of the database with a unique license number assigned to it. Sometimes it is possible to negotiate a site license for the runtime package if demand is sufficiently high. One way to totally eliminate the overhead is to write the database software using a language or package that does not require royalty payments.

Technical Support for the Software

Support should be available whenever users encounter technical problems with the database. There are many ways to provide this support via telephone, mail, electronic mail, and fax machines, but the approach should be consistent no matter which route the user takes. Technical support should be courteous, punctual, and correct.

Courtesy is usually the first attribute to be dismissed when an angry user requests support, but it is surprising how often good ideas and comments arise from such encounters when one reins in the temptation to simply respond to the user in kind.

Punctual response to a request for assistance is another indicator of the care that has gone into developing the product. This assures the user that the support request is taken seriously.

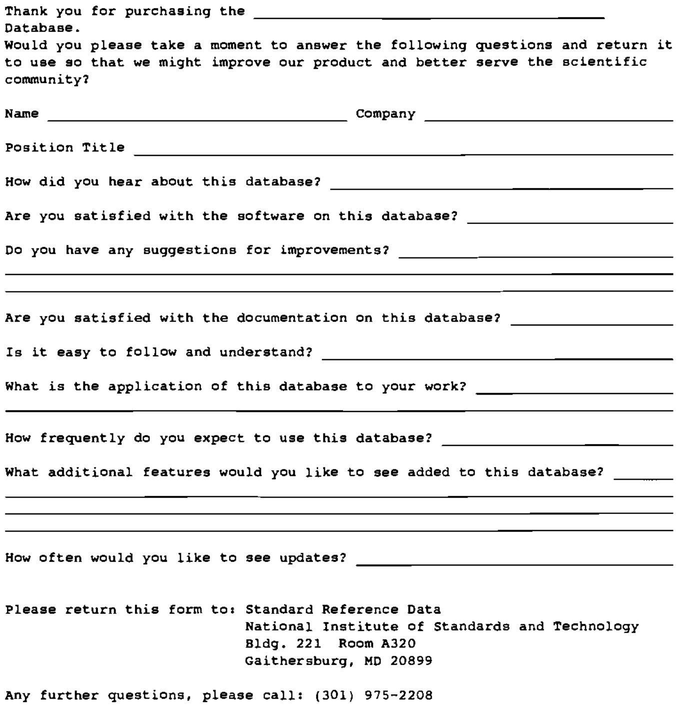
Failing to provide correct technical responses to users' questions undermines their confidence in the database system and may lead them to discard it entirely. If a correct response cannot be made promptly, explain this to the user and give a conservative but reasonable estimate of the time it will take to provide the necessary help.

It is best to place technical support contact information in the printed user manual for the software.

Technical Support for the Data

Separate support should be provided for the data since it is likely that the software engineer does not also have in-depth knowledge of the information in the database. The same rules of support—courtesy, punctuality, and correctness—apply, however, and should not be ignored.

Data support contact information should also appear in a convenient location in the printed manual.



Thank you for purchasing the ____________________ Database.
Would you please take a moment to answer the following questions and return it to use so that we might improve our product and better serve the scientific community?

Name ____________________ Company ____________________

Position Title ____________________

How did you hear about this database? ____________________

Are you satisfied with the software on this database? ____________________

Do you have any suggestions for improvements? ____________________

Are you satisfied with the documentation on this database? ____________________

Is it easy to follow and understand? ____________________

What is the application of this database to your work? ____________________

How frequently do you expect to use this database? ____________________

What additional features would you like to see added to this database? ____________________

How often would you like to see updates? ____________________

Please return this form to: Standard Reference Data
National Institute of Standards and Technology
Bldg. 221 Room A320
Gaithersburg, MD 20899

Any further questions, please call: (301) 975-2208

FIG. 2.8—An example of a feedback questionnaire used by the Standard Reference Data program at NIST.

Announcing Future Releases of Software

Announcing new releases of the software requires some organization with regard to mailing lists and user feedback. It is important to learn who purchased the software and why. An effective method of gathering this information is by distributing a user feedback form with the software package. Figure 2.8 is a feedback questionnaire and used by the Standard Reference Data program at NIST. It covers all the basics such as user identification, how the database is being used, what problems exist, and which features are missing. In general, approximately 10% of the questionnaires will be completed and returned, but even this small number will provide valuable insight into how the user community perceives the product and what changes should be incorporated into future versions. As a consequence, announcements for new releases will reflect responsiveness to user needs.

MAINTENANCE

Database developers must be prepared to maintain their software following public release of the system. For large-scale projects, maintenance time can eventually exceed development time, and project managers must be prepared to allocate the necessary resources for this activity. Data and code may be added, removed, or altered, which means that the latest version of the database will need to be tested and debugged before being released to the public. The project manager will need to establish a timetable for new releases of the database.

Demonstrating the System

The topic of demonstrating the database system is not strictly germane to program infrastructure, but the author considers it extremely important. He has witnessed too many projects where the developers committed the egregious error of not thoroughly preparing the demonstrations. This careless attitude not only insults the audience, it can scuttle the entire project or significantly change its management and direction.

The guiding principle is to never demonstrate anything that has not been previously tried using the current version of the system. Prepare a demonstration that thoroughly covers all the highlights of the database but politely decline invitations from the audience to demonstrate features that had not been explored before the demonstration. Honesty and humility are more graciously received than humiliation.

GLOSSARY OF DATABASE JARGON

Compound Key—A key field comprised of any combination of fields from a record.

Database—An organized collection of related files.

Data Element—see **Field**.

Field—Also known as a data field or data element; it represents the basic unit of information storage in a database and is always defined to be an element of a record. A field has attributed associated with it, such as name, type (for example, character, numeric, date), and length.

File—The primary physical storage unit into which a database is organized. Database records are stored in files.

Index—A set of key values, similar to the index of a book, which enable rapid retrieval of records from a database.

Pointer—An address or a key value in an index record, which provides the information necessary for locating a record in the database.

Record—A collection of related data fields.

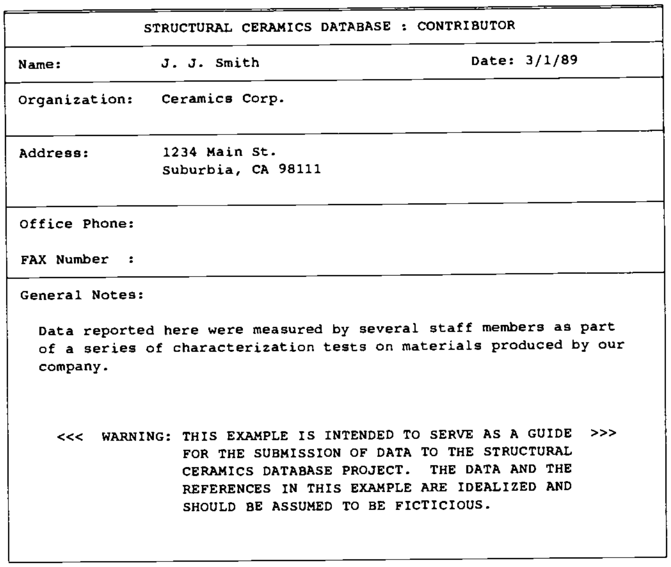
Redundant Data—Identical data that is stored in multiple locations in a database.

Schema—A conceptual model of the structure of the database, which defines the data contents and relationships.

Virtual Memory—An input/output management technique that keeps the most recently and most often accessed information in memory during execution of a database application program. It reduces the amount of required actual disk I/O, resulting in improved performance.

APPENDIX

Figures 2.9 through 2.13 represent an example of the data acquisition format developed for the Structural Ceramics Database project. Included are completed forms for contributor and bibliographic information, material specification, measurement method, and property measurements.



STRUCTURAL CERAMICS DATABASE : CONTRIBUTOR

Name: J. J. Smith Date: 3/1/89

Organization: Ceramics Corp.

Address: 1234 Main St.
Suburbia, CA 98111

Office Phone:

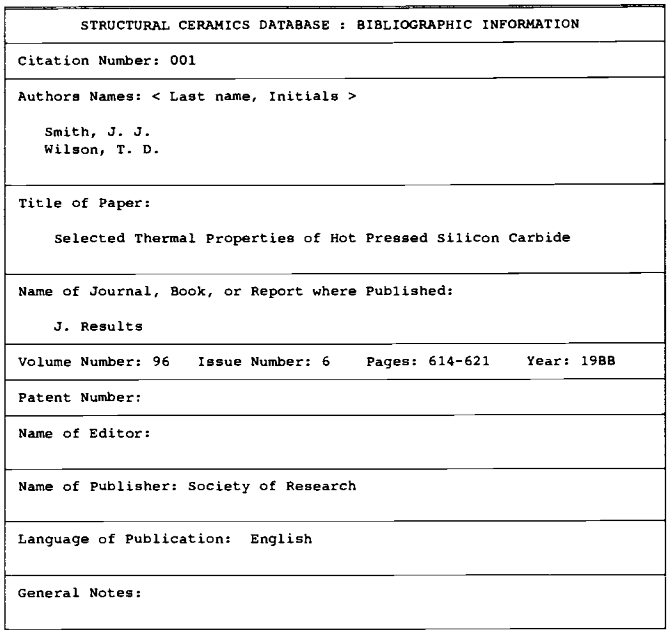
FAX Number :

General Notes:

Data reported here were measured by several staff members as part of a series of characterization tests on materials produced by our company.

<<< WARNING: THIS EXAMPLE IS INTENDED TO SERVE AS A GUIDE >>> FOR THE SUBMISSION OF DATA TO THE STRUCTURAL CERAMICS DATABASE PROJECT. THE DATA AND THE REFERENCES IN THIS EXAMPLE ARE IDEALIZED AND SHOULD BE ASSUMED TO BE FICTICIOUS.

FIG. 2.9—The first page of the SCD data format documents the name and contact information of the contributor.



STRUCTURAL CERAMICS DATABASE : BIBLIOGRAPHIC INFORMATION

Citation Number: 001

Authors Names: < Last name, Initials >

Smith, J. J.
Wilson, T. D.

Title of Paper:

Selected Thermal Properties of Hot Pressed Silicon Carbide

Name of Journal, Book, or Report where Published:

J. Results

Volume Number: 96 Issue Number: 6 Pages: 614-621 Year: 1988

Patent Number:

Name of Editor:

Name of Publisher: Society of Research

Language of Publication: English

General Notes:

FIG. 2.10—Published sources of data are documented as bibliographic information in the SCD data format.

STRUCTURAL CERAMICS DATABASE : MATERIAL SPECIFICATION	
Material Identification Number: SC-123-A	Citation Number: 001
Material Class	monolithic ceramic
Structure Class	polycrystalline
Chemical Class	carbide
Generic Name	silicon carbide
Formula	SiC
Manufacturer Material Name Lot Number Product Date	Company Name, Inc. SCX-123 N123456 9/1/88
Fabrication Process Form Notes	hot pressed billet 5 hrs. at 2100 °C and 150 MPa
Primary Composition Wt. Percent Standard Dev.	Si; C 68.5; 29.5 0.5; 0.5
Material Phases Wt. Percent Standard Dev.	alpha; beta 96; 2 1.0; 0.5
Sintering Aids Wt. Percent Standard Dev.	Al2O3 2.0 0.5
Other Dopants Wt. Percent Standard Dev.	None
Impurities Wt. Percent Standard Dev.	Al; Ca; Cl; F; Fe 0.03; 0.008; 0.01; 0.1; 0.007 0.01; 0.001; 0.002; 0.02; 0.001
Density Unit Percent Theoretic Measurement Method	3.10 +/- 0.03 g/cm^3 96.5 ASTM C373-72(1977)
Porosity (%)	0.2 +/- 0.05
Grain Size Unit	0.8 +/- 0.3 10^(-6) m
Application Notes	heat exchangers

FIG. 2.11—General and detailed data are used to specify monolithic ceramics.

STRUCTURAL CERAMICS DATABASE : MEASUREMENT METHOD	
Measured Property: specific heat	Citation Number: 001
Method Name: differential scanning calorimetry (DSC)	
Measurement Environment: argon	
Specimen Preparation: A sample of approximately 1 cm in length was cut from a rod of diameter 0.625 cm and then crushed into a powder.	
Measurement Procedure: A specimen of 0.1 g of powder was placed in a gold pan. A reference material, alpha alumina powder, was placed on a separate gold pan in the DSC. The two materials were heated simultaneously at a rate of 10 °C/min from 25 °C to 500 °C.	
Other Measurement Notes: A detailed discussion of the measurement procedure is given in "DSC Measurement of Specific Heat" by J. J. Smith and K. L. Jones, J. Results 94, 134-142 (1986).	

FIG. 2.12—Nonstandard methods need detailed preparation and procedural information and preferably a reference.

STRUCTURAL CERAMICS DATABASE : PROPERTY DATA			
Material Identification Number: SC-123-A		Citation Number: 001	
Property Name: specific heat			
Measurement Method Name: DSC (≤ 500 °C) and Drop Calorimetry (> 500 °C)			
Cautions and Special Observations: Accuracy decreases with increasing temperature. Temperatures are mid-point values. Property values are given with a 95% confidence range based on five independent measurements at each temperature.			
Variable:	Temperature	Property	
Unit:	°C	J g^-1 K^-1	
Values:	25 +/- 5	0.670 +/- 0.001	
	100 +/- 5	0.840 +/- 0.002	
	500 +/- 5	1.12 +/- 0.02	
	1000 +/- 50	1.26 +/- 0.06	
	1500 +/- 50	1.36 +/- 0.07	

FIG. 2.13—Precision and accuracy statements are expected entries in the SCD data format.

REFERENCES

[*1*] Date, C. J., *An Introduction to Database Systems*, Vols. 1 & 2, 4th ed., Addision-Wesley Publishing Co., Reading, MA, 1987.

[*2*] Martin, D., *Advanced Database Techniques*, The MIT Press, Cambridge, MA, 1986.

[*3*] Ullman, J. D., *Principles of Database and Knowledge-Base Systems*, Computer Science Press, Rockville, MD, 1988.

[*4*] Wiederhold, G., *Database Design*, McGraw-Hill Book Company, New York, NY, 1983.

[*5*] Atre, S., *Database: Structured Techniques for Design, Performance, and Management*, 2nd ed., John Wiley & Sons, New York, NY, p. 59, 1988.

[*6*] Munro, R. G. and Begley, E. F., "The Structural Ceramics Database: Data Acquisition Format for Monolithic Ceramics," *Journal of Testing and Evaluation*, Vol. 19, No. 3, 1991, pp. 250–255.

[*7*] Begley, E. F. and Munro, R. G., "The Structural Ceramics Database: Topical Report," NISTIR 4601, The National Institute of Standards and Technology, Gaithersburg, MD, 1991.

[*8*] Galitz, W. O., *Handbook of Screen Format Design*, QED Information Sciences, Inc., Wellesley, MA, 1985, p. 21.

[*9*] Rumble, J. R. and Smith, F. J., *Database Systems in Science and Engineering*, IOP Publishing Ltd., Bristol, England, 1990, p. 70.

[*10*] Rumble, J. R. and Smith, F. J., *Database Systems in Science and Engineering*, IOP Publishing Ltd., Bristol, England, 1990, pp. 14–17.

[*11*] Grattidge, W., Westbrook, J., McCarthy, J., Northrup, C., Rumble, J., "Materials Information for Science & Technology (MIST): Project Overview," NBSSP 726, U.S. Government Printing Office, Washington, DC, 1986.

[*12*] Grattidge, W., Westbrook, J., McCarthy, J., Northrup, C., Rumble, J., *Database Systems in Science and Engineering*, IOP Publishing Ltd., Bristol, England, 1990, pp. 32–33.

[*13*] Kaufman, J. G., MPD Network, Thesaurus, private communications, 1988.

3

Types of Materials Databases

John R. Rumble, Jr.[1]

INTRODUCTION

The purpose of this chapter is to define different types of materials databases using several classification schemes. The schemes include: material data type, user type, application type, and access type.

The range of materials information is wide and the process of providing computer access must reflect this diversity. The use of materials data is more than going into the laboratory, performing a measurement, and using that test result in a design. Materials information is more like the flow of a slow moving river in which individual test measurements are collected together and over time aggregated into commonly accepted "property" values that are found in handbooks and design manuals.

One goal of this chapter is to define the flow of materials information from its generation to its use and to demonstrate how this flow affects computerization. The flow consists of four primary stages: generation, analysis, aggregation, and analysis. The most obvious effect of this flow is the difference among publications associated with each stage. As we computerize materials information, each stage will produce databases that reflect the specific nature of that particular stage. Databases with raw test results will look very different from application databases.

A second type of classification of materials databases relates to the user community. A database may be intended for a single user, a group, an institution, or the public. Each class of users will present different demands on the database project.

Another important consideration in classifying materials databases relates to their use in specific applications. Today almost every engineering and scientific activity is computerized, and the use of materials data as found in databases is not confined to simple search and retrieval. Computerized material data resources must be integrated with other engineering software such as expert systems, finite-element analysis, process control, design, and product maintenance. An understanding of the different types of materials databases needed to support this software will allow computerization to proceed and be integrated more effectively.

Materials databases can be classified with respect to their access, whether as a stand-alone personal computer database, as part of an online system, or as integrated into an engineering workstation. The access characteristics affect the database building process not only from the viewpoint of hardware and software, but also with respect to the integration of a materials database with other materials databases.

Thus this chapter will examine the types of materials databases in four ways: (1) the type of materials data included, (2) the community of users, (3) the application of the database in engineering and scientific work, and (4) the different types of access and dissemination. This will be done to provide the database builder with an understanding of the different types of materials databases and what the impact of the type chosen will be on the final database [*1*].

There are two basic types of databases concerning materials: (1) those that deal primarily with the description of materials and (2) those that contain information on the properties and performance of materials. While these databases are usually linked together in some way, this does not always occur. The description of materials in databases is covered in detail in Chapter 4 and will not be discussed further in this chapter. Note that the considerations in the later sections of this chapter apply to both basic types of databases.

CLASSIFICATION BY TYPE OF MATERIAL DATA

Data on the properties and performance of materials is dynamic information, initially generated over a period of time and changing and improving as time passes. This happens both because a material becomes better defined in terms of composition or processing and because test methods are improved or extended. Most materials properties used in engineering are not intrinsic properties but are dependent on the test method. These materials properties generally result from a specific test procedure that has been designed to capture some aspect of material service and can be used to predict performance.

As a material is subjected to more and more testing and as test results are analyzed more completely, individual test results are aggregated together into commonly accepted val-

---

[1] Program Manager, Standard Reference Data, National Institute of Standards and Technology (NIST), A323 Physics Building, Gaithersburg, MD 20899.

Copyright© 1993 by ASTM International www.astm.org

TABLE 3.1—Types of technical databases.

Source of Data	Type of Database
Measurement	Laboratory Notebook
Analysis	Report
Aggregation	Handbooks
Application	Application

ues. The flow of material property data, while fuzzy at the boundaries, is readily identifiable in the bulk. Four distinct stages exist: test generation, analysis, aggregation into property data, and use in application. The databases that are generated in each stage are quite distinct and have features quite different from databases from other stages (Table 3.1). The ideas incorporated in the flow model have been previously discussed by Bullock et al. [*2*], Mindlin and Smith [*3*], Westbrook [*4*], and Rumble et al. [*5*].

Laboratory Notebook Databases

The first stage for material property data is their generation, usually by testing or experiment. Presently, almost all materials test equipment is computerized so that all data collected is already computerized. Before computerized test equipment, these data were collected on printed forms that required considerable diligence and patience to fill out completely for each test. Now that the information is collected by computer, most test results come with a much more complete test record. Round-robin testing to establish the validity of a test method or to compare test machines can often find the reason for differences more easily since more independent factors can be correlated.

The primary function of collecting and storing test data is the preservation of the measurement results. This might seem obvious; however, some measurements are made simply to support instantaneous decisions. Examples are determining the temperature of a solution to see if the next processing stage can begin or measuring the hardness of a steel to see if a shipment meets its specifications. Many experiments or tests yield data worth saving. It is often costly to reproduce these data and sometimes impossible to do so. In the past, experimental data were collected and preserved in laboratory notebooks. In the last 30 years, computers have become prevalent in almost every type of technical and test experiment and, in effect, have replaced laboratory notebooks. Computers can respond faster than a human observer and are more accurate; they are also capable of monitoring tests for long time periods. They certainly can collect and process large amounts of data more easily.

The computerized collections of test results data are called "Laboratory Notebook Databases." Although most testers and researchers do not consider them to be databases, these data collections are usually treated in the manner of databases, that is, they are searched, analyzed, edited, updated, manipulated, and displayed. The primary features of laboratory notebook databases are the amount of data contained and their completeness. Test results include the following:

(1) information on the material from which the test specimen was taken,
(2) details of the test specimen,
(3) test parameters that were established initially and not varied,
(4) test parameters that vary and are monitored throughout the test,
(5) details of the material behavior and appearance during the test,
(6) the raw test results,
(7) information of the validity of the test results, and
(8) information on the analysis of the raw test results, such as conversion of a voltage change to a strain.

The key to successful use of test results data is the existence of supporting data that must be kept to make further use of the results possible. The more expensive or unique the test, the more important these data become. Such ancillary data are often well-defined, either from long practice or by standards. From the earliest stages, researchers and research managers must be mindful that data collection is for both the immediate purpose as well as long-term use. It is more difficult to make notes on a computer file than it is to write in a laboratory notebook. Either a database must have all data items or data fields for likely information that might be included or have a large free-text comments section. Free-form comments are difficult to manipulate and a well-thought-out structure is preferable. The use of a laboratory notebook database to replace a written laboratory notebook as a legal document, say for patent purposes, is just now being explored.

The successful use of computers to collect test data requires two kinds of computer-related tools that are now available in laboratories: database management systems (DBMS) and data recording standards. The DBMS must have the features and capabilities necessary for the collecting of test data, as discussed in Chapter 2. Data standards refer to recording of test data as discussed in Chapter 5.

In summary, laboratory notebook databases contain original test results data. If the measurement results are used solely by the generating group, only abbreviated or brief additional information needs to be stored. If results will be made available on a wider basis in the future or if the original user will return to the database after a lengthy period of time, more complete information must be added.

Report Databases

Once a set of material test results are gathered, they are analyzed and made available. Analysis might be a simple graphical comparison to determine suspect data points or might be sophisticated correlation analysis. Most types of materials test methods have an established analysis procedure, usually computerized, either as part of a test instrument or as a separate software package.

In the analysis, often much of the individual test detail is not included, especially once the result of the analysis is written up in a report or a paper. Many analysis procedures produce a set of coefficients or parameters that essentially fit the test results to one or more independent variables, and the scatter among the test data may be lost. Analysis may also include determination of the precision and bias associated either with individual test results or with a set of results.

Databases associated with the data analysis process are

TABLE 3.2—Functions of technical data analysis and reporting.

Derivation of properties
Increased usability
Improved understanding
Extension of data domain
Quality assurance
Uniformity
Presentation of information

called "Report Databases" and contain the analyzed results. Today most of this data analysis is done on computers, and the use of databases to help the process and store the results is a logical consequence.

There are several purposes for the data analysis and reporting process as shown in Table 3.2. In each of these cases, the test data are examined in different ways, depending on the discipline, and the results are published as journal articles, reports, and other technical literature. The derived data are the form primarily included in handbooks and used in applications.

At the present time, report databases are just beginning to be built because they generally do not contain many data. A typical published report of analyzed data contains only a few data tables and graphs, and the typical report database contains only a few analyzed data sets. Generally speaking, when analyzed data are aggregated into handbooks, as discussed in the next section, materials databases then become large enough to be worth distributing.

The importance of databases in the data analysis stage cannot be minimized, however, because of the need for preserving analysis results as well as documenting the analysis technique itself. In many cases, the test measurements are converted into derived data, for example, by statistical analysis, especially to determine the influence of certain variables. In some instances, the number of measurements of an easily changed independent variable, for example, temperature, is large, and the amount of data is also very large. Data reduction techniques, such as curve fitting, can reflect detailed variations in the experimental data while providing compact expressions for further use.

Mechanisms are needed for preserving these small materials databases that should contain both the derived data and the original test results for future data aggregation. Presently data analysis results are often published on paper even though the data were initially collected on a computer, all analysis was done on a computer, and all the tables and graphs included in the publications were made on the computer. Users often need the published data in computerized form for use in simulation or modeling software and must retype the data and verify the accuracy. Such a process is of course subject to errors, inefficiency, and incompleteness and can be improved by creating and disseminating report databases.

In summary, report databases contain the results of an analysis procedure applied to a set of test data. Only in some cases are the original data included. The supporting information should include enough details on the analysis procedure that users can determine if it is acceptable or correct. Typical data resulting from analysis are properties or statistically analyzed best values.

Handbook Databases

After test data have been collected, analyzed, and reported, they are used for a variety of applications. Over the years, scientists and engineers have found that retrieving published data is a difficult process. The data in the original literature are often not easy to use. Similar data are reported in a variety of units. Different test methods may have been used to measure the same properties with varying results. Materials data are published in many sources, often difficult to find.

Consequently, data are often aggregated together with other results into collections, usually in the form of a handbook or data evaluation compilations. Aggregated data sources have become the most important source of materials data in many situations, thereby greatly reducing the cost and improving the efficiency of data accessibility. A typical example of an aggregated data source is a handbook. These contain a wide range of data for large numbers of materials, and they come in a variety of formats with many tables and graphs. Usually handbook data have been evaluated to some degree, in many instances, with individual editors responsible for small sections.

An important type of materials database, called a "Handbook Database," corresponds to these data compilations. These generally, but not always, contain a wide range of data for a number of materials and represent a compilation and selection of available analyzed data.

Many data users are not experts in data measurement and are not particularly adept at determining the quality of data from original sources. In data compilations that are unevaluated or for which the evaluation process is not documented, determining data quality is even more difficult because experimental details have been left out. Therefore, evaluated data compilations have become very important to users.

Data evaluation represents the efforts of neutral critical evaluators who assess the quality of a given set of data regardless of the origin. Evaluation is usually based on three criteria: (1) documentation of the test method, (2) comparison to known physical and empirical laws, and (3) comparison to other measurements of the same quantity.

Handbooks and other data compilations are usually the data source of first choice. Unfortunately, few handbook databases now exist, and it is a void that is keenly felt. Not all handbook databases contain evaluated data. For example, material producing companies are now producing databases that primarily contain information on their materials. The data might be wide in coverage though simply a collection of test measurements made by the company. These materials databases are similar to published data sheets and often provide important summary information to materials selectors, for example, in computer-aided design. However, the data included in these databases reflect the needs of the producing company and not of the consuming scientific public. Data can be issued primarily to build sales or make products look attractive. Users must be aware of the purpose of a database.

Application Databases

The work of materials engineers and scientists involves problem solving. Rapid access to pertinent materials information is a major factor in determining solutions in a timely fashion. A data collection targeted to one specific application area containing relevant data from a wide range of sources is called an "Applications Database." When materials information is completely computerized, applications databases will be of primary importance.

Specific data collections for individual applications are created (1) for convenience because the work they do often requires data from a wide variety of different sources and (2) for quality because the data have already been tailored to meet their needs. Data in these specialized collections are much changed from the original measurements and represent the highest refinement that technical data undergo.

Applications databases are now being built for materials applications, directly comparable to the specialized publications mentioned above. The decisions taken in building such databases are entirely dependent on the application. The user interface and the search and display strategies are optimized to the application.

In other circumstances, applications databases will be put together for a particular problem using different data sources, both published and computerized, along with new data directly related to that problem. For example, a materials specialist might have to determine the structural integrity of bolts and other fasteners within a nuclear power plant. Data must be taken from the design, plant performance, and materials databases, then combined perhaps with additional test results.

These working databases often take on a value and life of their own in the sense that they remain after a project is completed. The more intense the work and the longer it takes, the more likely such a database will become important. When its value is recognized, steps must be taken to preserve it and expand its use. Often this would require additional resources or time that cannot be easily justified. Critical decisions must then be made regarding the future of the database. An analogous situation exists with respect to paper data collections. Often these are put in files never to be used again because of the high cost of cleaning up the data or adding full documentation. The same applies to databases if they are only archived in a tape library.

One of the most expensive and difficult aspects of building a database is retro-fitting, that is, adding or changing information. Working databases intended for short-term projects should be reviewed at the earliest possible stage for possible preservation or long-term use. Preservation can be achieved, but not without planning and resources. It does not just happen. The decision to change the nature of such a database must be made consciously and with careful planning.

Many materials applications databases, namely those that are becoming successful, are products of well-planned and deliberate efforts to appeal to a given market. The time and effort that have gone into them are considerable, but their developers have made conscious decisions and know their goals. Other application databases that result from wishful thinking and lack adequate support are failures. If resources are not readily available, database builders need seriously to consider stopping the project before wasting time and money.

TABLE 3.3—Classification of databases by types of users.

Personal (one person)
Group
Institutional
Collegial
Public
Archival

CLASSIFICATION BY USER GROUP

A second classification scheme for materials databases is made with respect to their user community (Table 3.3).

A "personal database" is intended only for its creator. Its use may be intensive or sporadic. Because it is aimed at only the builder, many short cuts and abbreviations may be used, depending on the memory or habits of the builder.

A "group database" is used by a group working on the same problem or using the same experimental equipment or computer software. The users may be connected by telecommunications. The contents are characterized by their brevity and the informality of conventions and documentation. Depending on the size and closeness of the group involved, these will still be more formal than for a personal database.

When a materials database becomes an "institutional database," a different level of support is involved, and more formal conventions are needed. Included are databases used by several groups, by a company, or even by a large corporation. At this level, good formal documentation is needed and careful planning and design are important to accommodate multiple needs. However, some conventions are still likely, reflecting common institutional practice. For example, materials might be referred to by trade names only.

A "collegial database" is one used across institutions, by both small and large numbers of people working in a related materials area, but usually on a fairly formal basis. The data contents may use general terminology, thereby avoiding tradename problems or proprietary concerns. Formats for data contributors may be well-defined. Documentation quality can vary, but the larger the community of colleagues, the more extensive it will be.

"Public databases" are those materials databases made available to the public or to a significant portion thereof. Since these databases are often used differently than anticipated or intended, documentation needs to be complete, and if wide usage is intended, the contents should use commonly accepted terminology.

"Archival databases" place primary emphasis on managing and saving materials data for future use. Sometimes a need for these databases arises from the sheer volume of materials data. Data may also be archived because the immediate demand has passed but a future demand is foreseen.

The level of usage is as important as the stage of technical information flow in characterizing a technical database. A database moves from one level of usage to the next only with difficulty if it is done without planning. The mere existence of a group database does not imply that it can be used by an institution as a whole or be distributed to colleagues. References, documentation, and metadata may have to be retrofitted, and this can be one of the most expensive and time-consuming acts related to database building. This is why

TABLE 3.4—Uses of materials information [7].

Calculation of properties
Evaluation of properties
Design engineering
Materials selection
Materials performance
Materials development
Production engineering
Quality assurance
Failure analysis
Product information
Legislation

planning and designing a materials database is so important: to identify whether more widespread use is probable and, if so, to take this into account from the beginning. After consideration, it may be decided that indeed such wider use will not occur or is not worth the expected effort or expense. Planning may take time, but usually only a few days, a meager cost compared to the time spent retro-fitting a database.

MOVING A MATERIALS DATABASE BETWEEN TYPES

Materials databases are built in the normal course of the creation and use of materials data. For each stage as previously described, databases can logically and easily be built but for different reasons, with different characteristics, and for different types of use. Databases arising in one stage may be inappropriate for use in another stage. People needing materials data may find that a materials database created for a different stage or for a different user group may have too little or too much information.

A problem that has not been addressed by materials database builders is how data might flow through the system. Report databases do not now get their data from laboratory notebook databases. Handbook database builders certainly do not extract data from report databases. In the future this will be a major consideration because, as we have indicated, computers have taken over all aspects of technical work.

Readers should now be in a position to classify their materials database efforts with respect to the flow of technical information and the user group and to assess the possibility that the database will cross from one stage to another. Careful planning is needed to make sure that the database will support its intended uses. Search paths, data items included, and output displays all change among types and need to be reviewed carefully.

OVERVIEW OF TYPES OF MATERIALS APPLICATIONS

Information about materials is used in many different ways, and this is reflected in the wide range of materials databases that are possible and have been built (Table 3.4). The databases that are associated with each of these uses contain different amounts and types of supporting information. The database schema and user interface also vary. Details for each application must be worked out in conjunction with the appropriate user community as described in later chapters.

The use of materials databases with the computer software used in different applications has general features. First a brief discussion of data transfer between engineering software is given. Then, integration of materials databases with two types of software, expert systems and numerical modeling, will be briefly discussed. Finally, the use of materials databases by nonexperts will be commented on.

DATA TRANSFER BETWEEN MATERIALS APPLICATIONS

The fundamental nature of a materials database is its use within an engineering activity. One framework for integrating materials databases with other computerized engineering tools is ISO 10303 on Industrial Automation Systems and Integration—Product Data Representation and Exchange, called "STEP." This international standard defines the data used in the life cycle of a manufactured product including materials data. The purpose of STEP is to facilitate transfer of the information generated in any engineering activity related to a manufactured product by means of a neutral format. The standard will be discussed in Chapter 8.

The physical transfer of data through a STEP file format will be accomplished more easily than the interpretation and understanding of materials by a "computerized" engineer. If the primary use of a database will be in an integrated environment, considerable control will be needed for materials-related terminology and the meaning of different material property data. The database developer must work with software developers to ensure that users who access materials databases from other engineering software are able to understand the terminology and translate their needs into the language of the materials databases. If a materials database is intended to support a large number of diverse applications, this will have considerable influence on the database design. Almost every enterprise has evolved a specialized terminology for materials that must fit together with the data transfer standards.

Expert Systems

A discussion of expert systems themselves is beyond the scope of this manual. Generally expert systems are developed using specialized software, usually in the form of expert system shells that have a database capability built directly into their software. Use of the database by the expert system is then automatic and is solely for the support of the expert system. Materials databases built using this capability are directly attached to the expert system.

A second situation is one where an expert system tries to use an existing materials database that has been built without this use in mind. At present, a few expert system software packages can "link" to outside database management systems. Use of these linkages requires considerable programming to make sure the database schema is intelligible to the expert system. If this use is to derive new rules related to materials performance, great care must be taken and an appropriate materials expert needs to be intimately involved.

Numerical Modeling

The situation with regards to numerical modeling is better than that for expert systems because many database management systems support outside calls and easily provide a data value that can be used by modeling software. For a stable database environment, that is, when the same numerical software interrogates the same materials database over and over again, the problem can be solved fairly easily by use of the STEP materials model. For situations where a variety of materials databases are accessed, each with its own schema, full use of the STEP capability will be needed.

Materials Data for the Nonexpert

Engineering materials are complex substances that are chosen to perform well in both expected and unexpected circumstances. Because of the inherent variation in a material, test results also show a variation that is not always meaningful to a nonmaterials expert. These variations lead to safety factors or property multipliers, so unexpected failure is avoided. The suitability of a particular material for a given application is often dependent on a limiting factor that is not immediately obvious. The body of experience for successful materials utilization is slowly being transferred to expert systems, but these systems usually focus on a very narrow application. Ashby [*6*], among others, is developing a set of software tools that provide a nonexpert system alterative for materials selection and utilization, but these tools are just in their infancy. A considerable period of time will likely pass before materials selection can be computerized beyond one application.

The STEP materials model will not solve these problems because it is just a data transfer mechanism. The "knowledge" associated with a materials expert cannot be transferred so easily.

OVERVIEW OF ACCESS METHODS

Databases are built to be used, and their access has an important impact on the entire building process. Access choices must be actively considered at the beginning of the building process because they require different actions. The primary options are as follows:

(1) personal computer and workstation packages,
(2) online systems, and
(3) mainframe packages.

Each has advantages and disadvantages, and the choice depends on the user community, application, and the kinds of computers the users will have.

Personal Computer and Workstation Packages

The widespread availability of personal computers (PCs) has opened tremendous possibilities for materials databases. PCs now allow databases to reach end users and be totally under the user's control. Their self-contained nature makes them attractive; generally the user simply has to put a diskette into a PC, type a few commands, and the database is loaded and ready to use. Occasionally additional software, such as a graphics package, is required. PC databases also allow for appealing user interfaces. Workstations can be viewed as very powerful PCs and, in fact, the distinction has become blurred. Basically the same considerations apply to workstation databases as to PC databases.

PC databases are distributed on a variety of diskettes and for different operating systems. Many combinations confront the database vendor, and usually only a few configurations and diskette types are supported. If the database uses a commercial database management system, suitable licensing agreements must be made. Most DBMS vendors do allow third-party distribution of a run-time version of their product for a small fee. A key consideration is whether the user will be allowed to make changes in the database, such as adding new fields or additional data. If this is the case, some vendors feel that the users are no longer using just a run-time version, but instead are doing their own database management, and the vendors charge the full licensing fee. It is important to settle these issues before work has progressed too far. Of course, home-built databases can be disseminated without such licensing problems.

Online Systems

A materials database can be made widely available by an online system, often through a third-party vendor. Materials database builders must work closely with the vendors to achieve compatibility. A database may need to be modified substantially to make it suitable for the vendor's system, a process that takes time, heavy involvement of the builder, and some cost. Rarely is a database so attractive financially that a vendor does this work free. The usual pattern is for the database builder to waive royalties for a period of time, rather than actually transfer funds. The transformation work can easily take several months for a mid-sized database. Several factors must be considered in adding a database to an online system:

(1) correct interpretation of the data; each data field must be understood and handled accurately,
(2) determination of the equivalence of data and data fields with other databases on the system,
(3) transformation of the physical data structure to a new DBMS, and
(4) display of the data on the online system.

Making a materials database available via an online system has real advantages. Materials users often must turn to several different sources to find all the information needed to solve a problem. An online system offers the possibility that the entire set of databases needed can be found through one access point. The online system usually integrates different databases together with a common terminology and materials equivalency tables so using several databases is the same as using one. An online system also shares the costs, both developmental and operational, over all databases on their service.

Many groups are opting to make materials databases available both online and in PC format to satisfy different user groups.

Mainframe Packages

Often databases are intended for use on mainframe computers, allowing access by numerous users. In this case, magnetic tapes have been the medium of choice to date to load the database onto the mainframe. Magnetic tapes have been well-standardized, and hardware and operating system problems are rare. With the widespread availability of local area networks, uploading databases from a PC is possible, and the database can be sent on floppy disks.

One primary difference between packages for PCs and those for mainframes is the way databases are used and maintained because of the absence of direct control by users on a mainframe. Starting and stopping the software, the handling of errors, as well as file management take on new dimensions. The licensing of DBMS software packages can also be more complicated on a mainframe. Another problem is a lack of a user interface. Because the database will be loaded onto an existing system with its own interface, a specialized interface is not needed. Of course, the PC version could be loaded onto a mainframe if an emulation package is available. Emulation packages can be very slow and usually do not support interface features such as full-screen addressing.

SUMMARY

Materials databases can be classified according to several different methods: data type, user groups, applications, and access methods. Each database type distinctly changes the nature of a materials database, not only in the schema, but also in the user interface and data displays. A materials database builder should classify each materials database before beginning work. This assessment must be done with the users so that the resulting database will have the greatest acceptance. In some cases, this planning may show the database will not be cost effective. Finally, prior to using a database, a user should be aware of its purpose.

REFERENCES

[*1*] Rumble, J. R. Jr. and Smith, F. J., *Database Systems in Science and Engineering*, Adam Hilger, Bristol, England, 1984.
[*2*] Bullock, E., Kröckel, H., and van de Voorde, M., "Data Systems for Engineering Materials, the Materials Engineer's Point of View," *Materials Data Systems for Engineering, A CODATA Workshop*, Westbrook, J. H. et al, Eds., Fachinformationszentrum Energie, Physik, Mathematik GmbH, Karlsruhe, Germany, 1986.
[*3*] Mindlin, H. and Smith, S. H., "Database System Considerations in Engineering Design," *Managing Engineering Data: The Competitive Edge*, R. E. Fulton, Ed., American Society of Mechanical Engineers, New York, 1987.
[*4*] Westbrook, J. H., "Some Considerations in the Design of Properties Files for a Computerized Materials Information System," *The Role of Data in Scientific Progress*, P. S. Glaeser, Ed., North-Holland, Amsterdam, 1985.
[*5*] Rumble, J., Sauerwein, J., and Pennell, S., *Scientific and Technical Factual Databases for Energy Research and Development*, No. DOE/TC/40017-1, U.S. Department of Energy, Office of Scientific and Technical Information, Oak Ridge, TN, 1986.
[*6*] Ashby, M. F., "Materials and Shape," *Acta Metallurgica*, Vol. 39, No. 6, 1991, pp. 1025–1039.

MNL19-EB/Nov. 1993

Nomenclature and Current Standards for Identification of Engineering Materials

4

Bert Moniz[1]

IDENTIFICATION OF AN ENGINEERING MATERIAL

The identification or description of a material is one of the most important features of a materials property database. It is the primary method by which users enter and search any database and by which materials property developers recognize or code their materials.

If the material identifier or identifiers used in the search are too broad, we are inundated with irrelevant data. If it is too narrow, we may be denied extremely relevant data. Equally important, what is the value of any data if we cannot understand the description of the material? Identification is a complex subject and is difficult to relegate to a single field in a database.

Suppose we are searching for hardenability data on American Iron and Steel Institute (AISI) 4340 low alloy steel. This low alloy steel is often used in demanding applications where ability to strengthen (harden) large section sizes is an important criterion. Searching by "steel" or "low alloy" steel is obviously too broad a category. On the other hand, use of the identifier G43400 (the Unified Numbering System designation for AISI 4340) may not produce all the data we require. This is because H43400, an alternate designation for 4340, is used to describe a special version of 4340, developed to meet specified hardenability requirements. If we can experience such potential problems searching for data on a well-established alloy, imagine the difficulty of obtaining all the relevant data on materials having specialized applications, such as advanced ceramics or experimental polymers, where identification schemes are not well established.

Human and institutional factors complicate the problem of identification. We tend to associate materials with names common to our end use and expect to be able to search databases using these friendly names. For example, the term rubber has a strict identification, which is thermoset elastomer, yet most engineers would begin a general search using the term rubber. Consequently, there must be provisions for finding data using both the common and the more rigorous identifications.

Composites are a class of materials that usually contain discrete amounts of two or more different families of engineering materials. An adequate description of composites requires us to separately distinguish the basic materials that form the composite, plus describe the arrangement, shape and product form of the reinforcing component of the composite.

The same identification nomenclature may have entirely different meanings in different industries, such as the descriptions for end products in the aluminum and steel industry [*1*]. An industry's internally focused view of itself may also complicate the identification process, as with the use of the adjective, "advanced," in composites and ceramics. Advanced composites and advanced ceramics are highly engineered materials, but should be identified in computerized databases as composites and ceramics, respectively, since the qualifier "advanced" is bound to disappear with time.

OBJECTIVES OF IDENTIFICATION

To allow information from different databases to be compared, it is important to define the material identification features that are considered essential to any database. The number of essential fields could be reduced significantly if universal coding systems for different families of materials existed and were maintained.

The four objectives of describing a material in a form suitable for computerizing are as follows:

- To ensure that each material is unique
- To ensure that material equivalency can be determined to the level desired
- To ensure that the material can be found again
- To ensure that different identification nomenclature systems are supported (because there will always be different sets of nomenclature systems)

ASTM Subcommittee E49.01 or Materials Designation is chartered to develop guidelines for materials identification in computerized material property databases. The subcommittee has approached the task by producing generic guidelines for the major classes of engineering materials. These have led to more specific guidelines for materials or groups of materials within the major classes. Work is also proceeding on describing structural joints between materials and coatings or linings of one material upon another.

Subcommittee E49.01 has also reviewed and encouraged the development of unified identification codes for materials.

---

[1] Consultant, DuPont Company, P.O. Box 6090, Newark DE 19714-6090.

Copyright© 1993 by ASTM International www.astm.org

The development of these codes would greatly simplify data entry, exchange, and searching. Some unified coding systems are well accepted in specific geographic locations, such as the UNS in North America. There are many barriers, political and commercial, to the development of unified coding systems. Consequently, there is no ideal "one-step description system" for the various classes of materials, and no short-term solution to the problem.

GENERIC DESCRIPTION OF ENGINEERING MATERIALS

The information required to completely describe an engineering material may be placed into nine segments.

- Primary Identifiers
- Specifications
- Characterization
- Reference Test Results
- Source
- Processing History
- Product Form
- Fabrication
- Supplemental Information

The boundaries between these segments are flexible and vary with the experience and viewpoint of specific technical communities.

The nine segments have evolved through a process of iteration. Subcommittee E49.01 initially developed description formats for the four major groups of engineering materials (metals, ceramics, polymers, and composites), analyzed the types of fields and records in them, and began to place the information into various segments. As emphasized previously, the segments are not rigid and the order of individual fields is based on the protocols of the technical communities or industries that support any specific database.

Primary Identifiers

Primary identifiers are fields that broadly distinguish a material from other classes of materials or describe a particular material within a class. The four major classes of engineering materials are metals and alloys, polymers, ceramics, and composites. Although most primary identification fields are self evident, the standards provide guidance, when it is necessary, to decide the class in which a material should be placed. For example, most polymers contain additives and fillers to modify or improve their properties. Strictly speaking, this places them in the class known as composites; however, industry recognizes them as polymers or plastics. The draft Standard Guide for the Identification of Polymers (exclusive of Thermoset Elastomers) describes the differences between a reinforced polymer and a polymer matrix composite to help classify a material. The relevant format depends on the application of the database and the viewpoint of the users, as discussed in ASTM Guide for the Identification of Composite Materials in Computerized Material Property Databases (E 1309). The general differentiation between a homogeneous material and a composite material depends on the class of materials, for example, metal versus metal-matrix composites, and so forth.

Specifications

Specification fields contain information that is recorded on drawings, requisitions, and design data in order to process or fabricate a material. Where relevant specifications for a material exist, it is important to reference them. They help describe the material in its commercially available form(s) produced to industry standards such as ASTM. The number of applicable specifications usually grows with the commercial maturity of a material.

In commercial transactions, the year or revision code for a standard is usually omitted, for example, we write ASTM D 2466 for PVC Plastic Pipe Fittings, not D 2466-89. With computerized material property databases it is essential to include the year or revision code because properties may be changed as a result of revisions to a standard.

Characterization

Characterization fields contain structural, compositional, and other details that narrow the definition of a material. Measured properties in response to chemical, thermal, mechanical, or other stimuli, which are commonly used to describe a material, are placed in the Reference Test Results segment. For emerging engineering materials, characterization may include many types of information, such as description of grain size for a ceramic. With mature materials much of the important characterization data has been refined into standards and specifications, so there is less need for characterization data in the unique identification of a material.

Reference Test Results

Reference test results fields describe data from chemical, thermal, mechanical, or other tests used to characterize a material, especially when used and recognized, to demonstrate that a material has properties associated with a particular specification. Examples include a maximum hardness value for some steels or melt flow index for polyethylene.

Material Source

Material source fields contain information on where, when, and by whom a material was produced. The amount and type of information contained in this segment depend on the end usage of the database. End usage may be classified as internal or external, and experimental or commercial. Internal databases created by a material producer may contain information on experimental and commercial materials which are made available externally at the manufacturer's discretion. External databases usually contain only that information which the producer considers essential for unique materials identification and adequate understanding of the property data for external use. Experimental materials require more information about their source compared with commercial materials because information in segments, such as specifications and characterization, is less well developed or understood.

Processing History

Processing history fields contain information on how a material was processed and treated to achieve its primary shape and properties. These data are usually obtained from the primary producer of a material or product form. The amount and detail of processing history, as with the source segment, will depend on the end use of the database. Internal manufacturers' databases contain significantly more information or processing history, some of which may be proprietary or relevant to the competitive advantage of a product. Information that is not proprietary and which uniquely identifies a material is considered essential. For example, solution annealing temperature and method of cooling austenitic stainless steel are considered essential information because these items of data are critical to ensuring the corrosion resistance of the product.

Part or Sample Detail

Part or sample detail fields contain information on the produced shape or part, or the sample obtained from it to create a test specimen. The amount and type of information in this segment depends upon whether the material is an engineered product or a primary form. Engineered products encompass materials like composites, specific aluminum alloys, and some ceramics. These materials are manufactured close to, or in, the final form in which they are used, and are often tailored to a specific application. A considerable amount of information may be required to describe an engineered product. Primary forms are basic shapes made by primary manufacturers, such as steel sections or green ceramic parts. These are further processed by secondary manufacturing steps, and usually by an organization other than the primary manufacturer. Primary forms are well defined and require relatively little supplementary information to completely describe them. The part or sample detail segment contains no information on the test specimen, because it is either described in the specifications segment or it exists with the body of data used to record test results and properties. See Chapter 6.

Fabrication and Service History

Fabrication and service history fields contain information on the joining, machining, or shaping technique employed on the part, plus applicable service experience, which cannot be entered elsewhere. The information in this segment depends on the type of material: engineered product versus primary form. Engineered products (see previous section) may require relatively little additional fabrication information. Primary forms require a relatively large amount of fabrication information, especially if the database is developed by an end user of the material.

Supplemental Information

Supplemental information fields contain other relevant information that does not fit elsewhere.

TABLE 4.1—Format for identification of metals and alloys (ASTM E 1338).

Field Number[a]	Field Name
Primary identifiers	
1	material class
2	family name
3	family subclass
4	application group
5	product group
Specifications	
6	specification organization[b]
7	specification number[b]
8	specification version[b]
9	specification designation[b]
Characterization	
10	UNS number[b]
11	common name
12	compositional detail (key)[b]
13	elemental symbol[b]
14	measured weight percent[b]
15	minimum weight percent[b]
16	maximum weight percent[b]
Material source	
17	manufacturer
18	country of origin
19	manufacturer's plant location
20	production date
21	manufacturer's designation
22	lot identification
23	additional detail (key)
Processing history	
24	primary process type[b]
25	primary process detail (key)
26	secondary process type[b]
27	secondary process detail (key)
Part of sample detail	
28	part identification number[b]
29	geometric shape[b]
30	thickness[b]
31	width
32	length
Fabrication and service history	
33	fabrication history[b]
34	fabrication details (key)
35	service history[b]
36	service details (key)
Supplemental information	
37	supplementary notes

[a] Field numbers are for discussion purposes only.
[b] Essential field, if applicable.

Essential Fields

There must be equivalency between specific data sets and metadata in order to compare information between databases. The formats for identifying materials contain all the elements required for complete identification. For each type of material, there is a minimum amount of information that must be included to achieve the four objectives for describing a material outlined at the beginning of this chapter. Consequently, certain fields are designated essential in all the guidelines produced. For further discussion see Chapter 1.

STANDARDS FOR IDENTIFYING METALS AND ALLOYS

Metals are the most commonly used class of engineering materials; their properties are well understood, and they are supported by an enormous industrial infrastructure, which shapes, forms, and joins them. ASTM Guide for the Identification of Metals and Alloys in Computerized Material Property Databases (E 1338) is a generic guide applicable to all metals and alloys and is used to aid the database developer in defining essential and recommended fields to identify specific families of alloys (Table 4.1).

The essential fields in E 1338 relate to specifications; commonly used names or designations; chemical composition details; information on the primary and secondary manufacturing process and the shape produced; and fabrication and service history.

Metals and alloys are relatively mature products. Consequently, they are well characterized, and important features of the characterization segment have found their way into specifications. The principal area of characterization, which ought to be designated, is chemical composition (a range, or a maximum or minimum value). Reference test results (for example, a maximum allowable hardness value) may also be part of the identification.

Aluminum Alloys and Steels

Guidelines for aluminum alloys and steels have been drafted using the guidelines of the generic document. One of these has evolved into ASTM Guide for Identification of Aluminum Alloys and Parts in Computerized Material Property Databases (E 1339). See Table 4.2. Aluminum alloys are sometimes made as engineered products, details of which have to be included in the identification. To achieve this, ASTM E 1339 includes product and application group in the specifications segment and requires the application group to be an essential field.

A draft standard for steel products is currently being balloted (ASTM Draft Standard Guide for the Identification of Steel Alloys and Parts in Computerized Material Property Databases). See Table 4.3.

One widely used coding system, for metals and alloys, is the Unified Numbering System for Metals and Alloys [*2*], or UNS. It is based on chemical composition and consists of a letter followed by five numbers, for example, S30400 for 304 stainless steel or T30402 for D2 tool steel. Both materials are better recognized by their friendly shorthand descriptions, which are 304 and D2 from the original AISI designation system. The UNS is used principally for North American alloys and is not accepted globally. For several groups of nonferrous alloys (copper, aluminum, and magnesium), the UNS coupled with a temper designation provides an entry point into a database, which can uniquely describe a material, based on its allowable chemical composition and processing history (Table 4.4).

TABLE 4.2—Format for identification of aluminum alloys and parts (ASTM E 1339).

Field Number[a]	Field Name
Primary identifiers	
1	material class
2	family name
3	family subclass
Specifications	
4	specification organization[b]
5	specification number[b]
6	specification material[b] designation[b]
7	specification version[b]
8	UNS number[b]
9	common name[b]
10	application group
11	product group[b]
Characterization	
12[c]	element symbol
13[c]	actual weight percent
14[c]	maximum weight percent
15[c]	minimum weight percent
Material source	
16	manufacturer
17	country of origin number
18	manufacturer's plant location
19	production date
20	lot identification
Processing history	
21	temper[b]
22	temper detail
Part or sample detail	
23	part identification number
24	geometric shape
25	thickness[b]
26	width
27	length
Fabrication and service history	
28	fabrication history
29	service history
Supplementary information	
30	supplementary notes

[a] Field numbers, if applicable, are provided for information only.
[b] Essential field, if applicable.
[c] Fields 12 to 15 are repeated as often as needed.

STANDARDS FOR IDENTIFYING POLYMERS

Polymers consist of molecular chains and linkages characterized generally by the repetition of one or more types of monomeric units and usually based on carbon. Polymers are available as raw materials for formulation, as molding compounds or as shapes and parts, fabricated using a wide variety of processes.

ASTM Guide for Identification of Polymers (Excludes Thermoset Elastomers) in Computerized Material Property Databases (E 1308) is a generic guide for this class of materials (Table 4.5). It does not encompass thermoset elastomers (rubbers). They are different in nomenclature and processing requirements and will have their own identification format.

There are two levels of primary identifiers for polymers: (1) those that establish various polymer families and (2) those that distinguish members within a polymer family. The first level organizes information from ASTM Classification System for Specifying Plastic Materials (D 4000) (Fig. 4.1). The second level permits documentation of various types of compositional, structural, and application information, plus any reference test results commonly used to identify the polymers within one family. The last named are identified by a repeating field called, Classification Property (corre-

TABLE 4.3—Draft format for identification of steel alloys and parts.

Field Number[a]	Field Name
Primary identifiers	
1	material class
2	family name
3	family subclass
4	application group
5	product group
Specifications	
6	specification organization
7	specification number
8	specification material designation
9	specification version
10	UNS number
11	AISI designation
12	common name
13	grade
14	type
15	commercial name
Characterization	
16	element symbol
17	actual weight percent
18	maximum weight percent
19	minimum weight percent
Material source	
20	manufacturer
21	country of origin number
22	manufacturer's plant location
23	production date
24	heat number
Processing history	
25	melt practice
26	cast practice
Heat treatment	
27	thermal cycle
28	time of thermal cycle
29	temperature of thermal cycle
Part or sample detail	
30	part identification number
31	geometric shape
32	thickness
33	width
34	length
35	diameter type
36	diameter
37	weight
Reference test results	
38	ultimate tensile strength, min
39	ultimate tensile strength, actual
40	number of ultimate tensile tests
41	number of ultimate tensile tests
42	yield strength, min
43	yield strength, actual
44	yield strength, max
45	number of yield strength tests
46	elongation, min
47	elongation, actual
48	elongation, max
49	number of elongation tests
50	reduction of area, min
51	reduction of area, actual
52	reduction of area, max
53	number of reduction of area tests
Hardness	
54	hardness
55	hardness scale
Impact	
56	Charpy V-notch toughness value (average of 3 tests)
57	temperature for Charpy tests

[a] Field numbers, if applicable, are provided for information only.

TABLE 4.4—UNS and temper designation calls out much of the essential information on some alloys.

- A96061-T6
  Aluminum Alloy 6061 in artificially aged condition
- M11610-F
  Magnesium Alloy AZ61A in as-fabricated condition
- C82600-TB00
  Beryllium Copper Alloy 245C aged at 345°C for 3 h

TABLE 4.5—Format for identification of polymers (excluding thermoset elastomers).

Field Number[a]	Field Name
Primary identifiers	
(1)*	material class
(2)*	polymer class
(3)*	polymer family
(4)*	family abbreviation/code
Specifications	
(10)[a]	specifying institution
(10a)	specification number
(10b)	revision number or version
(10c)	issue date
(10d)	grade designation
(11)*	unified name (line call out)
(12)[a]	other naming systems
Characterization	
(5)*[a]	component(s)/composition
(5a)[a]	component attribute
(5b)	measured component quantity
(5c)	maximum component quantity
(5d)	minimum component quantity
(5e)	residual monomer content
(6)*[a]	attribute
(6a)	attribute
(6a) (Iteration 1)	attribute
(6a) (Iteration 2)	attribute
(6a) (Iteration 3)	attribute
(7)*[a]	modifier(s)/modifier composition
(7a)[a]	modifier attribute
(7b)	measured modifier quantity
(7c)	maximum modifier quantity
(7d)	minimum modifier quantity
(8)*[a]	application descriptors
(8a)	attribute
(8a) (Iteration 1)	attribute
(8a) (Iteration 2)	attribute
Reference test results	
(9)*[a]	classification property
(9a)	test method
(9b)	test specimen
(9c)	value or range
(9) (Iteration 1)*	classification property
(9a)	test method
(9b)	test specimen
(9c)	value or range
(9) (Iteration 2)*	classification property
(9a)	test method
(9b)	test specimen
(9c)	value or range
(9) (Iteration 3)*	classification property
(9a)	test method
(9b)	test specimen
(9c)	value or range
(9) (Iteration 4)*	classification property
(9a)	test method
(9b)	test specimen
(9c)	value or range

TABLE 4.5—Format for identification of polymers (excluding thermoset elastomers). (Continued)

Field Number[a]	Field Name
Source	
(13)[a]	polymer production
(13a)	monomer purity
(13c)	source–manufacturer
(13d)	source–not manufacturer
(14)	traceability: plant
(14a)	lot number
(14c)	date
(15)	molding compound production
(15a)	formulation
(15b)	processing
(15c)*	use of reclaim
(15d)*	polymerization state
(15e)*	form
(15f)*	commercial name/grade
(15g)*	source—manufacturer
(15h)	source—not manufacturer
(16)	traceability: plant
(16a)	lot number
(16b)	manufacturing date
(16c)	shipping date
Processing history	
(17)*	additional formulation
(17a)*	regrind, rework usage
(18)*[ae]	processing method
(18a)*[a]	processing conditions
(18b)	processor (molder)
(18c)	traceability: plant
(18d)	lot number
(18e)	date
Part or sample detail	
(19)*	sample form
(20)*	sample material state
(21)*[a]	sample modification
(21a)*	processing
(21b)*	treatment

(*) is used to mark essential fields or segments.

[a] Some fields and segments may need to be repeated for multiple item listings.

[b] Segment/Field Number/Designation is for convenience only.

[c] Units for data not resolved. Use of ISO units is favored for eventual harmonization.

[d] Specification D 4101 does not give the year of publication of the test methods it references. Documents associated with Classification D 4000 might include the publication year to better define the test methods. When known, the complete test method designation should be used, that is, Test Method D 638.

[e] Iteration example: Thermoforming could require sheet extrusion before part or shape fabrication.

sponding to Reference Test Results), for example, Flow Rate per ASTM Test Method for Flow Rates of Thermoplastics by Exclusion Plastometer (D 1238).

Essential fields for identifying polymers are shown in Table 4.3 and include information from the Primary Identifiers, Polymer Family Identifiers (Characterization), Specifications, Processing History, and Sample Detail segments.

Polymer Versus Polymer Matrix Composite

It is sometimes necessary to distinguish a reinforced polymer from a polymer matrix composite. In practical use, discrete second-phase ingredients (modifiers) are incorporated into many polymers to enhance properties or cost. Many, if not a majority of commercial "polymers" (plastics), should be strictly classified as "composites." The polymer identification document offers some guidelines for a database manager in distinguishing traditional reinforced polymers from polymer matrix composites.

The modifiers in a polymer are generally small relative to molded part size and are supplied in random orientation to the molding equipment as part of a molding compound. Although they can enhance the strength or other property of the polymeric part or shape, the polymer reinforcements do not require the detail of structure (for example, weave or braid) or precise geometric placement (for example, orientation or order of lay-up) in processing or identification as do reinforcements in composite materials. When such detail of reinforcement structure and geometric placement is required, the composites identification standard, ASTM E 1309, should be used.

STANDARDS FOR IDENTIFYING CERAMICS

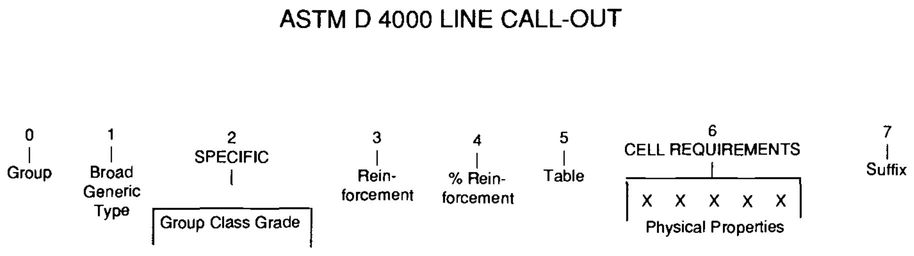
Ceramics encompass nonmetallic inorganic materials, generally formed by the application of heat. They include oxide and nonoxide compounds and are either crystalline or noncrystalline. For the purpose of identification, ceramics consist of two main groups, which are refractories and advanced ceramics. Refractories have widespread structural uses, but there has been little driving force to standardize data on their identification and properties. Advanced ceramics are processed into engineered products that extend the range of physical environments and mechanical conditions under which they can operate. The generic identification standard of ceramics is currently a draft format (ASTM Guide for the Identification of Ceramics in Computerized Material Property Databases). See Table 4.6. The word, "advanced," was removed from the title because it becomes redundant with time. However, the subject material covers what are currently known as advanced ceramics.

Historically, ceramics have been identified by designations related to their chemical composition. This is not sufficient to describe in many varieties of ceramics that are in use today. No singular or unified coding system has gained broad agreement, but a Versailles Agreement on Advanced Materials (VAMAS) group is working on it. Consequently, the number of primary identifiers is relatively large compared with other materials. Since ceramics are highly tailored to end usage, the applications group field is essential information.

Although some specifications exist for ceramics to facilitate their engineering application, the characterization segment is extremely important for an adequate description of the material, especially for the many ceramics that are tailored to specific end uses. The characterization fields included describe the chemical and physical features of the ceramic. For the same reason, processing information is often highly specific and uniquely defines a product. Many of the ceramics processing fields are essential.

STANDARDS FOR IDENTIFYING COMPOSITE MATERIALS

Composite materials, usually called composites or advanced composites, are the most complex engineering ma-



0 - One digit for expanded group, as needed.
1 - Two or more letters identify the generic family based on Abbreviations D 1600.
2 - Three digits identify the specific chemical group, the modification or use class, and the grade by viscosity or level of modification. A basic property table will provide property values.
3 - One letter indicates reinforcement type.
4 - Two digits indicate percent of reinforcement.
5 - One letter refers to a cell table listing of physical specifications and test methods.
6 - Five digits refer to the specific physical parameters listed in the cell table.
7 - Suffix codes indicate special requirements based on the application, and identify special tests.

FIG. 4.1—ASTM D 4000 designation system can provide detailed identification of polymers.

terials to identify. Composites consist of two or more materials that are insoluble in each other and are considered on a macroscopic scale to form a single engineering material that exhibits specific properties not possessed by the constituents. Composites are comprised of metal, polymer, and ceramic matrix (base) types. The are made into precursors (primary shapes later formed into end-use products) and engineered products.

ASTM E 1309 documents the format for the identification of all composite materials. It references ASTM Guide for the Identification of Fibers, Fillers, Core Materials in Computerized Databases (E 1471), which uses the metal, polymer, and ceramic formats for information on constituent materials. Presently E 1309 includes examples of application to filament-wound polymer-matrix composites, laminar polymer-matrix composites, and ceramic-matrix composites. An example of metal-matrix composites will be added when data are available with no limits on distribution.

ASTM E 1309 is a generic guide. Presently it includes examples of application to filament-wound, polymer-matrix composites, laminar polymer-matrix composites, and ceramic-matrix composites. It is being broadened to include metal-matrix components.

The identification of composites is complex because they are formed by combining different materials in varying amounts and configurations, resulting in an infinite number of possibilities. An effective description scheme must be capable of identifying the majority of possible combinations, without overburdening the system to include details relevant only to a limited number of material systems.

The primary objective of ASTM E 1309 is to describe the final composite material (engineered product) and its precursors, so that minimal information is contained on the constituent materials that form the composite matrix or the reinforcement. The constituent materials of the composite should themselves be fully described where possible, according to the applicable guides for metals, polymers, and ceramics or other materials.

ASTM E 1309 concentrates on describing the form, orientation, and ordering of the reinforcement component of a composite material. Essential fields include information on the precursor source, matrix identifier, reinforcement identifier, and process description (Table 4.7).

ASTM E 1471 covers the description of fibers (continuous and discontinuous), fillers of various geometries, and core materials (foam, honeycomb, or naturally occurring materials such as balsa wood), which are used in sandwich-type composites. These materials are distinguished from bulk materials by the importance of their tailored geometric shape to their properties, which is reflected in the use of essential geometry fields as a major component in their characterization (Table 4.8).

STANDARDS FOR IDENTIFYING JOINTS BETWEEN MATERIALS

A joint between two or more materials requires a radically different identification format from those used to identify the materials themselves. Special classes of materials may be used to obtain permanent joints such as weld filler metals, brazes, solders, or adhesives. The joining parameters are significant components in the identification of a joint.

There are many types of joints between many kinds of materials. A draft format was developed for the most commonly used industrial joint, the arc weld. The format is published by the American Welding Society (AWS) Guide for Describing Arc Welds in Computerized Material Property and Nondestructive Examination Databases (AWS A9.1).

The materials description segments in AWS A9.1 consist of the first base metal, second base metal, backing material, filler material, additives, gases, and tungsten electrode. The welding procedure description segments consist of applicable standard, joint, welding process, welding technique, welding orientation, welding thermal cycle, postweld heat treatment, electrical characteristics, weld metal, and testing laboratory information (Table 4.9).

TABLE 4.6—Draft format for identification of ceramics.

Field Number	Field Name
Primary identifiers	
1	*material class
2	*structure class
3	*ceramic class
4	*traditional name
5	**Oxide type
6	*primary chemical constituent
7	*minimum purity of primary chemical constituent
8	*maximum purity of primary chemical constituent
9	*minor chemical constituents
10	*major application group
11	Secondary application group
12	Tertiary application group
13	Ceramic identification code
14	Code organization
Specifications	
15	*specification organization
16	*specification number
17	*specification version
18	*specification designation
19	supplemental specification information
Characterization	
20	chemical component
21	chemical symbol or formula
22	chemical component quantity
23	reported estimated standard deviation
24	measurement units
25	method of analysis
26	phase component
27	chemical formula or analysis of phase
28	phase quantity
29	repted estimated standard deviation
30	measurement units
31	grain boundary structure class
32	measured density
33	density test method
34	percent of theoretical density
35	theoretical density calculation method
36	percent porosity
37	type of porosity
38	mean grain size
39	grain size estimated standard deviation
40	grain size distribution
Material source	
41	manufacturer
42	country of origin
43	production date
44	lot number
45	manufacturer's designation
46	ceramic type
Processing history	
47	*manufacture of precursor
48	*manufacture of powder
49	*preparation of powder
50	*compaction
51	*shaping
52	consolidated form
53	*consolidation process
54	consolidation process conditions
55	*finishing process
56	finishing process conditions
57	joining process
58	joining process conditions
Part or sample detail	
59	product geometric form
60	product length dimension
61	product area dimension
62	product volume dimension
Fabrication and service history	
63	*service history
Supplementary information	
64	supplementary notes

STANDARDS FOR IDENTIFYING COATINGS AND LININGS

Coatings and linings are materials that may be applied to another material (substrate) for the purpose of improving the properties of the substrate. The coating or lining is usually relatively thin compared with the substrate and are made of metals and alloys, polymers, ceramics, or composites.

A draft format for the description of metallic coatings ASTM Draft Guide for Identification of Coatings of Engineering Materials in Computerized Material Property Databases, is in the balloting stage. Key differences between it and the engineering materials formats lie in the processing history segment, both of which require information on surface preparation and in the type of reference test results used.

The identification segments required for coatings and linings include generic field types that are not utilized in the formats previously described for engineering materials and joints. These include preparation of the identified substrate surface, method of application of the identified coating or lining, form of material applied, homogeneity of coating or lining, porosity, thickness, and uniformity (Table 4.10).

UNIFIED CODING SYSTEMS FOR MATERIALS

Many designation systems for engineering materials have been established this century, principally to serve the needs of commerce. For example, trade associations such as the Copper Development Association, were the first to develop orderly designation systems for their materials. The multiplicity of systems has sometimes hampered materials selection, but the problems have been tolerable because materials specialists have been able to interface and interpolate between designation systems as needed.

The emergence of computerized material property databases has seen a pressing need to harmonize designation systems. Without harmonization, it is more difficult to link in-

TABLE 4.7—Format for identification of composites (ASTM E 1309).

Field Number	Field Name
Primary identifiers	
1	material reference number
2	material class[a]
3	matrix class[a]
4	reinforcement class[a]
5	symmetry type
6	structural detail
7	ply count
8	lay-up code
Precursor identifiers[b]	
9	precursor material reference number
10	precursor type[a]
11	precursor commercial name[a]
12	precursor manufacturer[a]
13	precursor production date
14	precursor lot number
15	precursor specification organization
16	precursor specification number
17	precursor specification version
18	precursor specification designation
19	precursor dimension parameter[c]
20	precursor dimension value[c]
21	precursor reinforcement content by weight
22	precursor reinforcement orientation(s)
23	precursor additional information
24	precursor reinforcement area weight
25	precursor volatile content, by weight
26	volatile content test conditions
27	precursor matrix flow, by weight
28	matrix flow test conditions
29	precursor matrix gel time
30	matrix gel time test conditions
Matrix identifiers[b]	
31	matrix cross reference[a]
32	matrix subclass[a]
33	matrix chemical family[a]
34	matrix commercial name[a]
Reinforcement/core identifiers[b]	
35	reinforcement/core material[a] cross reference
36	reinforcement/core class[a]
37	reinforcement/core subclass[a]
38	reinforcement/core chemical[a] class
39	reinforcement form[a]
40	reinforcement/core commercial[a] name
41	reinforcement/core production date
Process descriptors[d]	
42	process specification organization
43	process specification number
44	process specification version
45	process stage type[a]
46	process equipment type
47	process stage conditions
48	processor[a]
49	process date
50	process records reference
Part descriptors	
51	part form[a]
52	part-dimension parameter[c]
53	part dimension value[c]
54	part specification organization
55	part specification number
56	part specific version
57	part specification designation
58	part reinforcement content, by volume[a]
59	part additional information
60	part void content, by volume
61	part volatile content, by weight
62	volatile content test conditions
63	part history[a]

[a] Essential field, if applicable.

[b] Composites may contain more than one precursor, matrix material, or reinforcement. These fields should be supplied for each.

[c] Dimension parameter and value should be included for each relevant dimension. Type is essential if value is given.

TABLE 4.8—Format for identification of fibers, fillers, and core materials (ASTM E 1471).

Field Number	Field Name
Primary identifiers	
1	material reference number
2	class[a]
3	subclass[a]
4	chemical family number[a]
Specifications	
5	common name[a]
6	additional name information
7	specification organization
8	specification number
9	specification version
10	specification designation
Characterization	
11	density[a]
12	cross section type
13	dimension parameter[b]
14	dimension value[b]
15	dimension distribution parameter[c]
16	dimension distribution parameter value
17	dimension distribution sample size
Material source	
18	manufacturer
19	manufacturer's identification
20	lot number
21	date of manufacture
Processing history	
22	process conditions
23	surface treatment type
24	surface treatment detail

[a] Essential field, if applicable.

[b] Dimension parameter and value should be given for each relevant dimension. Type is essential information if value is given.

[c] For each dimension where distribution width is relevant, the parameter is essential if parameter value is given.

TABLE 4.9—Format for identification of arc welds (AWS A9.1).

Material Information

- First base metal
  - specification organization
  - specification number
  - specification version
  - specification designation
  - UNS number
  - CAS number
  - common name
  - manufacturer/supplier
  - heat/lot identification
  - composition (repeat as many times as necessary)
    - elemental symbol
    - measured weight percent
    - method of analysis
  - material condition
    - manufacturing history
    - service history
  - product form
  - thickness (wall thickness for pipe)
  - diameter (for tubular sections)
- Second base metal (repeat if different from first base metal)
- Backing (repeat per first base metal if used)
- Filler metal (if more than one filler material, repeat for each)
  - specification organization
  - specification number
  - specification version
  - classification
  - UNS number
  - CAS number
  - common name (tradename)
  - manufacturer/supplier
  - heat/lot identification
  - date of receipt of material
  - composition (repeat as many times as necessary)
    - elemental symbol
    - measured weight percent
    - ymethod of analysis
  - product form
    - dimensions
  - other information
- Flux
  - specification organization
  - specification number
  - specification version
  - specification title
  - particle size
  - CAS number
  - common name/type (tradename/type)
  - manufacturer/supplier
  - heat/lot identification
  - date of receipt of material
  - composition (repeat as many times as necessary)
    - element/compound symbol
    - measured weight percent
    - method of analysis
  - additive (alloying materials, etc)
    - type
    - amount
    - placement
- Gases (repeat as necessary)
  - specification organization
  - specification number
  - specification version
  - specification title
  - CAS number
  - common name/type
  - manufacturer/supplier
  - lot identification
  - date of receipt of material
  - composition (repeat as many times as necessary)
    - elemental symbol
    - measured weight percent
    - method of analysis
  - dew point
  - gas nozzle inside diameter
  - gas nozzle flow rate
  - root shielding (back side) flow rate
  - trailing gas shield flow rate
  - plasma orifice gas flow
  - other information
- Tungsten electrode
  - specification organization
  - specification number
  - specification version
  - classification
  - UNS number
  - CAS number
  - common name/type
  - manufacturer/supplier
  - heat/lot identification
  - diameter
  - electrode tip geometry
  - comments (material and consumables)

Welding Procedure

- Applicable standard
  - specification organization
  - specification number
  - specification version
  - specification title
- Joint
  - joint type
  - joint preparation techniques
  - groove type
  - surface preparation techniques
  - backing placement
  - back gouging
  - other information
- Weld Type
  - welding process (repeat as many times as necessary)
  - method of application
- Setup
  - contact tip to work distance
  - electrode extension
  - travel speed
  - oscillation method
  - oscillation width
  - pass/layer sequence (and limits)
  - peening
  - interpass cleaning techniques
  - other information (multiple electrodes, hot or cold wire feed, tacking electrode extension/set back)
- Orientation
  - weld position (orientation of joint)
  - welding progression
  - travel angle
  - work angle
  - other information
- Welding Thermal Cycle
  - preheat temperature
  - preheat maintenance (temperature kept above minimum during duration of welding)
  - interpass temperature, max
  - postheat temperature (for example, promote hydrogen outgassing)
  - postheat time
  - other procedure details

TABLE 4.9—Format for identification of arc welds (AWS A9.1). (Continued)

- Postweld heat treatment
  - heating rate
  - holding temperature range
  - holding time
  - cooling procedure (cooling rate or quench procedure)
  - other procedure details (for example, step cooling procedure, furnace type, atmosphere)
- Electrical characteristics (repeat for multiple power sources and for change as the weld progresses)
  - manufacturer/supplier
  - common name
  - current type
  - polarity
  - welding current (rms)
  - wire feed speed
  - welding voltage
  - comments (for example, pulsed power parameters, machine settings, and conditions for parameter changes)
- Weld Metal
  - composition (in actual weld, including dilution) (repeat as many times as necessary)
  - elemental symbol
  - measured weight percent
  - method of analysis

Laboratory Testing Information

- Weld identification number
- Laboratory Which Performed Test
  - Laboratory PQR Identification
  - Date of Completion
- Comments general

formation from different databases and fully utilize the power of the computer to manage information. Currently, 30 to 70 fields are required to identify a material with as many as 10 to 20 being considered essential.

Current Situation

The current situation with respect to unified coding systems is given below.

- For metals and alloys, the compositionally based UNS has wide use in North America. It may be enhanced by including a temper designation (to include heat treatment and mechanical working information) in the identification. There is no international system. ISO tried to introduce one several years ago and did not gain sufficient international support.
- For polymers, the numbering systems given in ASTM D 4000 (for polymers) and ASTM Guide for Classifying and Specifying Adhesives (D 4800) (for adhesives) appear to be excellent starting points for coding systems.
- For ceramics there is no system. However, VAMAS in conjunction with ASTM Committee C-28 on advanced ceramics is attempting to develop a coding system [*3*].
- For composites, the complexity of the situation has defied anything but a few proposals for specific groups [*4*].

TABLE 4.10—Draft format for identification of coatings.

Field Number	Field Name
Primary identifiers	
1	material form
2	coating name[a]
3	coating subclass[a]
4	application group[b]
5	product group[b]
Specifications	
6	specification organization[a]
7	specification number[a]
8	specification version[a]
9	specification designation[a]
10	common name[a,b]
11	grade or type
12	commercial name
Characterization	
13	average coating number[b]
14	maximum coating thickness[b]
15	minimum coating thickness
16	identification of coating material[c]
Material source	
17	manufacturer
18	country of origin
19	manufacturer's plant location
20	production date
21	batch number
Processing history	
22	coating applicator
23	country of application
24	applicator's plant location
25	application date
26	application batch number
27	number of coatings
28	surface cleaning
29	surface profiling
30	surface priming
Part or sample detail	
31	part identification number
32	surface identification number
33	surface reference document code
34	surface area
Reference test results	
35	post coating surface resistance
36	post coating surface impedance
37	post coating surface thermal conductivity
38	post coating surface reflectivity
39	post coating hardness
40	scale for post coating hardness

[a] Essential field, if applicable.
[b] Field is repeated as often as needed.
[c] Use formats that uniquely identify the coating material, for example, ASTM 1338.

REFERENCES

[*1*] Westbrook, J. H., "Designation, Identification and Characterization of Metals and Alloys," *Computerization and Networking of Materials Databases*, STP 1017, American Society for Testing and Materials, Philadelphia, 1989, pp. 23–42.

[*2*] *Metals and Alloys in the Unified Numbering System, ASTM DS 56E*, American Society for Testing and Materials, Philadelphia, 1993.

[*3*] *ISR Newsletter*, Vol. 2 Issue 4, Dec. 1991, pp. 9–10.

[*4*] *Aluminum Association Nomenclature System for Aluminum Metal Matrix Composite Materials*, Aluminum Association.

MNL19-EB/Nov. 1993

5

Nomenclature and Current Standards for Recording of Test Results and Properties

Marilyn W. Wardle[1]

INTRODUCTION

One of the important tasks associated with the design phase of a materials properties database is defining and understanding the exact nature of the data to be stored in the database. Among the key questions are:

(1) What are the data types: character strings, numbers, numerical arrays, graphics, images?
(2) For numerical data, what is their format: integer, floating point, exponential? How many decimals? How many significant figures?
(3) What metadata need to be associated with the data, and what is their form?

Without a thorough understanding of the data themselves it is difficult for the database designer to create an adequate vehicle for storage and retrieval. Since in many cases the designer is a computer systems person or someone not intimately familiar with the materials and functions of the database, it is vitally important that this input be gathered from those who do know the materials and applications, namely, the people who will be supplying data to the database and those who will be using it.

One of the problems that the designer will rapidly discover is that within the materials community there is a wide spectrum of usage in things like property names, definitions, and data reporting practices, which are critical to the database design. The human mind has an innate capability for dealing with and resolving ambiguity and "fuzziness," which is not shared by the computer. Thus, while an engineer may cope easily with different names for the same property (elastic modulus, Young's modulus, initial modulus), the computer is much more limited. The computerization of material properties data requires a far higher level of standardization in this area than we have previously enjoyed.

The data reporting standards discussed in this chapter are the first steps in this direction. They attempt to capture the thinking of scientists and engineers working in the field and subject it to the consensus process for development of standards. Clearly the few documents discussed here represent a very limited subset of the entire range of materials properties that are measured and reported, and ongoing work in this area is required. Furthermore, these available documents are only guides, intended to aid the database designer in making decisions regarding database content and organization. They are by no means a complete solution nor do they eliminate the need for close consultation with the materials experts.

In most cases the standardization process in Committee E-49 began with material identification and description formats, so the standards discussed in the previous chapter are fairly well advanced at least for the common classes of materials. Data recording formats generally were not begun until the material identification formats were well along, so they are in a somewhat earlier stage of development. It is important to keep in mind throughout the discussion of these standards that we are looking at a "snapshot" of them in a particular state of development, and this snapshot may have changed by the time this manual is in the hands of the reader.

STANDARD FORMATS

Standardization is essential to extracting maximum value and utility from a database. For example, if a search for "compressive strength" values overlooks potentially useful data because they are stored as "compression strength," the database is clearly not delivering full value. Some latitude in property naming can be accommodated through an internal thesaurus, but this requires considerable forethought and sufficient knowledge of the materials and properties.

Standard data formats are important for two additional reasons. First, data exchange between database systems is greatly facilitated if the properties are defined and formatted the same. This is especially true if format information is not stored as metadata along with the data and may not be transmitted with them. Second, proper definition and use of standard formats preserves the intent and precision of the original measurements and allows proper display of numerical data. (It is important that the storage and display formats for numerical data not imply a level of precision higher than that justified by the original measurement technique, a factor that is commonly overlooked in database design, particularly where internal unit conversion and calculations are done.)

Data entry to the database is much easier if the data are presented at the entry point in the desired format. This means that the use of the standard formats should ideally be incorporated into all of the processes before entry to the da-

---

[1] Senior research associate, E.I. Du Pont de Nemours & Company, Inc., Advanced Materials Systems, Chestnut Run Plaza, Wilmington, DE 19880-0701.

Copyright© 1993 by ASTM International www.astm.org

TABLE 5.1—Issued standard guides for data recording.

Number	Title	Test Method(s)	Jurisdiction (Liaison)
E 1313-91A	Development of Standard Data Records for Computerization of Material Property Data	. . .	E-49 (E-28)
X1	Pin Type Bearing Test of Metallic Materials Bearing Test Data	E 238	E-49 (E-8)
X2	Plane-Strain Fracture Toughness of Metallic Materials	E 399	E-49 (E-28)
X3	Tension Testing of Metallic Materials	E 8	
X4	Compression Testing of Metallic Material at Room Temperature	E 9	
X5	Notched Bar Impact Testing of Metallic Materials	E 23	
E 1434-91	Development of Standard Data Records for Computerization of Mechanical Test Data for High-Modulus Fiber-Reinforced Composite Materials	various	D-30 (E-49)
A 9.2	Recording Arc Weld Material Property Data in Computerized Databases	. . .	American Welding Society (E-49)
G 107-91	Formats for Collection and Compilation of Corrosion Data for Metals for Computerized Database Input	. . .	G-1 (E-49)

tabase. For this reason these formats should be regarded as data recording formats and not just formats for computerization of data. The formats will be most effective if they are used all the way from the testing laboratory, where measurements are made, through analysis and statistical processing, to the database gateway. In this way one can be certain that what goes into the database has the same level of precision and reliability as that which was originally produced.

It should be clear by now that there is a great deal more to material properties data than simply numerical data values. Along with each piece of data, there is a body of accompanying metadata that is necessary to the understanding and interpretation of the numerical value. These pieces of information are as important as the numerical values themselves and need to be treated with as much care. Thus a great deal more of the data recording format is devoted to defining the metadata requirements than to the data themselves. Again, use of these formats throughout the data generation process is the best way to ensure that the metadata are complete and accurate.

A list of data reporting format guides which have been finalized is given in Table 5.1 along with the ASTM Standard Test Methods to which the guide applies. Table 5.2 lists guides that are known to be under development. Unless otherwise noted, jurisdiction lies with ASTM Committee E-49. In many cases other ASTM committees (shown as liaisons) are actively involved in the drafting of the guides. In most cases responsibility for the documents will eventually be transferred to their jurisdiction. It is the goal of Committee E-49 to transfer the data reporting standards to the appropriate technical committees for refinement and maintenance, since they represent the principle users of the standards. In one case of the document on arc weld properties, the guide was originally developed jointly by E-49 and the American Welding Society (AWS) and has now been turned over to the AWS for maintenance.

GENERIC FORMATS

ASTM Guide for the Development of Standard Data Records for Computerization of Material Property Data (E 1313)

Guide E 1313 identifies the types of metadata that should be included in a materials properties database. Since it is not material, test, or property type specific, as are the rest of the standards to be discussed, it is necessarily a very generic document posed at a high-level of abstraction. However, it gives us a framework for discussion and comparison of the other, more specific guides, and points the way for those yet to be written.

The annexes to ASTM E 1313 contain examples of material and test type specific guides that will be discussed later. (Whether these guides will remain as annexes to ASTM Guide E 1313 or be attached to the specific ASTM Test Methods to which they apply is still under discussion. Alternatively, ASTM Guide E 1313 may be amended to limit its scope to metallic materials, and all the metals data reporting formats may be incorporated as annexes to it. In any case, ASTM E 1313 is the generic parent from which all of the data reporting formats derive.)

One of the first points ASTM Guide E 1313 makes is that it, and all of the specific guides that spring from it, are not intended to stand alone, but are to be used in conjunction with the applicable guide for identification for the materials under discussion (see Chapter 4). Certain information fields appear in both the identification format and the data recording format. These provide the necessary linkage between the material description, its property data, and the data source. In an actual physical database this redundancy is not necessary if the relationship is otherwise defined by the database structure. Typically the redundant items are as follows:

- Generic type or class of material
- Specific material within class
- Material designation
- Material producer or source
- Lot number or other identification
- Data source

The last item, data source, does NOT appear in all of the material identification formats, in which case it should definitely be included in the data reporting format.

Another feature of the data recording formats is that, like the material identification formats, they identify both essential and desirable data fields. Essential fields are those without which the ability to make meaningful data comparisons is seriously compromised. An example of an essential field

TABLE 5.2—Standard guides for data recording in preparation.

Number	Title	Designation	Jurisdiction (Liaison)
E 1313	Development of Standard Data Records for Computerization of Material Property Data		E 49
X20	Measurement of Fatigue Crack Growth Rates	Test Method E 647	(E 08)
X21	Constant-Amplitude Low-Cycle Fatigue Testing	Practice E 606	(E 08)
X??	Creep, Creep Rupture, and Stress Rupture Tests of Metallic Materials; Time-for-Rupture Notch Tension Tests of Materials	Practice E 139; Practice E 292	(E 28)
	Recommended Data Format for Wear Test Data Suitable for Databases		(G 02)
	Development of Standard Data Records for Computerized Storage and Transfer of Digital Ultrasonic Test Data		(E 07)
	Recommended Standard Data Records for Computerized Storage and Transfer of Digital Radiologic Test Data		(E 07)
	Recording Property Data for High Explosives in Computerized Databases		(E 27)
	Tensile Properties of Plastics	Test Method D 638, D 638M	(D 20)
	Standard Fracture Toughness Test Data	New Method	E 08

for reporting data on a notched bar impact test is "specimen type," while an optional field is "testing machine type." In the opinion of the writers and reviewers of this guide, the conditions and constraints of the notched bar impact test make it meaningless to compare data from different sources if the specimen type is not known, while such a comparison can be made without knowledge of the testing machine. The comparison will be more definitive, however, if the test machine type is also known. Obviously, what is essential and what is merely desirable will vary from one test method and material to another, and is influenced by the prejudices and experiences of those who write and review the standards, which may differ from those of other experts in the field. This, however, is the nature of consensus standards and is one reason why these documents are presented as guides.

As we shall see, some of the material and test specific formats introduce a variant on the essential field theme, incorporating fields that are conditionally essential. That is, they are essential only if another field contains a particular value. An example of such a field might be, "calculation method," which is essential only if the calculated property is reported.

In general the formats include a few essential fields and a much larger number of desirable fields. One intent was to help prevent the database developer from inadvertently omitting something that might be of importance to the application. The number of fields depends on the complexity of the materials and the specificity of the tests. Thus the formats for recording static mechanical properties of composite materials typically have many more fields than those for metals, where the materials are not as varied and the test methods are more definitive.

In addition to the information common to the materials identification formats, the ASTM Guide E 1313 calls for the following classes of metadata:

Test and Specimen Description

It is unfortunately all too common to find materials properties data for which test and specimen description are not reported. For some materials, the results of property measurements are highly dependent on the details of the test and specimen geometry, and failure to report these makes the data useless for many purposes. Furthermore, while for established materials, such as metal alloys, simply citing the corresponding ASTM Test Method may completely define the test specimen geometry; this is often not the case for materials like composites and polymerics, where the existing test methods allow considerable latitude. Thus it is necessary to include this information explicitly in the data report.

Test Conditions

Along with the test description, of course, must go the environmental conditions under which the test was performed, and if not included in the material description, information on preconditioning or previous service history of the material tested. A problem peculiar to computerized systems arises in formatting fields for parameters like test temperature, since most database management systems require that fields be predefined as either numeric or character string. As long as actual numeric temperature values are consistently reported, this is not a problem. Unfortunately test temperature is commonly reported as "ambient" or "room temperature," where temperature during the test was not controlled. Translating these into a nominal numerical value may create a false impression of the degree of temperature control that was actually exercised, while storing temperature in a character field may result in a less efficient search procedure. This may seem a trivial concern; however, it is an illustration of the type of problem with which the database developer must grapple to create a system that is both efficient and easy to use.

Another potential source of confusion about test conditions, when dealing with laboratory data, is the distinction between nominal and actual measured values of parameters like test temperature. In some cases data may be reported with a nominal test temperature, that is, the temperature of the lab or the temperature at which the environmental chamber was held during the test. In other cases, the actual specimen temperature at the time of test may be recorded based on readings from a thermocouple or other monitoring device. The database design should make clear which of these is intended. In some cases, it may be desirable to include fields for both.

Property Descriptions

Another important part of the data record is the description of the properties being reported. This should include property name, units, definition, format, and precision. For numerical data, units should always be specified. Where the database must handle more than one set of units for a given property, the database developer must decide how to store the data and make conversions. One option is to store the data in a set of default units and use stored conversion factors and equivalence tables to convert at display time to the alternate units. This has the advantage of saving storage space, but may slow down retrieval and display. On the other hand, data may be stored in both forms, in which case conversions are made at or before data entry. In either case it is important that the conversion process preserve the precision of the original measurement. Also it is recommended that the number of significant figures and a flag that identifies the original units be stored along with the data.

Where the property is a derived result, it is important to specify the method used to determine it. For example, if the property is calculated from the slope of a two-dimensional plot, as is the case for Young's modulus and many rate parameters, specify how the slope was determined—by the chord, secant, linear fit, or "eyeball" method—and between what points on the curve the calculation was made.

Test Results

The type(s) of data to include in the database—"raw" (unanalyzed) data, calculated results, or the outcome of statistical analyses (means, allowables)—must be decided based on the purpose of the database and the type of data available. However, wherever possible, if the "raw" data are not included, sufficient information should be included to make it possible to trace them to their source. Where data are statistically analyzed, the type of analysis and statistical parameters measuring uncertainty (standard deviation, coefficient of variation, and confidence interval) should be reported along with the number of data points used for the analysis.

Test Observations—For mechanical property tests, a description of the failure mode is important to the interpretation of the results, especially for materials that exhibit a variety of types of failure behavior. In other types of tests, observations of what actually occurred during the test may be an important indicator of reliability and should be included.

For test observations and nonnumerical data, a category or value set, which is a list of acceptable entries, can be used to limit the variety of responses. For example, where a qualitative observation, such as failure location is called for, the category set might be limited to "gage section," "under grip section," and "between grip and gage section." Category sets are considered to be closed. That is, they contain all the possible acceptable entries for the given field. Value sets, on the other hand, are lists of typical or representative entries, but are not necessarily complete, leaving it again up to the developer to decide how to structure the actual database. Generally the standard format identifies which lists are category sets and which are value sets.

Validity Criteria

Appropriate indicators of data quality should be included. These may be "check offs" of items enumerated by the Test Method ("Did failure occur in the gage section, yes or no?") or the results of assessment of the data against some predefined quality criteria. (See Chapter 8 for more on this subject.) Where these criteria are not part of the Test Method they should be fully documented in the database.

PHYSICAL PROPERTY DATA RECORDING FORMATS

There are as yet no formats for specific physical property tests, so this is an area that needs further work. Many of the material identification formats include a selection of physical properties as part of the identification and, in some cases, even provide for limited metadata (for example, test method) in association with these. For example, the polymeric materials identification format includes density. This arises because materials are routinely classified based on properties as well as on composition. For the database developer this raises the interesting question of when is a property an identifier and when is it part of the data? Until specific criteria are developed, a recommended procedure is to include properties as identifiers only where nominal values, without additional documentation as to how they were measured, will suffice. Where complete documentation of the test, conditions, method, and so forth, are required, they are best treated as property data. (In some cases it may be desirable to make provision for both a nominal value of the property and the result of an actual confirmatory measurement.)

MECHANICAL PROPERTY DATA RECORDING FORMATS

In the mechanical property arena the data recording formats have been broken down by material class since in most cases there are significant differences in test methods between classes.

Metals

Annexes to ASTM E 1313

In the annexes to ASTM Guide E 1313 are formats for recording data from several mechanical property tests methods for metals. (See Tables 5.1 and 5.2.) Annex 1, the recording format for ASTM Method for Pin-Type Bearing Test of Metallic Materials (E 238) is relatively straightforward and provides a good example of the application of the generic guide. In all, 34 items of information are described with 11 identified as essential. It should be noted that there will not necessarily be a one-for-one correspondence between the items in the format and the actual fields in the database. For example, the first three items (material, lot, and data source identification) will actually be multiple fields as discussed in the material identification formats. Other items may require repeated fields. For example, if there is both an ASTM and an ISO (International Standards Organization) test method and it is desirable to store the numbers of both, two test method number fields may be needed.

Like the other formats we will examine, this one does not specify field format beyond the level of numeric/character string, nor is there a standard field name or mnemonic. It has been, unfortunately, extremely difficult to obtain consensus at this level of detail. Without this, the problem of data exchange from one system to another is still a difficult one. Also this format does not contain a number of the fields recommended by the generic guide, including precision and variance. It does however, have a number of validation fields from the reference test method. It is also a good example of the use of closed category sets and representative value sets to limit the responses for nonnumeric fields.

The second annex format, Plane Strain Fracture Toughness Test Data, is rather more complex, reflecting greater the complexity of the analysis and the larger number of input parameters required. There are, however, no significant new features in this format.

The remaining annexes to ASTM Guide E 1313 will not be discussed in detail. Some are still fairly early in the development cycle and may be subject to considerable modification before they are finally adopted. While some of the differences may be removed as they receive further editing and evaluation, it is instructive to compare them briefly in their present state. Whereas the Notched Bar Impact format, Annex 5, exemplifies the bare bones approach to metadata associated with a relatively simple, well standardized test method, the three annexes still in preparation (X20, X21, and the as yet unnumbered one for Creep, Creep Rupture and Stress Rupture) go to the other extreme, calling for not only the type of equipment used for the measurement, but also the make, model, and serial number! This level of detail is probably far greater than that required for most databases, but there may be cases where it is relevant, and it is part of the database developer's job to determine just how much detail is required to meet the present and future needs of the users.

The three draft annexes each list well over a 100 items of information. These are complex tests with many possible variables. Annex 20, Fatigue Crack Growth Rate, also has several examples of properties used as material descriptors (for example, hardness and ultimate tensile strength) with no associated metadata. In this case it might be better practice for a database to include these as material properties with complete metadata and cross reference them when the fatigue results are recorded. The creep format attempts to include two-dimensional data in what is essentially a one-dimensional data format. The latter would be easier to handle by choosing a database management system that is capable of storage and retrieval of graphical data.

Composite Materials

ASTM Guide for the Development of Standard Data Records for Computerization of Mechanical Test Data for High-Modulus Fiber-Reinforced Composite Materials (E 1434) was developed by ASTM Subcommittee E49.08 but has now been turned over to Committee D-30 on High-Modulus Fibers and Composites. The guide has some important limitations in scope:

- It is specific to tests of laminated composites and their laminae (plies).
- It covers only tension, compression, and shear tests.
- While it is intended to cover ceramic and metallic matrix composites, in addition to polymerics, all of the examples are for polymerics.

However, the format is sufficiently general that it should be easily extensible to other test methods and material forms, and wider use should be encouraged.

Testing of composite materials is not highly standardized. For example, there are several competing methods for measuring in-plane shear and compressive properties. For this reason these formats will probably not be associated with specific test methods but will likely remain grouped in an independent document. The ASTM test documents explicitly covered by this format are as follows:

(1) Tension Tests (ASTM Test Method for Tensile Properties of Fiber-Resin Composites [D 3039] and Test Method for Tensile Properties of Fiber-Reinforced Metal Matrix Composites [D 3552])

(2) Compression Tests (ASTM Test Method for Compressive Properties of Unidirectional or Crossply Fiber-Resin Composites [D 3410])

(3) In-Plane Shear Tests (ASTM Practice for In-plane Shear Stress-Strain Response of Unidirectional Reinforced Plastics [D 3518], Guide for Testing In-plane Shear Properties of Composites Laminates [D 4255], and a new method for V-notched beam specimens, now being balloted)

(4) Apparent Interlaminar Shear Test (ASTM Test Method for Apparent Interlaminar Shear Strength of Parallel Fiber Composites by Short Beam Method [D 2344])

Most of the composite test methods are relatively nonrestrictive. Parameters like specimen dimensions, for example, may not be completely specified by the test method. Thus this information must be included in the metadata. Because of this, the formats contain a relatively large number of fields.

This document illustrates the use of the conditionally required field, that is, a field that may be required, depending on the content of another field. Composite materials properties are frequently reported normalized to a common fiber volume percent. When the "normalization" field indicates that normalization by fiber volume has been performed, the field for "normalized fiber volume percent" must also be filled for the data to be complete.

This format also provides explicitly for reporting of either the results of measurements on individual specimens or statistically derived values (mean, standard deviation, and so forth) for an ensemble or set of nominally identical specimens. Either or both can be included, depending on the requirements of the database application.

The format calls for eight classes of information (compare to the discussion of ASTM Guide E 1313, above):

- Material identifiers
- Test procedure description
- Individual specimen description

- Individual specimen test parameters
- Individual specimens test results and analysis
- Ensemble (set of identical test specimens) description
- Ensemble test parameters
- Ensemble test results and analysis

Material Identifiers

As in the other formats, this one has some fields in common with the material identification format. It also includes information on preconditioning of the specimens before testing and results of nondestructive inspection (another example of a test result being used as a material descriptor). The latter is included because some form of nondestructive evaluation (NDE) is commonly performed as a quality control procedure on composite test specimens. More detailed formats for computerization of NDE results will be discussed later.

Test Procedure Description

The procedure description includes documentation not only of test method, geometry, and equipment but detailed records of by whom, when, and where the testing was performed. This may appear to be excessive, but in many of the industries using advanced composite materials, total traceability of the data is a requirement.

Individual Specimen Description

This section provides for recording each critical geometric dimension of the specimen. Also included is a field for specimen cross-sectional area used for stress calculations. Since this value is derived from the other dimensions, the database designer may decide not to include it in the database, preferring to calculate it from the stored data when required.

Individual Specimen Test Parameters

The only test parameters listed for the individual specimens are date, temperature, and humidity since other testing parameters (for example, strain rate) were included with the test procedure and, by inference, are only recorded once for the entire set of specimens. If, in practice, a parameter varies with each individual specimen (as, for example, specimen temperature or span-to-depth ratio) it should be recorded for each specimen.

We have already discussed the treatment of physical properties as descriptive characteristics. For composites it is important that the test procedures used to measure fiber volume and other physical properties be recorded; so they are best treated as physical property data and linked to the mechanical property data on the same material.

Ensemble Description, Test Parameters, and Results

These parallel the individual specimen information except that the average values and standard deviations for the entire set of specimens are recorded here.

Polymeric Materials

Presently, the only data recording formats under development for polymeric materials are those for tension tests, ASTM Standard Test Methods D 638 and D 638M. Additional formats are expected to be developed with time.

CORROSION, EROSION, AND WEAR DATA RECORDING FORMATS

ASTM Standard Guide for Formats for Collection and Compilation of Corrosion Data for Metals for Computerized Database Input (G 107)

Developed by ASTM Committee G-1 on Corrosion of Metals, in conjunction with Committee E-49, this guide provides formats for recording of data on corrosion of metals, excluding electrochemical corrosion. Because corrosion tests are conducted in a variety of ways with a variety of objectives, the format is necessarily lengthy and complicated. A comparison of this standard with the others we have discussed also reveals an interesting difference in emphasis. Here the primary focus is on the experiment or test, and not the material. Hence the test description occurs first and occupies a much more prominent position in this format. The drafters of this standard were more concerned with the methodology and interpretation of corrosion tests than with the materials themselves. They have identified nine categories of information to be recorded, which reflect this emphasis:

- Test identification—code or identifier associated with a specific set of specimens exposed to the same environment at the same time
- Test Type—standard, laboratory, or field test
- Test Emphasis—specific form of damage tested for, for example, corrosion, pitting
- Test Environment—chemical description of the environment
- Exposure Conditions—time, temperature, pH, and so forth
- Material Identification
- Specimen Identification
- Performance Data (results)
- Data Source or Reference

The difference between this format and the other formats is more one of form than content, since basically the same types of information are recorded, although the organization is different. This points up one of the challenges the database designer must face: different disciplines need to view the same data from different perspectives. A good data management system must be able to accommodate this.

ASTM Guide for Recommended Data Format for Wear Test Data Suitable for Databases

This format, being written in conjunction with ASTM Committee G-2 on Wear and Erosion, is still in draft form and needs considerable work. Presently it consists only of a list of essential and desirable fields without definition or discussion. The format appears to be aimed at recording of data from A-B comparison tests where two materials are evaluated side by side for their resistance to wear or erosion, but no related Standard Test Methods are cited.

NDE DATA RECORDING FORMATS

Dealing with the results of nondestructive evaluation in a computerized database presents some unique challenges because such data are typically in the form of images (X-rays, CT scans, ultrasonic images) and not simple one or two-dimensional numerical arrays. Unless the database is to be limited to recording the conclusions drawn from the evaluation (pass/fail or text description of observations), provision has to be made for storing and reconstructing images themselves. Unfortunately, the two existing draft guides for NDE data, ASTM Standard Guide for Recommended Standard Data Records for Computerized Storage and Transfer of Digital Ultrasonic Test Data and the sister document for Computerized Storage and Transfer of Digital Radiological Test Data do not provide a standard format for handling the image data themselves. Rather, they specify what accompanying identification and metadata should be stored with the data. Part of this information is a description of the format of the actual digital data file. Included in the format is the following:

- Traceability (header) Information
- Test Description
- Equipment Description
- Sample Description
- Coordinate System and Scan Description
- Test Parameters
- Data File Description
- Data File

Like the corrosion format discussed earlier, the focus here is on the test and methodology rather than the material. (This format is applicable to NDE results on any class of materials.) The equipment and procedures are specified in detail, while the sample gets only a minimal description.

The traceability information includes when, where, and by whom the test was performed along with reference to applicable standards or test methods. This kind of concern for traceability is common in industries where NDE is used for quality control of critical hardware.

Similarly, the equipment description calls for the exact type, manufacturer, and serial number of the equipment used as well as the values of setup parameters used to run the test. Because NDE procedures must be adapted to the requirements of the particular specimen, it is important that these details be recorded here.

The image file description supplies the information necessary to translate the numerical values stored in the file into meaningful information about the specimen giving, for example, the resolution and range of the data stored, along with the number of points and the physical format in which they are stored. From this information it should be possible to take the image data file and reconstruct the original image produced by the inspection.

OTHER DATA RECORDING FORMATS

This section discusses two additional draft formats for recording of data on specialized materials that do not fit the categories already covered: arc welds and high explosives. Both of these standards differ from those already discussed in one important aspect: they are not limited to the results of a specific type of test but attempt to include results from the whole range of characterization tests that might be performed on the subject materials.

Arc Welds

The Standard Guide for Recording Arc Weld Material Property Data in Computerized Databases was developed by the American Welding Society in cooperation with ASTM Committee E-49. The final standard has been issued and will be maintained by the AWS as Standard A9.2 along with the standards for identification and description of arc welds. It is intended to describe properties of the weld structure itself, not the properties of the parent materials or any filler material (although these might be described elsewhere in the database).

Unlike the other standards we have discussed, the primary purpose here is to record the information necessary to substantiate that the weldment under test meets the requirements of certain trade specifications, and as such, its scope is rather limited. In several cases the only result reported is a pass/fail flag based on the requirements of the cited specification. While for specialized application this may be adequate, actual numeric results, along with the specification maximum and minimum values for each measured property, should be included in the database for more general applications. (In the case of a post failure investigation, it might be useful to know if "pass" means "by the skin of the teeth" or "with a healthy margin.")

The format provides for recording the results of several types of mechanical and physical property tests, with minimal metadata. Also included are formats for results of microscopic and NDE examination (with only pass/fail results). For a more complete format for NDE results, see the previous section.

High Explosives

ASTM Guide for Recording Property Data of High Explosives in Computerized Databases, which is still in the very early draft stages, provides for recording a broad range of physical, thermal, mechanical, and explosive properties of a very specialized class of materials. The properties section of the format is still incomplete, consisting only of a list of properties without definitions, formats, or metadata specifications.

CONCLUSIONS

Data recording formats for most materials except metals and some composite materials are still very much in the developmental stages. However, the generic standard guide appears to have successfully established a pattern for the development of guides for specific material types and tests. There is clearly a great deal of work to be done even considering only those tests that have been successfully standardized through the efforts of ASTM and other organizations. While Committee E-49 has been leading the effort, it is anticipated that increasingly the technical committees having jurisdiction over the test methods will themselves recognize the value of standard reporting formats and assume responsibility for their preparation as a routine part of updating existing test methods or creating new ones.

While the principle driving force behind this effort has been the move to computerized data management, standardization of data reporting formats will have other benefits such as improving the traceability and useful lifetime of properties data as well as making it easier to compare data from different sources and arrive at meaningful conclusions.

6

Data Evaluation, Validation, and Quality

Anthony J. Barrett[1]

INTRODUCTION

Background

The term "data" means scientific or technical measurements, values, or facts, which can be represented by numbers or in other ways. These values or facts form the basis of calculations and of technical or scientific decisions. To be more specific, materials data should accurately represent such values as ultimate strength in tension, stiffness in torsion, hardness, and up to 50 or more similar mechanical properties in addition to numerous electrical, electronic, thermal, and other physical properties, corrosion, oxidation, and processing characteristics. Each data value corresponds to a particular loading condition, time of application, and set of values for each of several possible ambient environmental and other parameters. First, however, the material itself needs to be identified unambiguously in terms of its chemical composition, form, heat and other treatment, source, and more. Finally, beyond the mechanical and physical properties, data may be required on the cost of a target material, on its availability, its environmental impacts, and even, perhaps, its aesthetic qualities. Most of these data are numeric values, but they do not stand alone. The mechanical properties, for example, are dependent upon the test methods or predictive mathematical models that were used to determine them.

So the totality of materials data, including information about those data, or metadata, may be likened to a complex, multi-dimensional universe of "cells" containing data values, each one relating to one set of the various parameters described in the previous paragraph. The different ways of conveniently using these data manually, in calculations or as the basis for reaching decisions, might each require a different optimum arrangement of the cells. Contemplate, for example, the different ways in which the data need to be organized if one is searching for a particular property of a known material as distinct from looking for a material with a given, or minimum value, of some property. It is, of course, because of the complex nature of the materials data universe that capture and subsequent manipulation of all or part of it by computer is so attractive. There is good news and bad news for those who do this.

The good news is that the storage and manipulation of all the available data in the engineering materials universe is well within the technical capability of the computer hardware that is available. The bad news is that the same parts of the universe are seen differently by different observers. Returning to our simple analogy, different sources report more or less different values for the contents of the same cells in this universe. More confusing still, each observer, who is in fact a data generator or experimenter, often reports data appropriate to values of the parameters for the tests or predictions they each make, which are different than those used by other observers. So questions arise as to which sets of data should be stored in a database, how sets at variance may be utilized, how the reliability of the data that are to be stored may be improved, and how the quality of the data finally built into the database may be controlled and assessed. The purpose of this chapter is to address these and related questions.

Terminology

A small number of special terms will be described in practical language as they are needed. A short list of more authoritative definitions is given at the end of the chapter in Appendix B.

The Nature of Raw Data

At one extreme, the database builder may have available many large sets of values of different properties, derived by tests conducted in a number of different establishments, by different methods, under different ambient conditions; such a store of values is well described by the term "raw data." The task facing the database builder is to manipulate this store into a useful and readily accessible tool, to be able to describe its contents uniquely, and to be able to indicate its reliability and other qualities. Deciding what should or should not be stored in a materials database is a task requiring careful planning. The easy solution would be to store everything reported about each material, but this creates more problems than it solves if the database is to be an effective and efficient engineering design tool. Under these circumstances storing any and every materials data value that

---

[1] Consultant, formerly, Chairman and Chief Executive of Engineering Sciences Data Unit Ltd., 6 Deanway, Chalfont St. Giles, Bucks, HP8 4JH, United Kingdom.

Copyright© 1993 by ASTM International www.astm.org

might have been determined, without any judgment being made of the quality of those data, can hardly be in the interests of economy, and such a database, in the hands of an unskilled or unexperienced user, could lead to expensive mistakes or worse. Under other circumstances, such as in materials quality control or in research, raw data may be stored in a database. It is important that in all cases a trail should be provided that concerned users may follow to find the raw data from which the values in the database were derived.

Compilations of Known Pedigree

Quite frequently the database builder will be faced with a task not as daunting as that in which a start has to be made with completely raw data. A long established compilation of data, such as that in a handbook, a set of standards, or a company design manual, may have to be put into machine readable form. This is certainly a different task to that which starts with raw data. However, the computer is a hard task master, intolerant of inconsistency and the many irrationalities that may appear in a data compilation originally constructed for manual use! Whatever the starting point in the building of a database, with raw data or with an authoritative handbook type source, the need to refine the data in some way will often become evident if only so that some minimum basis of quality may be indicated for the reassurance of the eventual user.

RAW DATA RESOURCES

Data may originate from tests, which may include service experience, and theoretical predictions. In addition to the numeric values themselves, it is necessary to obtain and record their metadata. We will start by reviewing the origins of data and some of the causes for differences between sets of data obtained under nominally identical circumstances. The first steps in constructing a database, once its scope and applications have been determined in relation to user needs, include finding suitable sources of raw data, deciding which to use, and being cautious at every stage of collecting raw data. After reviewing these steps, this section of the chapter introduces the concept of data refinement and the processes involved.

Data from Tests

When a database is being constructed from scratch, a start has to be made with a vast collection of raw material. This raw material is frequently in the form of reports originating from basic research but may also include information from applied research that has been undertaken during the development of products. These records are generally numerous and their content includes commentaries on how the data were derived. If they were experimentally derived the report will, in addition to providing numerical results, contain more or less complete descriptions of such things as the design of the experiment, apparatus used, the test procedures employed, the readings taken, corrections made to these readings together with ancillary readings of ambient conditions, and much else. In ideal circumstances standard test procedures will have been available, used, and referenced in the reports that provide the raw data; but this is very much an ideal. Because it is necessary to cast a net widely in the search for adequate quantities of raw data, results have to be used from experiments made against different national and international standards as well as results from experiments made against no set of standards whatsoever. It will be many years before the work currently in hand to harmonize these standards will have progressed to a point where they will have been in wide and regular application long enough for it to be acceptable to ignore results from tests conducted before such standards were available.

Theoretically Predicted Data

In addition to data derived from tests, other data may be available that have been derived by theoretical prediction methods. Here one should expect to see the data accompanied by an adequate description of the assumptions and idealizations that have been made in order to make the prediction method amenable to mathematical modelling. Alternatively, one must expect that a standard prediction method has been employed and to find that method clearly specified.

Data About Data

The resources of raw data from which a database is built contain a great deal in addition to the data values themselves. Indeed, it may sometimes seem to those who apply data that those who generate data, namely, the experimentalist and theoretician, are more concerned with scientific method or mathematical principle than they are with the practical value of the results obtained. The database builder must resist the temptation simply to transfer numerical values into a database and ignore the other information, metadata, which hopefully accompanies the data in their original form. In the final analysis, it is only possible to assure and certify the quality of a database if all that is known about each of the data sets included has been properly assessed.

Inconsistencies between data sets, in terms of the information that accompanies them, must be taken properly into account, and any deficiencies or omissions must be consistently recorded. For example, there has been a vast improvement, over the last two or three decades, in the practices of measuring the fatigue lives of materials specimens and components. In the early days of materials fatigue testing it was not unknown for such "details" as whether a steady load has been superimposed on the fluctuating loading to escape mention in the test report or for the frequency of application of the fluctuating loading to be left unrecorded! The first of these omissions invariably renders the results unusable, and the lack of a record of the frequency may or may not be serious depending upon the material of the test specimen. This may seem to be an incredible state of affairs, and it brings home an important point. As time passes experimental techniques develop and new parameters are identified that can seriously affect the interpretation of test results and the value of the raw data obtained.

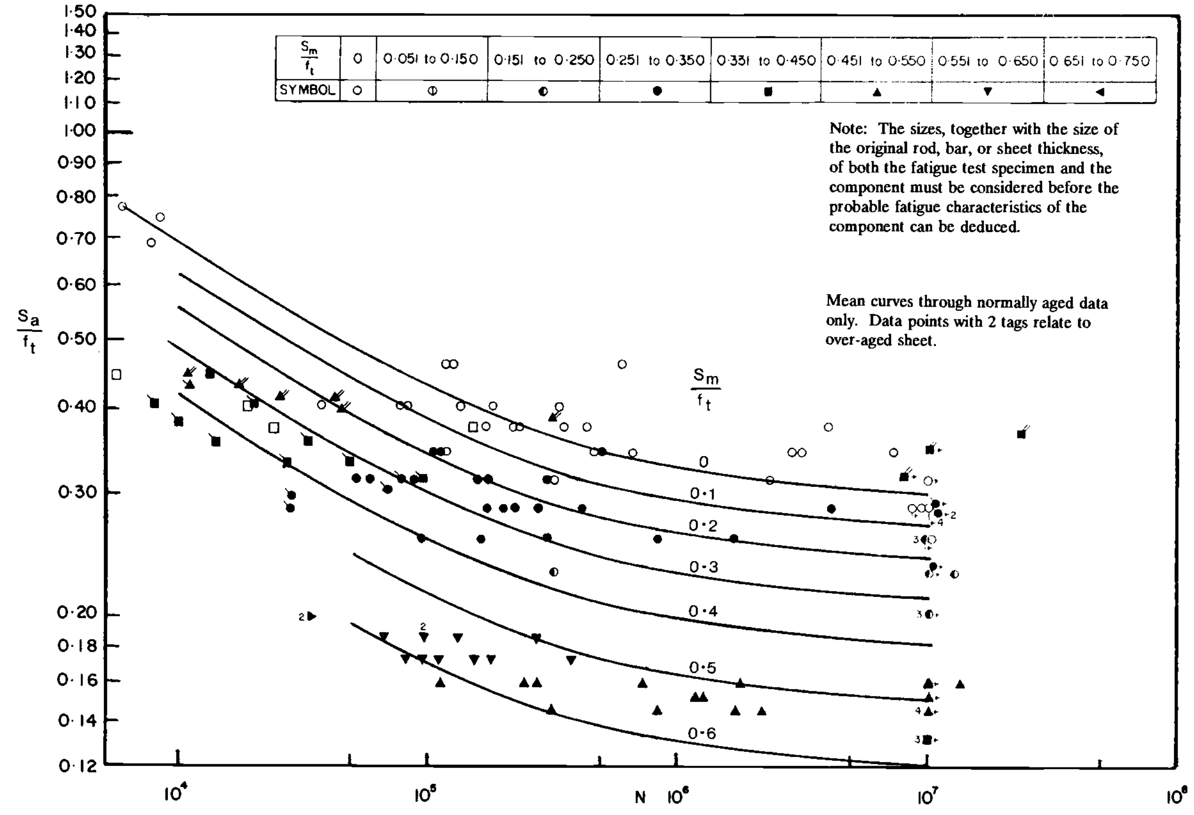


FIG. 6.1—Fatigue data for a titanium alloy (for illustration only).

Different Sets of Raw Data

The derivation of materials data from tests that have not been standardized allows much scope for "error" or, to put it at its best, inconsistency. The researcher may be unaware of these inconsistencies, and they may never come to light unless several sets of raw data on the same properties of the same material have been collected together. In general a group of such data sets will exhibit a cloud of points, when presented in the graphical form, rather than the neat relationship that might be expected or that a mathematical model might predict.

Even when different data sets have been reduced to a consistent set of conditions and characterized in terms of, say, their loading states, scatter will remain within the set and different sets will overlap. This is illustrated (Fig. 6.1) by the results of fatigue tests on a particular titanium alloy [*1*]. Values of fatigue endurance cycles N are plotted for different values of the normalized alternating stress S_a/f_t. Each set of data appropriate to a different mean stress level S_m/f_t uses a different symbol. Several small "clouds" of data may be detected. To produce the mean curves that are the resultant evaluated data from these clouds requires the application of correlation techniques, statistics, and a great deal of skill and experience. Incidentally, it can be seen (Fig. 6.1) that important metadata are provided in the two passages of notes, which also refer to further information in the original text [*1*].

Many different sets of raw data provide the material from which a database might be constructed. The point has already been made that although a database can be constructed just to store and retrieve the raw data, such a system is of little interest to those wishing to use materials property data in engineering applications. It is the experience of the author that the vast majority of the regular users of an engineering data service, with which they are familiar and which they trust, have no interest in the original raw data. Such users are actually seeking reliable data of adequate quality, which they can access quickly and easily.

Locating Sources of Data for Use in Materials Databases

The primary sources of data on engineering materials are, of course, the laboratories of materials producing companies, user companies, government laboratories and the laboratories of universities and independent organizations that are often employed by official bodies to provide data in support of national handbooks and defense and commercial programs. Directories of laboratories are generally available in most industrialized nations, and engineering librarians and information specialists are familiar with the report series that organizations, at least those in their own countries, produce.

Collections of data on engineering materials, world-wide, are fairly numerous though not always very visible. Wawrousek et al. have produced an indexed catalog [*2*] of some 1250 different sources of data on materials properties. Their description [*3*] of the philosophy and scope of that work references a number of directories to sources of materials data in North America and several countries overseas including France, Germany, Israel, Japan, and the United Kingdom. Other recently produced directories concentrate on sources in France [*4*] and on those originating, or readily available to users, in the United Kingdom [*5*]. A number of authoritative sources of materials property data in countries of the former USSR and of the onetime Council for Mutual Economic Assistance (COMECON) have been reviewed by Kozlov [*6*]. At the present time, the great majority of the sources included in these directories provide data in the traditional report or other printed forms. A directory that relates only to numeric databases in the machine readable form has been produced by the International Council for Scientific and Technical Information (ICSTI) [*7*]; it covers nine scientific disciplines. The largest group among these relates to engineering materials and consists of 70 entries.

Directories on most technical subjects become out of date in a very short time; this does not render them totally useless by any means. Generating materials data is a long-term exercise, and the interests of the specialists involved and the addresses where they are located do not change frequently. So, if one is prepared to be patient and to make a few inquiries, even a badly outdated directory can be of considerable help. Direct contact will in any case be necessary in order to obtain all the detail that eventually will be required by anyone with a serious interest in obtaining data from a particular source. Out-of-date directory information on scope, names of personnel, and all other details can be corrected when that contact is made.

Precautions when Collecting Data for Materials Databases

The location of sources, from which it is reasonable to expect to be able to obtain data of quality, is the first task that has to be undertaken once the scope and content of a new database have been decided. The task of deciding that scope in terms of the materials, specifications, forms, heat treatment conditions, and the properties, over what ranges and types of mechanical, thermal, and other loadings, and under what corrosive or other ambient environmental conditions, to be included is principally a management responsibility. The decisions made will need to take account of proven user needs and a range of what are essentially marketing factors.

The case of producing a database to cover only the contents of an already existing handbook is quite different but still requires that the scope of the database be clearly specified so that potential users may be properly informed. The more common case of using an already existing handbook compilation as one source among several requires that the scope of the projected database be specified independently of the known scope of the handbook. Under these circumstances the handbook data must be accredited according to exactly the same criteria as any other source being included.

Data Refinement

It is useful at this point to review the reasons why refinement is necessary and to introduce the processes by which this will be done. Generally speaking the resources of raw data, such as the sets of values obtained in experimental investigations, are vast and widely scattered. Often individual sets of results are presented in research reports prepared by research workers with their peers in mind; others wishing to use these data may have difficulty in understanding the language and may lack sympathy with the researcher's preoccupation with method rather than result. Sets of data from different sources, though nominally relating to the same physical phenomenon, will often be found to be inconsistent. The development with time of the precision of the research techniques available, the understanding of the parameters that may affect a particular material property, and the error and bias that the individual investigator may introduce are but a few of the reasons why inconsistencies arise. Before it may be used, this vast, untidy resource has to be extracted from whatever media it may inhabit and be refined. Refinement comprises evaluation to ensure that the data values have scientific integrity. It also includes validation where the use of data in some specified class of application has to be assured, and it may also include certification where use in a specific application has to be authorized. It also involves ensuring that the associated metadata are available and sound.

At the present time, most of the available experience of conducting evaluation and validation processes relates to the production of printed data compilations. From time to time reference will be made in what follows to some of the most well established of these printed media. They are valid exemplars since, up to the point of embodying data into a machine readable form, the same principles and practices relating to evaluation and validation continue to be relevant and applicable.

DATA EVALUATION METHODS AND PROCEDURES

Data evaluation has to do with establishing the basic scientific integrity of data. It is a process with different stages, each with its own procedures and methods. These stages include:

(1) Critical assessment of sources of data sets.
(2) Examination of individual data sets.
(3) Examination of groups of nominally compatible data sets.
(4) Dealing with gaps in the available data.
(5) Correlation, application of statistical tools, and mathematical models.

To some extent, these stages are interdependent. For example, the quality and acceptability of a data source, about which little else may be known, will become increasingly apparent as data sets that it has produced are examined. More will be revealed again as those sets are compared with nominally compatible data sets from other sources, particularly if those other sources have already established a standing in the experience of the evaluator. Gaps in the coverage provided by a source that is a generator of data, a testing establishment, for example, may tell the evaluator much about the abilities and equipment of the source; this may not necessarily be derogatory. Dealing with gaps in the available test data may often involve the use of mathematical models.

So, although in the day-to-day practice of evaluation these stages are never completely independent, nor are they necessarily carried out in the order displayed above, it is convenient to address them one by one.

Critical Assessment of Sources of Data Sets

The precepts by which sources of data may be assessed are examined following the proposals of Kaufman [*8*] with additions and a few modifications. For simplicity, and to familiarize the reader with the idea of making systematic quantitative assessments of quality, a number of check items are presented on each of which a data evaluator would make an assessment. This assessment might be in the form of a rating on a scale of, say, 1 (lowest standard) to 5 (highest standard); finer scales would not be justified. Alternatively, a more qualitative system might be preferred. On each of the items on the check list the evaluator will need to frame questions of the type illustrated and take a view based upon his or her own knowledge and experience, or that of close colleagues.

Rating the Source as an Establishment

1. Accountability—Is the source a government lab, a producer reporting on its own materials, a producer reporting on another's materials, an independent laboratory, or what? If its data were found to be defective, to whom would the source be accountable?
2. Experience—How long has the source been in the business of generating materials data?
3. Integrity and bias—Has the source a track record in these respects?
4. Equipment and calibration practices—How consistently are these reported? Would the source welcome an on-site review?
5. Management attitude to use of standards—How aware is management of relevant standards, and how supportive are they of their application within the establishment?

Rating the Personnel Associated with the Source

1. Qualifications and experience of personnel—Are these recorded, for example, in the source's annual reports? Are they known from personal contact?
2. Accessibility of personnel—Does the management encourage outside contact with its technically qualified staff? Are those staff encouraged to take part in outside activities?
3. Attitude of personnel to outside inquiries—When contact is made, what is the attitude of technically qualified staff? Do they welcome comment or are they overly defensive when criticism is offered?
4. Demarcation between duties of data generators and any data evaluators employed—Is this clear or do the same personnel generate data and evaluate them together with the data of other investigators? Has the source a history of being able to handle its own data and those of others in an objective manner?

The most telling demonstration of the quality of a source is, of course, the quality of the data that are produced over the long term. So all aspects of the products (data) of the source need to be kept under review continuously. Evaluators will normally renew their assessments of the quality of a source each time they examine a data set originating from that source.

It is not to be expected that the assessment of even the most renowned source of experimental materials property data will be found to have a perfect record in terms of satisfying these criteria to a high level. They are presented as an ideal, to give the reader an idea of the qualities to look for in the sources of raw data and, over a period of years, to make simple quantitative assessments involving these criteria and to build up a coherent perception of the abilities, strengths, and weaknesses of the sources available.

Many of the principles that apply to the search for reliable sources of raw data may also be applied to the identification of reliable databases. Accordingly, many of the tests that database builders should apply to the sources of the data they use are the same as those by which the quality of their databases will eventually be judged!

Having found sources of raw data and made some assessment of their reliability and quality, the evaluator may turn to the manipulation of the data themselves.

Examination of Individual Data Sets

In the present context, by "data set" is meant a single group of numeric data and associated metadata for one property of a single material. A simple example might be a set of related pairs of stress and strain values obtained from one run of a tension test on a coupon of a well identified material. The first stage in evaluating such a data set is to assess its quality in terms of (1) completeness of material description, (2) completeness of test method description, and (3) completeness of test data reported. Each of these headings relates to a range of attributes that the data and associated metadata in a set should cover. These attributes are listed at the end of this chapter in Appendix A in the manner prescribed by Kaufman [*8*]. The role of metadata in the design and operation of materials databases is described by Westbrook and Grattidge [*9*].

Once the provenance and pedigree of the data set have been established a preliminary, critical, visual examination should be undertaken. Graphical presentation of the data set is usually most convenient. Some of the more common telltale features that such preliminary examination may reveal are illustrated in simple terms in Fig. 6.2.

The diagnosis of the likely cause of some of the features illustrated depends upon whether the data set of Fig. 6.2 relates to a continuous test of a single specimen, as in a tension test, or to a series of tests on many specimens, one for each point on the figure, such as a series of fatigue tests. In the following diagnosis *s* relates to a data set involving single specimens and *m* applies to sets each involving multiple specimens.

Discontinuity

- ambient conditions varied (*s*,*m*)
- specimens drawn from more than one population *m*
- type, or point, of failure changed *m*

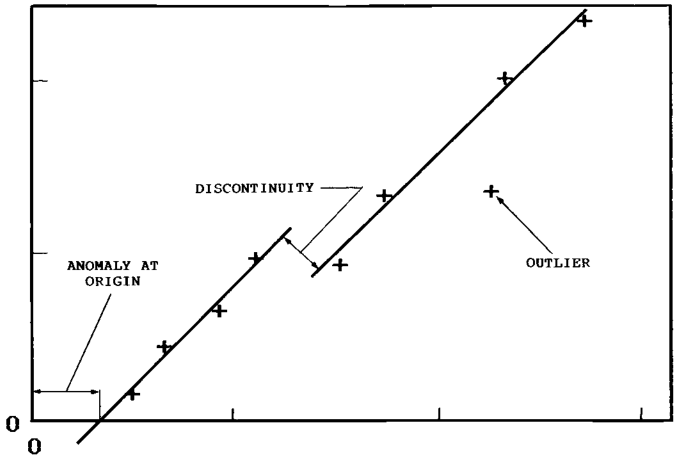


FIG. 6.2—Telltale symptoms in raw data (arbitrary axes).

Anomaly at Origin

- instrumentation errors, for example, backlash (*s*,*m*)
- systematic reading errors (*s*,*m*)

Outlying Value

- reading error (*s*,*m*)
- instrumentation error *m*
- rogue specimen *m*

Values Exceeding Known Physical Limits (not illustrated)

- instrumentation or reading errors (*s*,*m*)
- specimens drawn from more than one population *m*

Discontinuities may appear as a step change in the value and also as a step change in the gradient of the best fit to the values on either side. The outlying value featured in Fig. 6.2 calls for further investigation on two counts. Its ordinate falls well short of the best fit through its companion points and also there is an unusually long interval, on the horizontal axis, from its nearest companion to the left. This strongly suggests a misreading of at least the function represented by the abscissa.

It would be unrealistic to suppose that many of the data sets on the engineering properties of materials, which an evaluator is likely to encounter currently, will satisfy all the requirements that are specified in this chapter. Some will be deficient in the completeness of the description of the materials involved, for example, and the evaluator will have to make a decision on the extent to which missing items of information should eliminate the data set from any further consideration or limit the level of quality the data are ascribed.

In some cases, it may be reasonable to decide that corrections can be made. The best place to seek correction or the replacement of missing information is with those who generated the results in the first place, and the evaluator must make every effort to contact the source of the data set before reaching any decision alone. Of course, it may be many years since the tests on which the data are based were run, and experimental techniques may well have changed. The original apparatus may not be available for inspection, and the staff who used it may no longer be accessible for interrogation. Despite these difficulties the evaluator may still be able to apply corrections, for example, to reduce results to standard ambient conditions. In other cases it may be decided to accept results from an inadequately characterized material if those results are to be pooled with others, or if the set represents all that are available. *But under all such circumstances the outcome must be clearly recorded and qualified. Subsequent validation of what has been done must take such qualification fully into account.*

In the real world of data evaluation and validation, there will be many instances where a particular data set will be highly deficient of quality in the data values and in their metadata. These deficiencies may be beyond rectification for all sorts of reasons and almost the only virtue that the set may have is that it is the only set available that relates to a particular group of parameters. It is in relation to such data sets that evaluators will have to draw upon their experience, that of colleagues, and perhaps a number of more or less subjective conclusions in deciding how, and if at all, such a data set may be usefully and safely employed. Often the knowledge that the data are eventually to be validated for use under particular circumstances will be of some assistance in making this decision; see the later section of this chapter entitled "**Validation.**"

Once a data set has been accepted, albeit after correction and qualification of its quality, it will next be considered in relation to other data sets that may exist and apply to the same material, form, condition, and test parameters.

Examination of Groups of Nominally Compatible Data Sets

Different sets of raw data may be brought together in the evaluation process for one of a number of reasons. Part of the process of evaluating a single set may be to compare it with one or more other sets of already established integrity. Or, different sets of data may be brought together to investigate whether a test parameter, varied among them, appears to be of any real consequence. Or, different sets may be brought together with the prospect of combining them into one set that comprises more samples, covers a wider range of test parameters, or relates to a wider range of ambient conditions than would be otherwise available. At best, the careful examination [*10*] of several data sets together may add considerably to the store of knowledge on a material and give added confidence in the numerical values representing its properties. At worst, the simple "pooling" of several data sets can be imprudent, misleading, and dangerous.

The procedure with groups of nominally compatible data sets is as follows:

Separate Examination of Each Set

Follow the procedures described in the earlier sub-section of this chapter entitled "Examination of Individual Data Sets." This is to ensure that there are no "rotten apples," which might contaminate the validity of the subsequent stages.

Data Harmonization

Data harmonization is the conversion of different data sets to common terms, units, test conditions, analytical techniques, values of fundamental constants, and so forth. Any data set in the group for which this cannot be done should be held as suspect.

Intercomparison and Judgmental Evaluation

This part of the procedure can be assisted by graphical, analytical, or statistical methods.

Extraction of a Unified Evaluated Data Set

This should be done with an assessment of its quality in appropriate terms.

The last two stages are key to the examination of groups of data sets. Intercomparison may often be carried out by plotting individual values in standardized graphical forms. Authoritative descriptions of this part of the process will be found in the introductory sections of Refs *11* and *12*. Alternatively, computerized methods may be employed. These methods involve the comparison of individual data sets with mathematical models, or correlating equations, and the subsequent comparison and manipulation of the parameters of those equations. There is a wide variety of such models. Some are represented by relatively simple empirical expressions [*13,14*], such as those used to represent uniaxial stress-strain relationships for monotonic loading into the post elastic range. Others are more complex, such as those relating to the parameters involved in forming and machining processes [*15*] and those that relate to creep strain-time data.

A review of relationships for creep strain-time data, models relating to creep-rupture data, high-cycle fatigue data and fatigue crack growth data, has been made for the Versailles Project on Advanced Materials and Standards (VAMAS) by Nishijima et al. [*16*]. A most revealing study of the results obtained by different evaluators using some of these models has also been made for VAMAS [*17*]. This study concluded that the results of applying even the same model are not unique because of the different ways in which it is possible to implement it in actual practice. So there is much still to be done before even the correlation methods used in materials data evaluation can be thought of as acceptably repeatable between different evaluators.

The procedures of rectifying or explaining differences between data sets, and of untangling incompatibilities between data sets and well authenticated correlating equations, run parallel to those discussed in an earlier subsection of this chapter entitled "Examination of Individual Data Sets." They involve seeking confirmation of identical ambient conditions, points of failure, instrumentation usage and so forth. Where corrections can be applied, the evaluator should do so, keeping careful records of what was done and why it was necessary. Reference should always be made back to the originators of data sets that seem incompatible with others in the group. Those who generated the data may well not be aware of other data sets that an evaluator may have discovered, and they will usually work diligently with the evaluator to examine, explain or rectify any incompatibilities. If all else fails, the evaluator must avoid the temptation simply to pool results, take averages, and otherwise manipulate a group of data sets in which one or more sets may be suspect. Even the pooling of data sets that seem totally compatible may lead to questionable results [*18*].

Dealing with Gaps in the Available Data

A few data values or fragmentary data sets may be insufficient to produce an evaluated set covering as wide a range as might be needed of the independent variables governing a particular material property. Data synthesis from such fragmentary data may be possible. In brief, the procedure involves, first, the critical examination of each fragmented data set according to the methods already described. Assuming the fragmented data are credible and survive preliminary examination, they may then be used with well established empirical equations, theoretical methods, and semiempirical techniques that the evaluator has available and that enjoy his or her confidence. A number of these methods and techniques were referenced [*16*] in the previous subsection of this chapter. Experienced evaluators will add their own variants to the published methods and regard them very much as part of their "tools of the trade."

The objective of filling gaps is to ensure that the resulting set of recommended values is not only internally consistent but also covers as wide a range of the controlling parameters as possible. This can generally be done, to acceptable engineering accuracy, providing that the basic fragmented data are not too sparse. The VAMAS "round-robin" comparison [*17*] of data correlation methods used by some 15 experts in 5 different countries showed that, within the extremes of the data values to which a variety of correlating equations were applied, the variations between different equations and different implementations of the same equation were not extreme enough, in general, to make interpolated data values unacceptable for most engineering applications. However, among values that were extrapolated beyond the range of the experimental data the differences were unacceptably high. Accordingly, in general, the evaluator must look upon extrapolated data with extreme caution; any such data that are passed on for use in data compilations or in databases must be qualified most carefully.

Raw data on materials properties are often not available in the quantities and over the ranges that will permit evaluations to be produced and designated to high levels of quality. The cost of evaluated data sets is conditioned very much by the quality level that is required by the user and what it is reasonable to demand for use with engineering design and analytical processes. These processes may themselves be surrounded by considerable uncertainties and vague suppositions. Quality indicators in relation to the manner in which data are to be applied is a concern of the process of validation and will be described later. Methods of achieving quality, at whatever level, are inherent to the process of evaluation. Statistical analysis is one of the most powerful tools involved.

Statistical Tools, Quality Indicators, and Correlation

Experimentation in the engineering setting can be very expensive. Materials specimens, the maintenance of ambient conditions, and testing under static or fluctuating loads over possibly long periods are costly. This may limit the number of tests, for example those relating to creep or fatigue, to what in statistical terms may be very small samples. The statistical techniques used in analyzing such data need careful definition, and the evaluator needs access to them in terms that will be familiar and readily comprehensible [*12,19*].

Where there is a need for consistent interpretation and application of materials data under legislative or contractual circumstances, validated sets, such as those in officially sponsored compilations [*11,12*] and their machine readable forms, will be qualified according to broad groupings based on criteria expressed in terms of probability and confidence levels. Where such qualifications are not given, other less specific valuations may be indicated. For example, ASTM Guide for Formatting and Use of Material and Chemical Property Data and Database Quality Indicators (E 1484) ascribes the following codes for the quality descriptor that indicates the statistical basis or other foundation of a database value or set:

A = 95% confidence that 99% of values will equal or exceed (the database) value,
B = 95% confidence that 90% of values will equal or exceed (the database) value,
S = specification limit values,
D = combination of A, B, and S values,
C = other statistical basis,
T = theoretical value,
P = predicted value,
I = interpolated value,
E = extrapolated value,
M = mean or average values,
N = nominal or typical values,
U = unprocessed single-point test values, raw data, and
X = unknown.

The process of determining how the groupings should be specified is part of validation; see the later section of this chapter entitled "Validation." The process of determining the allocation of experimental results among such groups is primarily a task that is part of evaluation. As part of the process of evaluation, sets of data may be correlated to derive relationships that are then used to calculate data values to one of the qualities listed above. It is common practice to express these relationships as "best-fit" lines. The form of the line and the concept of what constitutes "best fit" are somewhat arbitrary; indeed, for most practical cases, perfect correlation between point values and a line may be obtained if that line is described by a polynomial of sufficiently high degree. But it does not follow that the coefficients of that polynomial would be of any physical significance!

The formalized methods of correlation have the important advantage of reproducibility though it must be remembered that they have no greater inherent legitimacy than "best-fit" lines drawn under the eye of an experienced evaluator. The formal methods also have the advantage that they may be highly automated, though it is prudent to do this only if there is every opportunity for human surveillance and intervention. The supervision of correlation, and of many other procedures in the evaluation process, will continue to depend upon subjective human influence. This will not be replaced until effective and reliable expert systems become available that are appropriate to the field of materials property data manipulation. Similar subjective influences have an even greater role to play in the companion process of validation, to which the reader's attention is now turned.

VALIDATION

Validation has to do with authenticating the soundness and defensibility of data and very often also with approving them as "validated data" for application under specified circumstances. Beyond this there may be legal implications regarding the acceptability of compilations or databases of such data when they are to be used under particular contractual conditions. Such validated data may be classed as "certified." At the other extreme of the scale of quality, there may be data that are simply offered as appropriate, say, for use in education, in preliminary design, for the purposes of comparative study, or for the initial selection of materials; these should be described as being of "limited validity."

This section of the chapter deals with the process of validation in general terms. The terminology used has a wide international acceptance [*20*], and most of the procedures described have been extensively employed for many years. Some of the terms used to describe these procedures may not correspond exactly with those that may be in current use in particular localities. The position in relation to ASTM use of the term "validation" is commented upon at the end of the chapter in Appendix B.

Users of materials data are becoming increasingly sensitive to the need for quality in data as they are also to the costs of providing that quality. This could in turn encourage the dubious practice of attempting to select data of a quality, and therefore cost, just adequate to suit the job in hand. Under such circumstances it will be particularly necessary that the exact status of their validation be declared and understood by all who may have access to them.

Validation May Remedy Limitations in the Evaluation Process

The skilled evaluator working on materials property data using well established standards for the characterization of materials, seeking well defined experimental methods for which there are generally accepted methods of correction, and which are known to be repeatable within well accepted limits, might at first sight be thought to be capable of fulfilling the requirements of the process of validation. In fact, such a person would be working to validate standards predetermined by whatever group of experts provided the evidence behind the qualifications "well established, generally accepted, known to be repeatable," and so forth. But such a lone evaluator has the severe handicap of not being likely to recognize personal errors or bias to which he or she might have succumbed. Evaluators, in common with data generators, have no special immunity to these limitations, and it is one of the purposes of the validation process to eliminate, as far as is humanly possible, all residual weaknesses in the data before they are committed to a database.

Validation as a Group Activity

Whereas evaluation is a process that usually involves only one person, validation is essentially a group activity. Judgments have to be made that will be defensible among a wide population of users of the data and among those who may in any way be affected by products designed and constructed in materials to which the data relate. Validation must also lead to a common interpretation of the data within a specified range of applications and under specified circumstances. It is in the interests of efficiency, economy, and safety that government contractors and partners in collaborative industrial enterprises should share the same validated data resources. Where this is not the case, incompatibilities between the materials selections and the property values adopted in different establishments may prove costly. Such inconsistencies may hinder the auditing of failures in teamed projects.

In some cases engineering companies and their subcontractors may have no choice in the matter because they are constrained to use particular evaluated and validated compilations. For example, the *Metallic Materials Data Handbook* [*12*] automatically becomes a contract document when the contract calls for the Design Requirements of the U.K. Ministry of Defence to be applied. Similarly, most of the data contained in Mil-Hdbk-5 [*11*] are formally recognized as being acceptable to the U.S. Department of Defense (DoD) and the Federal Aviation Agency (FAA). Such regulations have an important long-term influence on the standards of quality that are expected of data, and this expectation eventually extends well beyond the contractors whom they were originally intended to control. A migration of this effect to involve the computerized versions of such officially sponsored data compilations may be expected though probably not for a considerable time. A survey [*21*] of users of Mil-Hdbk-5 revealed a somewhat ambivalent attitude in the aerospace industry itself towards the suggestion that Military and Federal Aviation Agency regulations should be changed in order to encourage the use of a computerized version. Less than 20% of respondents were prepared to contemplate the phasing out of the printed handbook at any time in the future. Since that survey, data from the Handbook have become available on the Material Property Data Network, but it is too early to judge whether there has yet been a change in the basic attitudes of industry.

Validation is a matter of concern not only to users of materials databases but also to those who create them. Offering a database may expose its producers and distributors to litigation. It may be important, in the defense of a liability suit, to be able to produce evidence that data alleged to be faulty in some respect have been validated under the scrutiny of a group of recognized, independent experts. So the methods used and the management of the people employed in the process of validation are of direct concern not only to users but to all others involved with materials databases. In these respects quality is not just something to be provided in response to a market demand for it is also needed to help to protect the interests of those serving that market, whether commercially or otherwise.

Methodology

Quite apart from their role as quality assessors, groups of independent experts engaged in the validation process can make a substantial contribution to the quality of the data themselves; this point and the choice of members of these groups will be discussed after first examining the things that they are expected to do.

The validation group has a number of duties including responsibility for ensuring the following:

(1) The processes of evaluation have been properly applied.
(2) All known sources of relevant raw data have been included.
(3) The reasons for discarding any particular data are sound.
(4) Any outliers have been properly accounted for.
(5) Data needing statistical qualification have been identified.
(6) Statistical analysis has been properly applied.
(7) Probability and confidence levels used are those relating to current practice.
(8) Limitations to the applicability of the data are declared.
(9) The presentation of data is clear, convenient and appropriate to the abilities and needs of the intended user.

It is in relation to decisions relating to items towards the bottom of the above list that skill and experience in making qualitative judgments, as much as an ability to be strictly objective, will be exercised. Subjective influences figure less prominently in the evaluation process though they are never totally absent. Exposure of any subjective bias that may have come into play during the evaluation process and examination to assess its rationality are also among the tasks of the validation team.

Management

Example Organizations

The basic objectives of validation are achieved in some of the world's leading materials data compilations by organizations suited to local circumstances. Most of these organizations have yet to tailor and apply their methods to the validation of the machine readable database format. However, it seems likely that the validation of data compilations up to the point of their embodiment in machine readable form will continue to be achieved by much the same groups and the same methods that have been successful over many years with the hardcopy (printed) form. A few examples of how validation is managed for some high-quality materials data compilations will illustrate that there is some room for variation.

In the United States, administrative and technical management of Mil-Hdbk-5 [*11*] is the responsibility of government agencies working in conjunction with a subcommittee of the Federal Aircraft Design Criteria Committee. This arrangement carries the data through to include certification (see later section of this chapter entitled "Certification"). Similarly, in the United Kingdom, data in the *Metallic Materials Data Handbook* [*12*] are validated and also certified by a panel approved and sponsored by the Aerospace Industry, the Ministry of Defence, and the Civil Aviation Authority. It comprises experts drawn from industry, Government Defence and Civil Aviation agencies, Government research establishments, as well as two independent members.

In the international setting, the former COMECON countries and the USSR employed a meticulous procedure [*6*] for the validation of materials and substances databases and their certification for inclusion in the COMECON Standard Reference Data System (SRD). Under the supervision of a working group that met twice a year, a candidate database developed in one of the participating countries would be demonstrated at an international seminar. Then, for an experimental period of 12 to 18 months, the database would be used to answer requests for data, free of charge, from organizations in the COMECON countries. In addition to assisting the validation and certification processes, this arrangement revealed potential users of the database, and tested the complete database system, its field performance, serviceability, and general user friendliness. At the end of the experimental period the working group would discuss the results and make final recommendations to the COMECON Commission on Standardization regarding the inclusion of the database in the SRD.

In the United Kingdom the Engineering Sciences Data Unit (ESDU International) and its forbears have employed refined procedures for the validation of data for over 50 years. The general engineering properties [*12*] and fatigue properties of materials [*22*] are presently included in this service. The organization and management of the ESDU validating groups, which work over a wide range of different engineering disciplines, are discussed in Ref *23*. Perhaps the most important requirement identified is that these groups should be managed as consensus seeking bodies, and be carefully constituted, if they are to be of maximum efficacy.

A Consensus Seeking Group Model

The term, "consensus," is here used in its original sense, meaning, unanimous agreement on all matters of substance. A model of a true consensus process has been set up on the basis of a number of simplifying assumptions concerning the way in which the group applying the process is managed and the absence of any personality differences among its members. This model illustrates Ref *23*, for example, that the reliability of an individual expert's judgment may be magnified by a factor of 1000 when applied in a group of six such experts. There is, of course, a price to be paid; it is a price in the time taken by the group to reach a decision. The group will reach decisions at one quarter the rate achieved by the individual; even so, what other management system offers such good value?

The time taken to reach consensus is increased for decision tasks in which the reliability level of each individual participant is lower than that for the tasks with which they may be more familiar. Thus, for example, agreement on validation questions concerning any data sets that are towards the boundaries of an expert group's experience takes much longer to obtain than on questions that involve more familiar sets. Another effect that this simplified model exposes is that the reliability of decisions taken by such a group is disproportionately weakened by the inclusion in the group of any individual whose professional credibility is suspect because, say, of inexperience, extreme bias, or other causes. One might well expect such a result. The model also illustrates that the inclusion of but one such biased member can extend the time taken to reach decisions to an alarming degree. This too coincides with the general experience of those who have spent much time in any sort of committee situation, consensus seeking or not. The phenomena that this simple model illustrates are of a type and scale entirely consistent with the author's experience in working with true consensus seeking groups.

In summary, the validation process using groups of experts will achieve high reliability in the judgments that are made when those groups are managed so as to:

(1) Base decisions on a true consensus among their members.
(2) Include only members of as uniformly high a judgmental reliability, experience, and ability as possible.
(3) Avoid inclusion of any member of suspect reliability, of limited experience or known to be motivated by bias.
(4) Work well within the limits of their knowledge and experience.

In addition, these groups need to be supported by a well trained, competent technical staff who can present the issues on which decisions have to be taken, clearly and as unambiguously as possible. That staff function also embraces the exposition of technical issues in different ways until consensus is achieved. High caliber validation processes, like those for evaluation, require substantial investments of time and money. The resultant quality of the data sets and systems that are achieved by these processes can comfortably justify the investment made in them where they offer sufficient benefits in accordance with market needs and where those benefits can be clearly identified and demonstrated. Such an investment may also have to be made if there is the possibility that the producer or distributor of a database or system may be exposed to claims of legal liability. Suitable markets depend upon the existence of a potential customer population that is involved with the choice and application of materials in critical design situations and that is aware of, or can be educated to appreciate, the dependence of the quality of a product upon the quality of all things, including the data, that are used in connection with its design, production, and service life. Further discussion of costs, value, and benefits will be found in Refs *23* and *24*.

CERTIFICATION

Data values or data sets that have been recognized by a warranting authority are known as certified data [*20*]. As was suggested in the introductory paragraph of the previous section, certification implies the acceptance by some controlling authority of a data set, compilation, or system for specified applications for which that authority is in some way responsible. That responsibility may be for regulation, as in the case of civil aviation, or for funding, as in the case of defense contracting, or for prescribing safe design procedures, as in the case of pressure vessel codes, and so on. There is clearly some scope for a certification body to select and further manipulate evaluated and validated data sets and to limit their use to prescribed circumstances. An authority responsible for the safety of a particular class of engineering component, for example, may certify only a small range of design values derived from evaluated data to which have been applied more or less arbitrary safety factors based upon favorable service experience with that class of component. Thus there are no universal procedures beyond the stage of validation for the development of certified data; each controlling authority will have its own objectives and philosophies.

Even though objectives and philosophies will vary from one controlling authority to another, there are certain items of information that must accompany any compilation of certified data. These are as follows:

(1) Name and address of controlling authority.
(2) Names of any other authorities associated with the certified data.
(3) Names and affiliations of members of committees or similar groups responsible for supervising certification.
(4) Applications and purpose(s) for which the data are certified.
(5) References to regulatory, contractual, or similar associated documents.
(6) Sources of data used.
(7) Name, address, phone, and fax numbers of technical contact.
(8) Dates of issue and latest amendments.
(9) Arrangements for updating.

DATA/RECORD/DATABASE QUALITY INDICATIONS

False impressions of the quality of a data value, or of a complete database, may be given by the way in which those data are presented to a potential user. Scruffy manuscript presentations, and slovenly printed work, are often mistrusted just as high resolution graphical presentations give the impression of being more trustworthy than those which are less precise. Both reactions may be completely unjustified. So, in this section of the chapter, before considering quality indicators as such, the need for care in choosing appropriate levels of precision and accuracy will be discussed. Circumstances where confusion in this matter is often encountered—in databases derived from handbooks and in the procedures that are sometimes offered as added value features to some computerized databases—will be considered before describing a systematic procedure for the assessment of database quality.

Matching Presentation to Reality

The ability of the computer to store, calculate, and display numerical values to extreme levels of precision is intrinsically valuable but can be meaningless within the context of materials property data. Most property data are derived from physical tests of one sort or another. The precision with which the results of those tests are recorded may be very high; 1 part in a 100 000 or more is not unusual. However, many types of property tests are unlikely to be repeatable, even on the same apparatus and with all other conditions remaining constant, so as to give results varying by less than about 1% (an accuracy of 1 part in 100). So while it may be rational to record individual results to whatever precision the instrumentation may have offered during the tests, it is not rational to deliver those results from a database for application to a design or engineering problem at a precision greater than the real accuracy that can be claimed.

Databases Derived from Handbooks

High-quality materials property handbooks quote most property values to a precision of three significant figures, restricting their accuracy to less than 0.1%. Where statistical measures are applied, the highest grade indicates values above which at least 99% of values are expected to fall with a confidence level of 95%. Representation of these values by more than three significant figures would not be justified. Yet it is not unknown for properties, based upon such handbook input, to be retrieved by computerized databases to a misleading precision of six or more significant figures!

Unit Conversions and Other Manipulations

Many materials property databases offer a choice of the system of units, for example, customary units or metric units, in which the property data shall be retrieved and displayed. The values required may be stored in the database in a system that is not what a user elects to use. In this case, a simple conversion subroutine will be applied to the values between retrieval and display. It can happen that data stored to, say, three significant figures in customary units emerge from a conversion subroutine to a meaningless precision of seven or more digits according to the precision level of the arithmetic to which the computer is set! Whenever the data held in the database have been manipulated, and always before they are presented to a VDU or printer or interfaced with a CAD/CAM program, data values should have excess digits removed.

A number of materials property and design allowable values cannot be measured by direct tests. They may have to be obtained by graphical methods, mathematical models operating on test results, or from parameter relationships. Examples of such data include the elastic moduli, offset yield or proof stress values, fatigue limits and crack propagation data, and many more. Values produced by these means are not likely to have an accuracy at all comparable to the precision with which they may emerge from the methods by which they have been calculated.

Care should be taken to display data to a precision no greater than that justified by the accuracy of the procedures used to derive them. Rarely, in the case of materials property and design data, will this exceed three significant figures. The credibility of otherwise high-quality databases and of application software is undermined by a failure to observe these simple precautions.

Quality Indicators

Quality, as applied to data records or databases, is the totality of features and characteristics that are significant in determining their capability of satisfying the user's needs. Quantifiable and acceptable criteria will have to be defined in each application area to achieve a universal assessment of quality level [*20*].

A guide for the formatting and the use of material and chemical property data and database quality indicators has recently been issued as ASTM Guide E 1484. It provides a range of quality indicators relating to the following:

Data Quality

(1) Source—handbook, government, producer, and so forth.
(2) Statistical basis—(see a previous subsection of this chapter entitled "Statistical Tools, Quality Indicators and Correlation."
(3) Material status—in production, experimental, or obsolete.
(4) Evaluation status—if evaluated and by whom.
(5) Validation status—if validated and by whom.
(6) Certification status—if certified and by whom.

Database Quality

(1) Completeness of information—material form, condition and processing.
(2) Test procedures—if standard test procedures, or not, or if derived from service experience.
(3) Support status of database—if and how supported.

ASTM Guide E 1484 includes informative guidelines on the use of the above indicators. In addition, the following criteria relating to the operation of the database itself should not be overlooked:

Database Operation

(1) Availability—unrestricted, proprietary, government classified, and so forth.
(2) Distribution medium—on-line, floppy disk, CD-ROM, and so forth.
(3) Payment arrangements—on-line charge, subscription, and so forth.
(4) Access source—details of supplier.

A Suggested Procedure

Each of the indicators, listed in the previous subsection of this chapter, such as source, statistical basis, and so on, relate to attributes of a database that potential users will scrutinize when assessing the quality of that database in relation to the extent to which it will fill their needs and suit their types of application. A simple, systematic procedure for doing this is first to make a list of indicators that are pertinent, ignoring those to which one is indifferent. For example, under the operation heading, a government contractor might well be indifferent to availability if his security status is already well established. By contrast, the circumstances under which a database might be made available could be a matter of great concern to a university researcher.

The indicators in the final list may each be regarded as a "dimension" of a multidimensional space. Although attributes such as the different types of medium of distribution are not quantifiable on a continuous scale, the different media can be grouped into those that are acceptable and those that are not; or it may even be possible to rank them in terms of their appeal to the user. Finally, a comparison is made, dimension by dimension, of the working space offered by the database and what is required to satisfy the user's ideal. Various indifference levels may be tried. Each time the number of dimensions that are satisfied by a candidate system may be counted, and the dimensions (that is, the attributes represented by the indicators) that are satisfactory may be identified.

Simple, quasi-quantifiable methods, such as the method just outlined, are generally preferable to those that build up a total score by summing numerical ratings awarded to each attribute in turn. This is because the type of method outlined gives not only a simple numerical comparison, but also retains the information about which attributes are critical. The potential user can thus immediately see the practical compromises that may have to be made when using the system and its data. A similar procedure could be adapted as an alternative to that proposed earlier in this chapter in the subsection entitled, "Critical Assessment of Sources of Data Sets."

Whatever method is used for assessing the quality of a set of data, a database, or a complete system in relation to the requirements of a user, or for comparing two competing databases, that method should be devised in association with potential users. The user also has an important role in the testing of newly constructed database systems.

TESTING

Testing is an important stage both in quality control and in quality assessment. Database systems should be subjected to phases of testing similar to those that may be familiar to the reader in relation to the testing of software:

> *Alpha Testing* is an initial functional test agenda worked through by a small in-house group after the system has been developed past the initial debugging stage.
>
> *Beta Testing* is usually a secondary level of testing, carried out by offsite users before making the system generally available or marketing it.

DATABASE MAINTENANCE

The possible deterioration of the quality of a database through the "aging" of the data it contains has to be guarded against, particularly if the data are likely to be accessed by users involved in critical design or analytical projects. Arrangements should be made, at the time the data are being built into the database in the first place, to provide an audit trail back to the origins and date of generation of each data set. All data should be subjected to regular review and updating as necessary. For the benefit of users engaged in highly critical design, such as aerospace, which is controlled by official regulations, arrangements should be made to record the contents of the entire database at regular recorded intervals for archival purposes.

SUMMARY

The chapter has presented an introduction to the concepts of refining data and improving their quality as part of the task of building materials property databases. The processes involved, together with the practical procedures and methods that can be applied, have been described, and these are themselves being refined as experience is built up in what is still a relatively new field. Among all scientific databases, the construction of those concerned with materials property data presents some of the most profound challenges. There are many reasons for this. Materials data are highly multivariable in nature. They can be obtained and derived by a number of different routes and, unlike many other numeric data, they have to interface a diversity of levels of skill among their human users. They also have to interface reliably with external computerized systems.

It has been suggested [*26*] that perhaps as much as 80% of the raw data on materials that will be available at the start of the next century are already in existence, so it is already too late to apply quality control to their derivation where this has not already taken place. What it may be possible to achieve by adding value to those data, the assurance of their quality, and the costs of so doing, are governed by this reality.

The ultimate evidence of the quality of a materials database system will be found in the quality, economy, safety, and social acceptability of the products that it plays a crucial role in helping to create and provide. Engineers and scientists who are involved with materials property database systems will be at the leading edge of information technology for some time to come. They may transfer much to other scientific and technical disciplines from the privileged viewpoint that this location offers.

APPENDIX A

The matters to be considered when assessing individual sets of data and metadata provided by a single source as outlined in this chapter are expanded here in the form prescribed by Kaufman [*8*].

1. Completeness of Description of Material

(a) producer
(b) heat/lot identification
(c) status of material (e.g. commercial or experimental)
(d) name/UNS number
(e) specification(s)
(f) condition/temper/type/grade/class
(g) product form
(h) chemical analysis
(i) any special melting practice or source material, for example, vacuum cast, powder metallurgy
(j) process/thermal history
(k) microstructure
(l) background properties (i.e. tensile properties)

2. Completeness of Description of Test Method

(a) test standard
(b) specimen type, size, shape, finish
(c) specimen location and orientation
(d) specimen relation to end use situation
(e) loading rate
(f) temperature and method of measurement
(g) environment and method of monitoring
(h) statements of precision and accuracy

3. Completeness of Reporting of Test Data

(a) report format
(b) coverage of logical variables
(c) tests run to completion/according to test plan
(d) replications
(e) documentation of units and conversions
(f) consistency of results, (see also the introductory paragraphs of the section of this chapter entitled "Data Evaluation Methods and Procedures")
(g) failure type and description

APPENDIX B

Some Authoritative Definitions Concerning Data

The following authoritative definitions of terminology should help to draw attention to the different qualities of data, how they are related, and the circumstances of their application:

Data—Scientific or technical measurements, values calculated therefrom, observations, or facts that can be represented by numbers, tables, graphs, models, text, or symbols and which are used as a basis for reasoning or further calculation. Note "data" is a plural form; "datum" is the singular.

Raw Data—Data that have not been processed or reduced from their original form; (typically) experimental test results.[2]

Evaluation (of data)—The process of establishing the accuracy and integrity of materials property data. Evaluation involves the examination and appraisal of the data presented, (typically raw data), assessment of experimental technique and associated errors, consistency checks for allowed values and units, comparison with other experimental or theoretical values, reanalysis and recalculation of derived quantities as required, selection of best values, and assignment of probable error or reliability.[2]

Validation (of data)—The process of substantiating that data have been subjected to a process of evaluation, and assuring their soundness and defensibility leading to their ratification or confirmation, or to make them legally effective and binding in some specified application(s).[2]

Quality—As applied to data records or databases, the totality of features and characteristics that are significant in

---

[2] The definition has been simplified slightly from the original version to suit the purposes of this manual.

determining the capability of satisfying the user's needs. Quantifiable and acceptable criteria will have to be defined in each application area to achieve a universal assessment of quality level.

Metadata—Data about data. Description of data in a database to provide systematic information for users, application programs, and database management software. Metadata may also be manipulated and searched.

The above definitions are taken from Ref *20*, which, for most practical purposes and in the absence of an equally authoritative and comprehensive alternative, may be regarded as a de facto international standard. It includes many terms located in glossaries produced by national and international standardization bodies, terms proposed by members of the Committee on Data for Science and Technology of the International Council of Scientific Unions (CODATA), members of ASTM Committee E49.03 on Terminology for Materials Databases, and other eminent individual experts.

It should be noted that ASTM E 1484 uses the term, "validation," in effect as no more than the substantiation that evaluation has been properly carried out. ASTM E 1484 does not recognize the concept of validation of data for application to particular purposes except by an empowered authority, as is the case in data certification. It is believed that the terminology recorded by Westbrook and Grattidge [*20*], and as described throughout this chapter of the manual, will be found to be compatible with that most generally used internationally in the actual practice of data evaluation, validation, quality control, and assessment.

REFERENCES

The references, given below, comprise a reading list to assist those who wish to follow up specific points referenced throughout the text and to augment the material provided in this chapter. Particular items, marked {S}, are standards, drafts of standards, or authoritative compilations of ancillary information.

[*1*] ESDU, "Fatigue of Wrought and Cast Ti-6Al-4V Titanium Alloy (In Bending and Under Axial Loading)," Fatigue-Endurance Data Item 77027, ESDU International plc, London, 1977, p. 31.

[*2*] Wawrousek, H., Westbrook, J. H., and Grattidge, W., "Data Sources of Mechanical and Physical Properties of Engineering Materials," *Physik Daten*, No. 30–1, Fachinformationszentrum-Karlsruhe, Germany, 1989. {S}

[*3*] Wawrousek, H., Westbrook, J. H., and Grattidge, W., "Data Sources of Mechanical and Physical Properties of Engineering Materials," *Computerization and Networking of Materials Databases: Second Volume, ASTM STP 1106*, J. G. Kaufman and J. S. Glazman, Eds., American Society for Testing and Materials, Philadelphia, 1991, pp. 142–157. {S}

[*4*] Marx, B., "Banques de Données Matériaux, Pratique des Matériaux", Chapter 1.3, *Les Référentiels Dunod*, Dunod, Paris, Sept. 1990. {S}

[*5*] Reynard, K. W., *U.K. Materials Information Sources*, 2nd ed., Design Council, London, 1992. {S}

[*6*] Kozlov, A., "Materials and Substance Data Banks in COMECON Countries and in the USSR," The Provision of Materials Property Data via Computerized Systems, Report on a CODATA Symposium held at the Siberian Division of the USSR Academy of Sciences, *CODATA Bulletin*, Vol. 22, No. 2, Hemisphere Publishing Corp., New York, 1990.

[*7*] ICSTI Numeric Data Group, *Numeric Databases—A Directory*, International Council for Scientific and Technical Information, Paris, 1991. {S}

[*8*] Kaufman, J. G., "Standards for Computerized Material Property Data—ASTM Committee E-49," *Computerization and Networking of Materials DataBases, ASTM STP 1017*, J. S. Glazman and J. R. Rumble, Jr., Eds., American Society for Testing and Materials, Philadelphia, 1989, pp. 18–19. {S}

[*9*] Westbrook, J. H. and Grattidge, W., "The Role of Metadata in the Design and Operation of a Materials Database," *Computerization and Networking of Materials Databases: Second Volume, ASTM STP 1106*, J. G. Kaufman and J. S. Glazman, Eds., American Society for Testing and Materials, Philadelphia, 1991, pp. 84–102.

[*10*] Yu, M. T., Lu, F. J., Topper, T. H., and Irving, P., "Fatigue Data Pooling and Probabilistic Design," *Computerization and Networking of Materials Databases: Second Volume, ASTM STP 1106*, J. G. Kaufman and J. S. Glazman, Eds., American Society for Testing and Materials, Philadelphia, 1991, pp. 197–213.

[*11*] DoD/FAA, "Metallic Materials and Elements for Aerospace Vehicle Structures," *Military Standardization Handbook*, Mil-Hdbk-5E, U.S. Government Printing Office, Washington, DC, 1987, updated periodically. {S}

[*12*] ESDU, *Metallic Materials Data Handbook*, Ministry of Defence Standard 00-932, ESDU International plc, London, Oct. 1990, updated periodically. {S}

[*13*] Barrett, A. J. and Michael, M. E., "Generalised Stress-Strain Data for Aluminum Alloys and Certain Other Materials," *Journal of the Royal Aeronautical Society*, Vol. 59, No. 2, Royal Aeronautical Society, London, 1955, p. 152.

[*14*] Osgood, W. R., "Stress Strain Formulas," *Journal of the Aeronautical Sciences*, Vol. 13, No. 1, Institute of the Aeronautical Sciences, New York, 1946, p. 45.

[*15*] Westbrook, J. H. et al., Eds., "Materials Data Systems for Engineering," *Proceedings of a CODATA Workshop*, Schluchsee, Fachinformationszentrum - Karlsruhe, Germany, 1986, Chapter 16.

[*16*] Monma, Y., Kanazawa, K. and Nishijima, S., "Computational Models for Creep and Fatigue Data Analysis," VAMAS Technical Report No. 7, National Research Institute for Metals, Tokyo, Japan, Oct. 1990.

[*17*] Nishijima, S., Monma, Y. and Kanazawa, K., "Significance of Data Evaluation Models in Materials Databases," VAMAS Technical Report No. 6, National Research Institute for Metals, Tokyo, Japan, Oct. 1990.

[*18*] Barrett, A. J., "On the Evaluation and Validation of Engineering Data," *Computer Handling and Dissemination of Data*, P. S. Glaeser, Ed., Elsevier Science Publishers B.V., North Holland, 1987, p. 124. {S}

[*19*] ESDU, "Statistical Analysis of Fatigue Test Results," Fatigue-Endurance Data Items 68012 - 68017, ESDU International plc, London, 1968. {S}

[*20*] Westbrook, J. H. and Grattidge, W., "A Glossary of Terms Relating to Data, Data Capture, Data Manipulation and Databases," CODATA Bulletin, Vol. 23, Nos. 1–2, Hemisphere Publishing Corp., New York, 1991. {S}

[*21*] Westbrook, J. H. and McCreight, L. R., Eds., "Computerized Aerospace Materials Data," *Proceedings of a Workshop for the Aerospace Industry*, American Institute of Aeronautics and Astronautics Inc., New York, 1987.

[*22*] ESDU, Engineering Sciences Data Items on Fatigue-Endurance Data, 8 Volumes; and on Fatigue-Fracture Mechanics, 4 Volumes, ESDU International plc, London, 1992, updated frequently. {S}

[*23*] Barrett, A. J., "The Evaluation/Validation Process—Practical

Considerations and Methodology," *Development and Use of Numerical and Factual Databases*, AGARD LS-130, NATO Advisory Group for Aerospace Research and Development, Paris, 1983. pp. 5.1–5.14. {S}

[*24*] Barrett, A. J., "The Benefits and Economic Consequences of Materials Property Databases," *Computerization and Networking of Materials Databases: Second Volume, ASTM STP 1106*, J. G. Kaufman and J. S. Glazman, Eds., American Society for Testing and Materials, Philadelphia, 1991, pp. 17–25.

[*26*] Barrett, A. J., Discussion of Swindells, N. and Heine, G., "Information Engineering: Adding Value to Data," *Computerization and Networking of Materials Databases: Third Volume, ASTM STP 1104*, Thomas I. Barry and Keith W. Reynard, Eds., American Society for Testing and Materials, Philadelphia, 1992.

Management and Operation of Database Building and Distribution Functions

7

J. G. Kaufman[1]

PURPOSE, SCOPE, AND SUMMARY

The purpose of the material in this chapter is to provide guidelines for builders, maintainers, and distributors of computerized materials properties databases for any delivery system, for example, personal computer, workstation, mainframe, or online systems. It provides some basic tenets of database management and operating philosophy. These tenets are intended to ensure the development, maintenance, and distribution of machine-readable material property databases that are easily accessed and utilized, and are responsive to users' needs and expectations with respect to quality, reliability, and degree of documentation.

The scope, recommendations and guidelines, are provided as indicated in the paragraphs that follow. The organization is such that those concerned with management will find most of what they need to know in the section on Management and Operations, and those involved in data management will find most of what they need to know in the section on Data and Data Management, and so forth. The result there is some degree of overlap in a few areas, but this is in the best interests of completeness.

- Management of Operations
    - qualifications of personnel
    - standard procedures and practices
    - auditable quality assurance programs
    - maintenance and updating
- Data and Data Management
    - completeness and documentation of data
    - consistency and quality of data
    - data evaluation practices
    - data loading and reloading
- System Capabilities and Management
    - basic system capabilities
    - basic data system content
    - the user manual
    - ease of access and use of system
    - help services for users
    - testing of data system
- Security of Data
    - intentional alteration or removal
    - unintentional alteration or removal
- Costs of Operations

The information and guidelines herein related to quality and reliability of data and databases are of interest not only to database builders, but also to distributors of databases and database systems. The overriding guideline is to build the quality into the operation itself, and the result will be improved reliability in the database.

The information and guidelines herein are equally applicable, in most cases, to databases built and distributed in all types of media, including personal computer floppy or CD-ROM disks and online systems, as well as hardcopy data sources. When certain guidelines are applicable to a specific medium, it is generally specified when it may not be obvious from the content.

While not always apparent, the guidelines are often directed at the search and retrieval software instead of or in addition to the data themselves. Much of user acceptance of a data system is related to the host software as well as to the data.

Assessments of strength or weakness in several of the aspects of database operations covered by these guidelines are in some cases relatively subjective. In certain areas, it is impossible to state specific practices to be universally applied, though even in such cases, some general guidelines are of value.

In developing these guidelines, the recommendations of a variety of international groups and services have been consolidated and and in some cases amplified [*1–14*]. It is appropriate to note that the guidelines herein are evolutionary in nature, because of the relative newness of highly focused attention to computerization of detailed numeric properties data. More specific guidelines may be available in the future.

MANAGEMENT OF OPERATIONS

Qualifications of Personnel

It is essential to have experienced and knowledgable personnel involved in the process of locating, assembling, evaluating, and inputting data, as well as in the operation of a

---

[1] President, The National Materials Property Data Network, Inc., P.O. Box 02224, Columbus, OH 43202.

Copyright© 1993 by ASTM International www.astm.org

well-supported data source. While this aspect is among those to which largely subjective judgments must be employed in lieu of more stringent specifications, it is necessary that at least two areas of knowledge and experience be represented: the area of technical/scientific knowledge encompassed in the database and the area of the building, maintenance, and operation of an electronic (machine-readable) database.

It is inappropriate to specify limiting measures of the knowledge and experience required, and it is recognized that not every staff member can have extensive background in the areas noted above. However, it is clear that at least one engineer or scientist working on any specific database should be knowledgable in the technical field(s) involved, and at least one staff member should have some database management training or experience, or both.

Standard Procedures and Practices

A reliable database builder, maintainer, or supplier, or all of these, should have a regular set of procedures for all or most elements of their activity, as illustrated subsequently. These are, in effect, standard practices for the respective operations, ensuring a consistent quality to those operations. Further, the standard practices should be such that they assure the maintenance of high quality and reliability, both of the content of the database and of the service to the user of that database.

Procedures for collecting and for assessing quality and reliability should be consistent, unbiased, and carried out on bases that the user will understand and accept as reasonable and reliable.

Terminology, including nomenclature, abbreviations, symbols, units, and acronyms should be clearly defined and presented. It should reasonably conform to industry standards and expectations, and be used in a consistent manner.

Interfaces should be developed carefully and consistently to ensure user understanding and high standards of readability and clarity.

Attention should be given to ensuring that unreasonable precision is not suggested by extra, apparently significant figures resulting from unit conversions and other transformations of the data, or display programs.

For online systems, telecommunication links and the associated access and logon procedures should be based upon current industry practices, and should be consistently and easily carried out.

Auditing Quality Assurance Programs

The practices employed for assurance of quality regarding operation and consistency and quality of data should be auditable (see section on Security). Sufficient records should be kept documenting what, when, and by whom checks and corrective actions are taken so that it is possible for a person not directly involved in the process to ascertain that well-defined quality assurance practices are being utilized on a regular basis.

Among the most serious of complications facing users of database management systems are those associated with hardware and software changes. Several steps are recommended to assure maximum user understanding, acceptance, and the ability to deal smoothly with such changes.

(1) Provide ample warning in advance of the timing and nature of forthcoming changes in hardware or software.
(2) Provide provisions for holders of earlier versions of software packages to obtain updates quickly and easily.
(3) Make every effort to keep current in developing system enhancements to ensure a minimum of obsolescence for users of interactive databases and related software and hardware.

Maintenance and Updating

Employ regular maintenance operations, and provide adequate information and guidelines to users of your systems about their own needs for system maintenance, if any.

Provide for regular and easily used procedures for obtaining user feedback, and for assuring adequate attention to the input provided. Mechanisms for feedback may include mail or phone contacts, help lines, such as those discussed in the section on Help Services for User or, in the case of online systems, electronic mail direct from the service. All such input should feed through a single service group capable of relating it to feedback from others (perhaps indicative of trends or basic system weaknesses) and to recall database/system capability.

DATA AND DATA MANAGEMENT

Type or Statistical Significance of Numeric Values

There should be unambiguous descriptors of the type of numeric values given in a database, that is, whether they are individual test results, averages of several values, statistically derived values or values specified as minimum by some certification organization. When they are statistically derived, the specific statistical definition should be cited.

Completeness of Data

Consistent attention should be given to the completeness of the documentation of numeric data so that the utility and reliability of the data to the user are maximized. In general, it is appropriate to utilize standard data formats to ensure that there is sufficient and consistent background information on the material itself, the test methods, and the properties. These elements of information, though secondary in many respects to the specific performance data required by any given user, are essential to comparisons of data from various sources. See ASTM Standards E1308, E1309, E1338, and E1313 for more specific guidance on these points.

Material Descriptions

Complete descriptions of the materials for which detailed performance properties are given should also be included in the database. While the specific information needed will vary depending upon the application area, there is generally some minimum essential set of facts required to ensure adequate comparability. But for the broadest applicability additional detail may be highly desirable. In general, factors, such as product form, size and condition, must be considered when attempting to compare data from different sources and must be retained in a useful database.

Test Methods

It is essential to provide test method descriptors, notably standard test method references where applicable, plus such information as specimen type, size, and direction (orientation), and loading rate. Such information is even more important when dealing with data developed with nonstandard methods. The availability of test method documentation enables users to judge the comparability of data generated by different laboratories. Factors used in the reduction and analysis of data, such as percent offset for yield strength measurements and gage length for elongation measurement in tension tests, are also important for comparisons of data from different countries.

In the event that key variables, such as temperature, are known but not stated explicitly in a database as originally produced (for example, all data are for "room temperature"), consideration should be given to their addition before the database is widely distributed.

Test Conditions

Independent variables, such as temperature, humidity, environment, and exposure times, should always be specified, even though they may not have been varied within the specific set of data being compiled; any or all may differ when one dataset is compared with another.

Test Results

All elements of the individual test results should be included in any database that is to be a source of "raw" test data (that is, the original individual test results as contrasted to calculated average or statistically derived values).

Validity Criteria

Some test results and calculated properties have specific criteria associated with them, which provide essential information as to whether or not they were correctly generated or derived. In all such cases, the validity criteria and the original data upon which the judgment of whether or not the criteria were met should be included in the database whenever possible. For example, the reporting of plane-strain fracture toughness data requires at minimum a descriptor indicating whether or not the requirements of ASTM Test Method for Plane-Strain Fracture Toughness of Metallic Materials (E 399) are met, and preferably would contain the specific results of the measurements required to deal with those requirements.

Consistency and Quality of Data

Specific practices for ensuring the quality and reliability of data and of service to users should be defined and consistently utilized. These will constitute a quality assurance program for both the database content and the operational aspects of the service provided to users of the system.

Practices should be defined for the identification of candidate types and sources of data for inclusion in any new database, and for the final selection of the sources.

Practices should be defined for the assembly and proof checking of data to assure that the best sources are identified, that all of the essential support information is included, and that the values are correctly transcribed from the original source to the new machine-readable database. In the latter case, computerized auditing and verification practices that automatically reject input in inappropriate forms, units, or numbers of decimal places are highly recommended.

Practices should also be defined for correcting any and all types of errors found in the database, and recording the nature of the correction, for example, by whom, when, and what was changed.

Practices should be defined for assessing, on a continuous basis, the level of user satisfaction with the content and the capability of the service provided, and for reacting to input from the users with regard to improved services and capabilities.

Data Evaluation Practices

Standard procedures and practices should be established and rigorously employed in the evaluation of data or data sets for user consumption, or both.

Evaluation refers to the processes of quantitative or qualitative review of individual test results or groups of test results (datasets) by acknowledged experts in the appropriate discipline. The object of the review is to determine the reasonableness, relation to theory, or relation to other datasets, or all of these. Another purpose may be to analyze the datasets to calculate average or statistically defined properties, parametric representations, or mathematical distributions.

Consistent practices should exist for dealing with the evaluation of the content and attributes of candidate databases for a network, as well as for the review of additional data to be added to existing databases.

In any situation where evaluated data are presented to the user, the fact that data have been evaluated, by whom and when they were evaluated, and the evaluation procedures applied should be identified.

Units

Units and unit conversions should be handled consistently and precisely in storing and distributing materials data.

It is recommended that data be stored in the original units in which they were generated, even in cases where the database is distributed in some converted system of units.

As noted above, in providing for conversion of units, care should be taken to avoid the inappropriate and incorrect impression of a greater level of precision of the converted numbers than of the original data by presenting a greater number of significant figures in the converted numbers than in the original data.

SYSTEM CAPABILITIES AND CONTENT

Basic System Capabilities

Once connected to an online data source or in possession of a PC data disk of any kind, the user should find a basic set of capabilities available for accessing, searching, retrieving, and utilizing the information. The following elements are recommended:

Straightforward procedures for logon, and procedures to follow if logoff or disconnection occurs.

Prompts and error messages should be self-explanatory or clearly explained.

Regular and clear help functions which anticipate reasonably well the questions likely to arise.

A limited, relatively simple, and unambiguous set of commands, consistently used throughout the system. Sufficient repetition of the basic commands so that users are comfortable and rarely if every "hung up" (that is, in a situation where no obvious command seems to work). In response to incorrect commands, the system should provide helpful error messages pointing the user to correct procedures.

A variety of search paths and options, including some which are clearly guided, menu-driven paths for the new or occasional user as well as an "expert" or "command" mode that bypasses most menus for the experienced user.

The ability to backtrack to review previous screens without losing a search path (a search path tracking system).

Clear descriptions of special capabilities for viewing data such as the utilization of "windows" or color graphics.

Complete descriptions of the types of data contained in the system, clearly stating indices of reliability (or lack thereof) in a factual (not subjective or judgmental) manner, records of the original source(s) of the data, and who the user may contact to learn more about the background of the information.

Terminology that conforms to user/market industry standards. The interface should either present a reasonable latitude in terminology and nomenclature or provide quick and easily followed guidance on the preferred terms. Abbreviations and units should be given the same attention.

Capability for conversion from one system of units to others.

The ability to download data electronically into a reasonable range of compatible software/database management systems.

Easily followed procedures for ending searches, returning to key menus, and exiting the system.

Basic Data System Content

Several basic elements of information associated with the data should be provided, at minimum:

Scope

Materials and properties covered by the data source.

Source

Original source of data, either the data generator or the prior producer/publisher of the data.

Type of Data

Raw test results, statistically analyzed data, design values, and so forth.

Test Procedures

How the data were generated.

Evaluation of Data

How and by whom are data evaluated?

Key Contact

Names and phone numbers or other access modes for individuals to contact if questions arise about the technical content beyond that covered in the database.

Nomenclature/Terminology

Basis and limits of terms and abbreviations used in the database.

Units

Units include system of units used in the database and options available for conversion (highly recommended). The International Standard System of units (SI) is recommended as the primary set (see ASTM Practice for Use of the International System of Units [SI] [The Modernized Metric System] [E 380)]).

The User Manual

A readable, straightforward user manual should be provided, containing at minimum some introductory information, a data system overview, access procedures, guidance on terminology capabilities and limitations, general explanations of user commands, and more detailed explanations of commands and search options available to the searcher. Among the specific things recommended for coverage in the manual are the following:

Introduction to the database or group of databases, covering the scope of the source (materials and properties) and types of data presented.

Installation and start-up procedures for personal computer or work station databases.

Access and logon procedures, and procedures to follow if logoff or disconnection occurs.

A clear description of the search options available in the database or network, preferably with a schematic or flow chart summary, and typical screen displays.

Information about the quality and reliability of the data in the individual databases in factual (nonsubjective) terms, including procedures for collecting and evaluating the data, the completeness of documentation of test methods, and the traceability of the data in situations where only summaries of evaluated data are presented.

Breadth of terminology usable in searching, including nomenclature, abbreviations, symbols, units, and acronyms.

A clear summary of the commands used throughout the system.

Help functions, including prompts and error messages that may appear on specific occasions.

Clear guidance on how to end searches, return to key menus, and exit the system.

One or more example runs of the system. These should start from the offline situation, giving the precise inputs required of the user, and showing exactly what the response of the system is at each stage.

Contacts for problem resolution.

An index to the contents of the user manual itself.

Ease of Access to and Use of Online Systems

Users should quite readily be able to establish connections to any online data source.

Networks

The system should be connected to a commercially available and widely accessible public packet-switched network operating within the X25 protocols. The means to connect should support the Open Systems Interconnection (OSI) model of the International Standards Organization (ISO).

Connection Capacity

The minimum capacity of the connection between the host and the public network should be 2400 baud. A capacity of 9600 baud is highly recommended.

Number of Users

The connecting device between the X25 network and the data source should allow for multiple users having simultaneous access to the system. While no specific limit is appropriate and needs will vary with database content, limits of less than five users are not recommended.

Logon Procedure

The logon procedure should be straightforward and clearly explained; prompts and error messages shall be either self-explanatory or clearly explained. Procedures for recovery from disconnection, if different from those for the original connection, should also be explained.

User Interface

The interface screens should be clear, concise, and readily understood, so that at any point during use of the data system, a user easily understands the options available, the decisions necessary, and the response syntax required to proceed.

Times of Normal Access

Access to any public system should match the normal business operations of the user community being served. At minimum, access shall be from 1000 to 1600 h Monday through Friday, in the countries of origin and of primary service. A system serving an international community must recognize the wide range of normal business hours involved.

Availability

During the hours of operation, the availability of the system to users should be at least 90%, averaged over a four-week working period.

Serviceability

The serviceability (percentage of time in operating condition) of the data system, computer, and supporting systems by its maintainers should be at least 98% averaged over a four-week working period.

Ease of Startup of Disk-Loaded Databases

Users of PC or CD-ROM disk-loaded databases should quite readily be able to load and startup the search and retrieval system, plus any other special capabilities of the software.

Hardware and Software Requirements

Specific hardware requirements for accessing the software and data files should be spelled out clearly and concisely in the accompanying documentation. Also, if other software is required to completely utilize or to supplement the use of the database software that should be clearly spelled out.

Loading Procedure

The procedure for loading the software and associated files onto the user's host system should be straightforward and clearly explained in literature accompanying the disks. Quick startup procedures should be provided for the experienced users and made readily visible in the documentation materials.

Accessing the Database

The procedure for initiating use of the search and retrieval software should be straightforward, logical, and clearly explained in the documentation. Prompts and error messages shall be either self-explanatory or clearly explained.

User Interface

The interface screens should be clear, concise, and readily understood, so that at any point during use of the data system, a user easily understands the options available, the decisions necessary, and the response syntax required to proceed.

Procedures for recovery from hangup or losing the software, if different from those for the original connection, should also be explained.

Help Services for the User

A complete user manual, online help services, and an offline ("hotline") communication service, and in the case of disk-loaded database system, reference contacts are minimum requirements for any operating data system.

User Manual

The user manual should contain an introduction and overview of the system, instructions for connecting to the system, general and detailed explanations of the commands, listings of the types of help available to the user, and example runs of the system. It should also contain some treatment of what the user needs to know about nomenclature, terminology, symbols, units and abbreviations in using the system, notably the breadth/limitations of the dictionary/thesaurus.

Online Help

Regularly utilized commands should be frequently, if not continuously, presented on screen. At any point in use of the system, the user should have the ability to access either a specific help screen tailored to that position in the search or a general help system, within which it is quite easy to locate helpful information. At a minimum, the help system should explain the function and syntax of commands recognized by the data system. When incorrect commands are entered, the user should get helpful error messages rather than a simple rejection of the command.

Offline Help

A help desk accessible by public phone connections should be maintained by the data system provider and be accessible during all normal operating hours of the system. The minimum level of assistance available in this manner should provide competent guidance to the user through the system operating procedure.

Reference Contacts

In the case of PC or CD-ROM disk-loaded databases, help and backup information should always be available for users from disk or database producers. Phone numbers, and where possible, the appropriate names of contacts should always be provided.

Testing of the Database or Network System

Before a database or network of databases is placed in the marketplace for general sale or distribution, or both, it should be carefully and completely tested with a finite, but significant number of representative users. Care should be directed at the selection of test participants to assure that all types and backgrounds of potential users are represented, especially if the system is directed at other than professional information specialists.

SECURITY

Security of Data Content

The data system should be protected to prevent intentional or unintentional contamination of the information contained therein, and inappropriate access to the system by unauthorized users.

Uploading Limitations

Uploading of data to original online producer-supplied databases should not be permitted. If it is the service provider's intent to carry user created files, they should be maintained completely independently of the original database.

Unintentional Alterations

PC disk-loaded databases should make it very difficult, if not impossible, for the user to inadvertently change files while browsing or retrieving data.

COSTS

The relatively high cost of some operations associated with building, maintaining or distributing materials databases, or both, is sometimes a deterrent to high-quality products and services in this field. It is well to be knowledgable about them at the outset, so they are not erroneously underestimated and budgeted, with resultant poor quality or performance.

Cost of Locating the Data

In most instances the data are widely spread and rather laborious to locate. Some may be identified by bibliographic database searching, and the extensive followup required to assure relevance; others must be tracked by working through original data records, frequently an ever more laborious process, but actually preferable because it offers the opportunity to avoid errors of multiple transfer, and often provides the opportunity to obtain better and more complete background information.

Cost of Assembling the Data

Dependent upon the path used to locate the data, obtaining access may be accomplished at the same time, but about equally often, additional steps to obtain more complete backup information or acquire it once identified from other parties is involved. This can be one of the most time-consuming steps in the process.

Sorting and Organizing the Data

Unless great care has been utilized in the original data assembly, sorting and organizing the available data are often labor-intensive processes, but are key to the value and utility of the resultant database. It is essential that the data unit plan and type of search and retrieval anticipated be well thought out as the first step of this process, and that all subsequent work be maintained to fit that plan carried out. The result of this step in the process may well be a good hard-copy version of the file.

Cost of Evaluating the Data

Of all of the costs, this may well be the most difficult to estimate unless a very specific analytical evaluation is what is required. As defined in Chapter 6, there are many types of evaluation processes, and the need will vary greatly with the type, quality, and reliability of the source data. The major point to recall is that this may well be a very time consuming process, and it will require individuals who are expert in the discipline involved.

Cost of Producing Machine-Readable Products

There may well be several components of cost embedded in the total cost of producing machine-readable product. If complex tables, graphs, and other figures are involved, a major part of the cost can be the table or graphic capture process, which is far more than a simple scanning since the intent is to enable the material to be searchable, not simply represented. This requires capture of each component of the table or figure in a database format itself, and organization of that material for searching. If tabular or graphical information is not involved, it may involve only keying or scanning the numeric data. Recall, however that most numeric data require an extensive amount of factual support information, including their units, and all must be included in the digitization process.

Cost of Peripheral Databases (Thesauri, Directories)

If the system happens to be an online data distribution system, in which a number of files are accessed and searched as a cluster, there may be a need for the addition of new terms and associated factual data in a cross-file thesaurus or directory type of database. There are powerful advantages to such systems, but also an intellectual effort.

Other Elements of Cost

Other elements of cost may very well be involved, dependent upon the status of the software involved, the amount of programming required to produce search and display files that meet the objectives of both the user and the host software, and the number of new features that may be needed because of the addition of a new file. Even if the software is well developed, there may be substantial programming involved.

As is obvious from the previous information, it is impossible to state even an approximate cost of building a database without a clear statement of specific scope, content, and intended use.

SUMMARY

Guidelines have been provided for the building, maintenance and distribution of computerized materials property databases. Key factors in the four major areas requiring management structure include:

(1) In overall management operation:
  (a) qualifications of personnel
  (b) standard procedures and practices
  (c) auditable quality assurance programs
  (d) maintenance and updating
(2) For data and data management:
  (a) completeness and documentation of data
  (b) consistency and quality of data
  (c) data evaluation practices
  (d) data loading and reloading
(3) For system capability and management:
  (a) basic system capabilities
  (b) basic data system content
  (c) the user manual
  (d) base of access and use of the system
  (e) help services for users
  (f) testing of the data system
(4) For security of data:
  (a) intentional alteration or removal
  (b) unintentional alteration or removal

Factors affecting the cost of materials properties database operations are also summarized.

REFERENCES

[*1*] Code of Practice for the Operation of Materials Data Systems, produced by the European Communities Network Demonstrator Programme, Petten Establishment, Netherlands.

[*2*] *"Materials Data Systems for Engineering,"* J. H. Westbrook et al., Eds., *Proceedings of a CODATA Workshop*, Schluchsee, Germany, Sept. 1985, Fachinformationszentrum (FIZ-Karlsruhe), 1986.

[*3*] Barrett, A. J., "On the Evaluation and Validation of Engineering Data," *Computer Handling and Dissemination of Data*, P. S. Glaser, Ed., CODATA, Elsevier Science Publishers, North Holland. 1987.

[*4*] *Guide to Material Property Database Management*, CODATA Task Group on Materials Database Management, J. G. Kaufman, Ed., CODATA Bulletin 69, June 1988.

[*5*] Kaufman, J. G., "Data Validation for User Reliability," *Materials Property Data: Applications and Access*, MPD-1, PVP-Vol 111, ASME, New York, 1986, pp. 149–158.

[*6*] *Computerized Materials Data Systems, Proceedings of the Fairfield Glade Conference*, J. H. Westbrook and J. R. Rumble, Eds., National Bureau of Standards, Gaithersburg, MD, 1983.

[*7*] *Materials Data for Engineering, Proceedings of a CODATA Workshop*, Schluchsee, Germany, J. H. Westbrook et al., Eds., FIZ-Karlsruhe, Germany, Sept. 1985.

[*8*] Kroeckel, H. and Gunter, S., "European Activities Towards the Integration and Harmonization of Materials Data Systems," *Computerization and Networking of Materials Property DataBases, STP 1017*, J. S. Glazman and J. R. Rumble, Eds., American Society for Testing and Materials, Philadelphia, April, 1989, pp. 63–74.

[*9*] Reynard, K. W., "Materials Designation Systems: Their Problems and Possible Solutions," *Computerization and Networking of Materials Databases, STP 1106*, J. G. Kaufman and J. S. Glazman, Eds., American Society for Testing and Materials, Philadelphia, 1991, pp. 57–74.

[*10*] Rumble, J., "Standards for Materials Databases: ASTM Committee E49," *Computerization and Networking of Materials Databases, Second Volume, STP 1106*, J. G. Kaufman and J. S. Glazman, Eds., American Society for Testing and Materials, Philadelphia, 1991, pp. 73–83.

[*11*] Westbrook, J. H. and Grattidge, W., "The Role of Metadata in the Design and Operation of a Materials Database," *Computerization and Networking of Materials Databases, Second Volume, STP 1106*, American Society for Testing and Materials, Philadelphia, 1991, pp. 84–102.

[*12*] Kaufman, J. G., "Increasing Data System Responsiveness to User Expectations," *Computerization and Networking of Materials Databases, Second Volume, STP 1106*, J. G. Kaufman and J. S. Glazman, Eds., American Society for Testing and Materials, Philadelphia, 1991, pp. 103–113.

[*13*] Ambler, E., "Engineering Property Data-A National Priority," *Standardization News*, Aug. 1985, pp. 46–50.

[*14*] Rumble, J. R., Jr. and Smith, F. J., *Database Systems in Science and Technology*, Adam Hilger, Bristol, England, 1990.

8

Data Transfer

Philip Sargent[1]

OVERVIEW

The major difficulty in transferring materials property information among database systems is the conceptual mismatch of the sending and receiving systems, between the ways in which (1) materials are designated and (2) properties are defined. The second major difficulty is that the two systems usually represent relationships between data and metadata [*1*] using different conventions. Third are the relatively minor technical problems of agreeing text or binary file formats, which relate names of terms to their values, and which represent numbers, symbols, units, and text. There are organizational problems of agreeing which and what kind of information should be transferred, what pricing scheme to use, determining who should pay for the development of translator software, and determining how it should be sold or distributed. These are even more problematic in practice than the technical issues. Transfer of all technical data has certain things in common with respect to scope, software, and organization.

Scope

Any transfer format should be capable of representing all the different varieties of information that are present in the sending and receiving systems; it must have a super-set of capabilities. However, it must also not be "multi-valent," that is, it should not be possible to represent the same information in more than one way, or the receiving system must be significantly more complex in order to recognize such equivalence.

These two requirements can both be satisfied for closed-world types of information, such as moves in the game of chess, or engineering drawings with a pre-defined list of allowed geometric entities. unfortunately they cannot be reconciled for open-world information where new entities and relationships can be created by any participant such as moves in the game of soccer or unique properties of new materials.

Open-world information can be handled in one of two ways: either by (1) pre-defining a limited set of entities and allowed relationships (transforming it to the simple closed-world type) and recognizing that some information will just not be capable of representation (as represented by the dark-shaded areas in Fig. 8.1 or by (2) attempting to find some more abstract set of entities from which all potential information can be derived. This latter technique (abstract closed-world) nearly always introduces multi-valency because any sufficiently abstract concepts can usually be combined in multitudinous ways to produce equivalent descriptions.

Software

Translator software costs money to develop, and historically the most successful transfer formats have emerged either by following a commercially successful product (such as the DXF drawing format from AutoDesk or the PostScript® page description format from Adobe) or from generous, long-term support from a single large government agency (such as TCP/IP networking protocols from the Department of Defense, or the Spatial Data Transfer Format from the Department of the Interior and Geological Survey).

Consortia-developed formats are more common but also generally more troublesome because different organizations choose to implement them in slightly different ways by adapting their own software in the cheapest manner. The Initial Graphics Exchange System (IGES) engineering drawing standard is notorious for this, as is error-handling in Standard Query Language (SQL)-based database systems (which is not covered by the SQL international standard).

It is unfortunately the case that engineers do not have the training in modern discrete mathematics that is necessary if nearly unambiguous specifications for technical data transfer formats are to be defined. (In any case, completely unambiguous specifications are a theoretical impossibility.) This means that "standard" data formats are defined in practice with respect to a reference implementation and not with respect to the specification itself. The number of companies claiming compatibility with IBM's use of SQL in its DB2 database is a case in point.

The news is not all bad. Occasionally a good, complete standard transfer format emerges from a communal effort which is widely accepted without modification or extension. The Electronic Data Interchange Format developed by a Stanford computer science professor with support from 14 Californian electronics companies has been very successful,

---

[1] Senior software manager, Quillon Systems Ltd., St. John's Innovation Centre, Cambridge, United Kingdom, CB4 4WS.

Copyright© 1993 by ASTM International www.astm.org

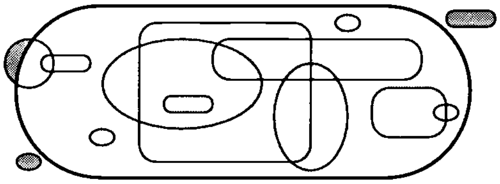


FIG. 8.1—Transfer formats have a superset of representations with respect to the databases between which they communicate, but some data cannot be transferred.

possibly because it was designed so that translator software would be easy to write and to maintain.

"Neutral" Data Formats

A "neutral" file format is one that is independent of any proprietary system but that is capable of transferring any materials information. From the above discussion on scope and software, the idea that there could be such a thing as an entirely "neutral" universal data transfer format for materials information can be seen to be fatally flawed, both in the abstract sense (closed-world assumption) and in the practical (the common need for reference software implementations). However this does not mean, now that we have been alerted to the dangers and difficulties, that it is impossible to produce useful data transfer formats.

The simple message to someone contemplating the construction of a materials database is that if materials data transfer with any other information systems is to be part of the project, then it must be planned and designed first because it is the most constraining influence on the structure of the database schema.

Passive Transfer and Active Interchange

This chapter discusses techniques for transferring materials information "passively," that is, without participation from the receiving system when the transmitting system is preparing the transfer. The information is assumed to be prepared, transmitted, and received as separate operations, and the information in transit contains everything required for its correct interpretation. Such passive data transfer systems are suitable for archiving information and for broadcasting information where the recipients are not known and indeed may not yet exist. Current examples are the free materials information systems produced by materials suppliers as part of their marketing efforts, and by technical associations as part of services to their members.

"Active" information transfer occurs when there is direct, real-time communication between the transmitting and receiving systems and where feedback is used to control the form of the information transferred. An intelligent agent using active interchange can extract far more meaning from a transmitting system than the basic data transfer format used would appear to allow, since queries and clarifications can add to the information's context and hence increase its meaningfulness. However, practical active transfer (without involving human agents) is many years away.

In the future, sophisticated software will be able to assume the role of an intelligent agent and to negotiate with a source database the best translation of terms and relationships into those of the receiving system. To a degree, commercial software of such sophistication already exists and is in intensive use (invisibly) in the routing and communication in Apple's local networking system for Macintoshes [*2*].

History of Materials Data Transfer

Successive international workshops and symposia have identified data transfer as a critical need, but it was at the (VAMAS) workshop in Petten in Nov. 1988 that several implications for materials data interchange resulting from the activities of materials database users became apparent. There were three main results [*3*] as follows:

1. Specific needs for data interchange were articulated.
2. Self education was acknowledged as a necessary precursor to intelligent discussion of data interchange formats, both technically and with respect to other relevant organizations.
3. It appeared that two or three formats was a feasible goal to aim for.

A meeting was organized in Sept. 1989 at Derby in the United Kingdom to identify a workable format and an informal, international round-robin attempt followed. Several important lessons were learned from the experiment, nearly all of which were independent of the format that was notionally the subject of the test.

There were problems of communication within individual participating organizations as to the precise purpose of the test. Commercial organizations, particularly the small companies that typify those involved with materials databases, find it very hard to provide data modeling and software support for a project with no immediate profitable outcome. The round robin was considered too "applied" for research funding, but significant research was nevertheless required since the different data schemata of the communicating databases presented a significant and intractable problem.

CLASSES OF MATERIALS DATA TRANSFER

Four major types of materials data transfer are generally acknowledged, and the early work in particular acknowledged a close relationship between data transfer and terminology consistency [*4–6*]:

(1) Initial data entry to a database.
(2) From a database to the user interface.
(3) From a database to specific software packages, such as for finite-element or other modeling software.
(4) From database to other databases or database gateways.

The idea behind such classifications as this is to identify those types of data transfer that could take advantage of the same data transfer formats. Classifying data by the type of data it is (economic, high-temperature, corrosion, and so forth) is usually felt to be less useful in this respect. However this four-fold classification does not handle the important issue of where in the data-product cycle the information is being transferred, and therefore what its level of refinement and abstraction is.

Data Entry

It is very rare for initial data entry to occur unless it is targeted at a specific database, and therefore it is almost invariably structured directly in the format of that database. The High Temperature Materials Database at Petten (Netherlands) makes available a software package that makes this fact concrete, that is, the user-friendly package guides data entry and produces files suitable for immediate review and uploading to the main database.

When transferring, once only, a large quantity of materials data from one (perhaps obsolete) database structure into another conceptually similar structure, it is often worthwhile to design a specific transfer format purely for that purpose. Such a transfer format can be very simple since it only has to allow for two different data structures, and often only a strict subset of some existing, more complex format is suitable. If the format is textual, then occasional untranslatable idioms from the source database can be hand-typed in this intermediate form thus saving the expense of producing complex translator software.

Data capture from text or diagrams requires special aids and a transfer format designed to cope with very complex interrelationships between different types of data points, lines, and curves, and between data and metadata [*1*].

Databases and User Interfaces

Chapter 2 covered program infrastructure and dealt in some depth with the complexities of user interface design. Even when the database and the front-end user interface programs are running on the same machine and were purchased from the same supplier, it makes matters clearer to consider them as distinct entities. Client-server architectures, where the two programs are distinct and where a shared database on one machine is accessed by several individuals interacting with their own personal computers, make the distinction concrete.

The background database and front-end user-interface programs require a communication of both materials data and control of the display of materials data. It is very widely agreed that it would be ideal if there were standards for both data and control information so that a user could access many different databases using the same, familiar front-end [*4,5*]. Such standards would have to be specific to materials data since there are too many different ways in which materials information can be represented in standard business-type databases, and the ambiguity leads to conceptual mismatch [7]. A demonstration common front-end system has been developed for metal-matrix composite materials data at the Concurrent Engineering Research Center, University of West Virginia, using data from several different proprietary databases running on different types of computer over a local area network.[2]

In the past two or three years, "scientific visualization" and data presentation software for personal computers has made great progress in capability, cheapness, configurability and user-friendliness. The specific needs of materials evaluators for a few functional presentations: Larson-Miller parameter fitting, Goodman fatigue plots, composites failure carpet plots and so forth, can now be configured from general purpose software in only a few hours. However because there are no materials data transfer standards every database requires a different configuration. In addition, data visualization packages do not handle metadata in any systematic way.

Downloading Data

The action of "downloading" data from a database into a user's personal spreadsheet, database, or data visualization package is usually perceived as forming part of the user interface, as indeed it is. However, for our purposes it is clearer to consider downloading in two halves: (1) setting up, selecting, and controlling the download, and (2) the format of the data that is downloaded. Currently the data download formats available are very rudimentary, usually only a single table at a time where the elements are numbers of various types or test strings, the columns of the table have names, and everything in the same column has to be of the same type (integer, number, text-string, and so forth). Examples of such formats are DIF, Sylk, Lotus WKS, or dBase DBF. This permits no complexity at all in the relationships between different data items; all metadata relationships have to be handled ad hoc using the setting up and control facilities (the "active" communication channel). Downloading data is further discussed under "SQL" below.

Standards exist for the transfer of all kinds of data: Diagrams can be encoded as Computer Graphic Metafiles, images as Tagged Image File Format or Group III Facsimile, chemical structures as Standard Molecular Data format, complex text as Structured Generalized Markup Language (SGML), and complete page layout in PostScript®. However, there are very few standards that relate to the meaning of the information being transferred and that is where the need exists for materials databases.

Database to Other Software Packages

Specific software packages are almost invariably highly restricted in the type and variety of materials data they require. Some configuration, however, may be possible, so that a finite-element system may be able to handle temperature-varying, anisotropic elastic moduli as an alternative to a single isotropic room-temperature value.

If any standard data transfer format existed then it would be cost effective to take pre-existing software packages that use materials data and to encapsulate them with a translator to transform the standard representation into package-specific form. New packages would be designed to take the standard format as direct input. This is the role envisaged for the proposed ISO Standard Product Information Representation and Exchange (the materials format is described in Part 45 of the Draft Proposal 10303, issued at Tokyo, 1990). It is universally referred to by its informal title "STEP: Standard for the Exchange of Product Data."

---

[2] Up to date information on this project is available be electronic mail from tad @ cerc. wvu.wvunet.edu; CERC publications can be obtained by anonymous file transfer from pub/techReports on babcock.cerc.wvu.wvunet.edu.

Between Databases

Database to database transfer is distinguished from database to other software because although the receiving database may be as restricted as any other software, it could have the capability to represent highly complex data and metadata interdependences. This is the most general class of data transfer problem, and all the other classes could be considered to be aspects of this case.

It is the transfer of relationships and associations and the degree of "associativity" that can be represented, which is the root of the problem. Most data transfer formats are at too "low" a level, that is, they deal in numbers or sets of numbers (tables), not in the relationship that one set of numbers has with another [7,8].

A significant relational materials database can contain 20 to 50 distinct tables where implicit data normalization relationships (not stored in the tables) are absolutely necessary to give the tables meaning [9,10]. An analogy with natural language translation is relevant: one can easily translate individual words from one language to another—though there are known difficulties such translations are rarely exact—but that does not produce a translated sentence.

SQL

SQL is a query language commonly used for the communication and control aspects of data transfer to and from relational databases. To handle the actual transfer of data, it must be embedded in a "host" language, such as C or Pascal, or some proprietary fourth generation database programming language [11]. SQL statements have to identify the precise names of the tables and fields (columns) in the database being queried, and there are no standards yet for automatically acquiring the relevant names when a database is first contacted. SQL produces the results of a query as represented in data structures in the host language. If the data is to be transferred into another database then another SQL program is required to do the "upload," using the (different) precise names for that database, assuming that there are direct one-for-one identities between the two sets of names with no omissions. Omitted columns can break the integrity relationships on which relational databases are founded [12,13].

SQL is therefore an appropriate language for communicating with a known and well-understood database where the SQL statements have been programmed specifically to access specific pre-defined queries, such as the interaction between a front-end user interface with fixed facilities and a single, specific background database.

If standard methods existed for describing in a catalogue the representation of materials data in relational databases, then sophisticated front-end programs could use that catalogue to construct appropriate SQL queries. This would be possible even if the database were one which was entirely new and had never been accessed before. However, use of SQL in this way, as one element in an effective data transfer system, requires that all participating materials databases strictly adhere to as-yet-unwritten rules defining how catalogues are to be structured. It is thus at best a long-term solution.

SQL as defined in ISO 9075:1987 (SQL1) is not a new language. It knows nothing about the handling of catalogues or data dictionaries as distinct from any other data structure. SQL1 is defined for two levels of implementation: Level 1 does not support NULL values for any field in the database, a vital requirement for most materials data, so the purchaser of an SQL system is advised to be aware of this. It also does not define what happens when errors occur or how error messages should be phrased or handled.

As a computer language SQL1 shows its age. It displays "a lack of orthogonality in expression and formats, a mismatch with its host language," has "mistakes" in the implementation of NULL values, and does not support at all some important aspects of the relational model, such as domains [11,12]. SQL also has many minor deficiencies, such as the lack of support for time and date information, and for variable length character strings (which is much more serious). It also reduces programming productivity because it was developed before many modern programming techniques became practicable and therefore retards their use.

The standard will be updated as SQL2 in the near future. It too is defined in several levels. Level 1 is simply a superset of SQL1, for compatibility, but now including standardized error handling. Levels 2 and 3 provide significantly better facilities and correct the major problems. Level 2 provides for variable length strings, support for multiple modules, a "match" operation, national character sets, comprehensive error analysis, outer join operator, datatype casts, CASE expression for NULL value conversion, additional consistency levels, and the ability to dynamically construct SQL queries at execution time instead of having to have them precompiled. SQL2/3 (Level 3) tidies the syntax to make it "cleaner," introduces domain checking, datatypes for dates and times, time-zone handling, deferred constraint checking, self-referencing delete and update cascades, scrolling cursors, and simple assertions.

A working draft of SQL3, the planned successor to SQL2, already exists though it is likely that before it is issued other activities (such as object-oriented and/or "semantic" databases, and perhaps the International Resource Dictionary System [IRDS] work) will have rendered the whole SQL approach obsolete. SQL3 is planned to introduce assertions triggered by general events, generalized tables and subtables, external procedure calls, user defined types, and perhaps up to 17 different types of NULL value (not present, irrelevant, inappropriate, and so forth).

Therefore, although SQL is a very significant achievement in raising the level of abstraction available for manipulating databases, there are many problems in representing materials data to which it is irrelevant. Most companies selling database management systems will include some of the facilities of SQL2 in their products before it becomes a standard, but none plan to implement it fully before then.

Object-Oriented Descriptions

Object-oriented databases are specifically designed to represent varieties of relationships between data explicitly. However, there is as yet no standard way of representing object relationships except through using a particular proprietary database. The standards that do exist, the languages C++ and the Common Lisp Object System (CLOS), are at too low a level. They represent named objects but do not have a standard way of naming the all-important relationships between objects. To suppose that producing a system using C++ will solve information transfer problems is equivalent to assuming that if an Englishman and, say, a Spaniard both learn Morse Code, then they may be able to converse intelligently. C++, like Morse Code, is a very low-level standard.

TRANSFER OF SPECIFIC TYPES OF MATERIALS DATA

If a datum is classified as to what it is then the set of relationships and concepts associated with each materials data class will be smaller than if it is classified by how it is used. However, although individual data transfer formats might be easier to produce, there will be many more of them, and there will always be the continuing need to translate from one type into another.

Product Data Cycle

The different uses of materials data in testing, aggregation, design, and analysis (including simulation of manufacturing processes) have already been described earlier in this book. Each stage of materials evaluation has characteristic relationships and descriptions that are more similar within each stage than between stages, but of course data necessarily have to progress between stages so designing transfer formats to be stage-specific is of limited use. Nevertheless some types of data do lend themselves to specific and useful treatment.

Raw Data

Strictly speaking, raw data are those produced directly from experimental measurement, but the term is also used to mean experimental data together with ancillary experimental details (metadata) and after preliminary validation (checking for conformance to standard test practices). Data resulting from any standard test are very close to being in a standard format, and ASTM is improving this by regularizing data recording (as described in Chapter 5). However, part of the problem is that ASTM data tables are not normalized are therefore not straightforwardly machine-interpretable. The terminology and precise meaning of individual data items for many ASTM tests is now defined, and it would be a very small step to now define a formal, written recording format. This would be most straightforward if it were an "item-based" format (see later in this chapter).

Nonstandard Test Data

Data resulting from nonstandard tests are more difficult. These could come either from innovative research, or from tests that were intended to be standard but which did not conform in some way (for example, though use of smaller specimens than usual because of limited material samples). The latter could be represented using the format for the standard test with an added annotation, but many database managers would see this as dangerous since receiving databases may omit the annotation. ASTM test records do include a formal field stating whether the test did conform to standard, so the danger can be overcome.

Data resulting from entirely nonstandard tests, or recorded from components in service, or from manufacturing processes, present the greatest problem. For tests, the appropriate ASTM abstract test recording proforma could be used to derive a test record format which, while not standardized itself, would be close to what would be produced were ASTM to produce such a standard. This could then be formatted using the same "item-based" system as might be used for real ASTM standard tests. Alternatively, the test (or service conditions, or process) could be described using EXPRESS, the data modeling language used to define the ISO 10303 Standard on Product Representation and Exchange, Part 11. This requires specialized expertise. It has been performed (once, by an expert) for a British Standard for the recording of "single-point" data for polymers.

General Issues in Raw Data Formats

Wherever a new format or a new variety of format is devised, it is extremely important that great care be taken with the use of terminology [*14–16*], that is, that existing terms not be used in cases where there is even a slightly different shade of meaning, and that new terms be documented and defined explicitly, preferably with reference to existing thesauri (for example, the ASTM and the ASM/Institute of Metals glossaries, and the European Commission's Common Reference Vocabulary for materials information [*17*]).

Whatever type of format is used, the fact that raw data are recorded implies that certain characteristics would be useful for each datum:

(1) It could be logically simple enough to be machine-generated directly from testing machines (complexity in the format would not be a problem now that modern testing machines are controlled by standard personal computers).
(2) It could be in text form and thus hand-edited to add conditions not monitored by the machine and typed by the operator where the machine is too old to do it itself.
(3) It could be interfaced usefully with existing computerized Laboratory Information Management Systems (LIMS).

Materials Index Transfer

One distinct type of materials information is indexes of materials databases. Any establishment of either a gateway system to mediate between users and a number of distinct, narrow databases, or a market in materials information requires the formulation of a format to support the exchange of indexes (meaning catalogues, data directories, data dictionaries, thesauri, data encyclopedias, information resource dictionary systems, and so forth), so that different systems can communicate their capabilities and limitations.

There are organizational difficulties in communicating catalogues since many database providers consider that their methods for structuring information convey competitive advantage, and while they do not mind their own customers taking advantage of such structuring, they would be loath to publicize it more widely. Note that the internal representation of any database, which presumably is tuned for efficiency and performance, is irrelevant. It is only the structure of the information as seen by a client user-interface program that is important.

Techno-Economic Materials Data

There are data that describes availability, pricing, and delivery timing of commodity materials products, but also data describing individual shipments and their quality control and quality assurance materials property data. These individual shipment data are similar to Laboratory Information Management Systems (LIMS) data.

Commodity Materials Data

Commodity materials information has to be up to date, and there is a niche here for on-line database services, but otherwise it is very similar to other materials property information defined for particular specifications of materials (metal alloys, polymer grades). There are no standards for transferring this information specifically, but the data relationships are simple: one grade at one delivery schedule has one price, so simple tables with named columns (fields) are adequate. Since the information is almost invariably read directly by a human being, formal definitions of terms enabling the information to be machine-manipulated are rarely necessary.

This type of data is becoming increasingly problematic since a continuingly increasing proportion of materials are not sold as commodities. Technical ceramics nearly always have to be made for specific components; long-fiber composites always are made for specific components; polymer grades are frequently developed for particular customers; and increasingly even metal alloys are being ordered to variant specifications that have been negotiated between material supplier and engineering customer [*18*]. Since negotiations depend on the current state of the materials supplier's processing equipment, more sophisticated automated pricing information will have to await significantly more complex software implementations and carefully controlled interfaces to suppliers' internal computer systems.

X.12 and EDIFACT

Shipment data are currently encoded using standard ASC X.12 in the United States and EDIFACT in the rest of the world. A number of proprietary and industry-specific formats, very similar to X.12, also exist. X.12 is approved by the American National Standards Institute (ANSI), but ANSI recommends transfer to the UN/ISO Electronic Data Interchange for Administration, Commerce and Transport (EDIFACT) standard, which is a development from it and from other industry standards.

These standards are not only highly specific, they are also in practice ambiguous for materials properties data transfer because the concepts they define do not precisely match those used in materials engineering. Thus different people might use different identifiers ("tags") to represent the same thing. The difficulty with them is that every different type of message has to go through the entire international standardization procedure. The formats are defined at a very low level (in terms of individual bytes, or "bit octets," as they are called in the EDIFACT standard: ISO 9735-1987).

Specifying Engineering Products

Materials information frequently appears in the specification of a complete engineering product as part of the definition of the product's components. Frequently a simple identification of a standard alloy and heat-treatment is all that is necessary, but, as discussed above under "Commodity Materials Data," it is increasingly necessary to define the material with reference to its processing history.

While a product specification may just require a specification of the material and required properties, a product description, where a particular individual instance of the product is concerned, will require substantial audit and specific history information if it is used in a safety critical application. Examples include pressure vessels, aircraft airframes and engines, and most nuclear engineering applications. Since the materials information is only a small part of the data defining the whole product, its transfer formats have to conform to those used for the product as a whole.

CALS

The American Department of Defense Computer-aided Acquisition and Logistical Support (CALS) program aims to completely computerize all technical data relating to the acquisition of weapon systems. The method is to adopt subsets of existing standards for data transfer and archiving, and to change these adopted subsets as new standards are developed. Thus currently geometric design data are required to be available in IGES form, but CALS will migrate to using ISO 10303 when it becomes available. Several American military organizations are funding development work on ISO 10303 (usually known as STEP) through the PDES organization (Product Data Exchange using STEP) to ensure that it can meet their requirements. Other adopted standards are SGML for structured text (Structured Generalized Markup Language, ISO 8879-1986), CGM (Computer Graphics Metafile) for diagrams, and Group III facsimile for images. Most Department of Defense (DoD) procurement activity centers around the production and transfer of documents, so SGML has a special role since it can have data encoded according to any other format embedded within it.

Nonmilitary industry has enthusiastically welcomed the CALS initiative because it permits greater control of engineering information, reduces waste, and should permit more rapid response to market forces. CALS also generally adopts international rather than purely domestic standards.

ISO 10303 (STEP)

The draft proposal for the ISO Product Information Representation and Exchange (ISO 10303) aims for a complete specification of engineering products, including civil engineering, ship-building, power electrics, electronics, machines, and process engineering. It covers project management, version control, and records of the engineering design process. The standard is written in formal data specification language, EXPRESS, which defines "entities," how they are composed of each other and how they relate to each other. It is strongly conditioned by its origins in the requirement to communicate geometry information (CAD data transfer), and how such information should be displayed. Defining geometry using a fixed number of entities uses, as defined earlier, the closed-world assumption.

The materials content of the standard when it is first issued (probably some time in 1994) will be confined to material designation and identification (including processing history), independent variables, and other metadata, and linear elasticity parameters suited for finite-element analysis. Note that input parameters to finite-element analysis are also amenable to the closed-world assumption.

The representation of material designation and identification is based on the abstract closed-world principle. Sufficient basic entities are defined that any process or procedure used to create or refine a material should be describable by use in combination.

No test data formats are yet defined, which is something of a problem since many materials are identified, or more properly distinguished, by characteristic values of certain of their properties that are measured for quality assurance.

While any particular type of materials data can be represented in EXPRESS, such representation is not straightforward and takes a long time. It has taken over two years to agree on a representation just for elastic moduli and necessary metadata.

EXPRESS

ISO 10303 uses the data modeling language, EXPRESS, to define and describe materials entities, such as material_ _product or material_spatial_structure. ISO 10303 also defines how a set of data conforming to these EXPRESS entity descriptions should be represented in a "physical file," that is, a stream of bytes. (The only byte codes used are those defined as ASCII so the physical file is human-readable and can be changed with a text editor). EXPRESS also has a graphical analogue, EXPRESS-G, which precisely represents in "lines and boxes" the entities and their relationships. Several commercial "upper-CASE" (Computer Aided Software Engineering) software tools are available, which can be configured to define and maintain consistent EXPRESS-G diagrams [*19*].

The EXPRESS description defines the precise relationships and dependencies between the data and metadata entities and is thus capable of being used as a database schema, the description of how a materials database is logically constructed. Although ISO 10303 has been developed principally as a data transfer and archiving mechanism, several research teams[3] have developed software that constructs an empty relational database directly from EXPRESS definitions (whether these definitions appear in the ISO 10303 or not). Software which accepts the ISO 10303 physical file data and uses it to populate such an empty database is under development.

Figure 8.2 gives something of the flavor of EXPRESS, though it has many more facilities than are shown here. This example defines a triangle as being made up of three lines, each line being defined by two points. Note that any entity, once defined (such as 3D_Point), can be used as a "type" for a component of another entity.

```
SCHEMA My_3D_Demo;

  ENTITY 3D_Point;
      X, Y, Z: REAL;
  END_ENTITY;

  ENTITY Sphere;
      Radius: REAL;
      Center: 3D_Point;
  WHERE
      IF (Radius < 0) THEN
          VIOLATION;
      ENDIF;
  END_ENTITY;

  ENTITY Line;
      p0, p1: 3D_Point;
  END_ENTITY;

  ENTITY Triangle;
      l0, l1, l2: Line;
  WHERE
      l1.p0 = l0.p1;
      l2.p0 = l1.p1;
      l0.p0 = l2.p1;
  END_ENTITY;

END_SCHEMA;
```

FIG. 8.2—Example of EXPRESS.

Separately defined is the standard translation of data into a physical file format (bytes on a tape). All technical information has a logical structure defined in EXPRESS. There is then only one way to convert that data to the actual transmission format, and it is the same for the entire ISO 10303 standard. This is the same sort of layered approach as used by Open System Interconnection (OSI) in that by defining interfaces correctly it is possible to reuse software and to reduce complexity at each layer (see later).

"Open-World" Information in ISO 10303

Materials information requires an "escape" mechanism so that open-ended problems can be represented. Some facilities along these lines are being designed by the Materials ISO 10303 committees but are being strenuously opposed by many other working groups, particularly those representing the geometric "core" of the standard. There are two mechanisms for allowing more general communication:

(1) the facility to declare a property variable as a text-string, and hence to communicate any property, and
(2) the standard facility to extend ISO 10303 (STEP) by transmitting EXPRESS-descriptions of not yet standardized entities, using the software apparatus developed for STEP and the same physical file format.

The danger with either of these is that new properties will have new names, and without agreed standardized property

---

[3] Rutherford Appleton SERC Laboratory (RAL) in the United Kingdom and the United States Navy. Contact mm@inf.ral.ac.uk for information on the RAL tools.

names no communication is possible. Thus these extensions will be useful for industry-specific groups that can agree on defined terminologies. Realistically this is the only option anyway. The terminology problem is so large that it can only be handled by industry-specific groups working independently.

ISO 10303 (STEP) Invisibility

In principle, very few people would actually be aware of the inner workings of ISO 10303. They would just interact with ISO 10303-compatible software packages, which would ensure that any data transfer that was necessary would be performed invisibly and painlessly. Such a scenario is unlikely to be appropriate for materials information because the materials data cycle, from raw to validated and then specification data, necessarily requires changes in the "data modeling" of the information as it progresses. This would seem to require more direct intervention into the precise definitions of types of information that the ISO 10303 process anticipates. However, for designers whose primary task is to interact with finite-element models of components, and who always use the same type of materials information that is supplied to them by a separate data evaluation team, ISO 10303 will do everything they require.

ISO 10303 Recommendations

Anyone now beginning to specify a new materials database who is concerned about data transfer and wishes to eventually make use of STEP technology should obtain a copy of Part 11 of ISO 10303 (or a more recent update) and learn the EXPRESS method for logically defining database entities. EXPRESS is now stable and will not be changed, so delay cannot be justified on those grounds.

A new database should attempt to use as many STEP-standard materials entities[4] as possible, particularly those covering material product (form, processing, and so forth), material characterization, material general descriptions (source, identification, and so forth), and the standard logical relationships between the materials data environment (defining parameters and independent variables) and sets of materials properties. This is because a great deal of effort has been expended on organizing these concepts, and it would be a waste of money to attempt to repeat this work needlessly.

OTHER RELEVANT TECHNICAL DATA TRANSFER STANDARDS

Open System Interconnection (OSI)

OSI is an internationally agreed design template for information technology. This 7-Layer Basic Reference Model (OIS 7498) splits the facilities required for network interconnection (Levels 1 to 4, see Table 8.1) from those required for software to actually interwork (Levels 5 to 7).

TABLE 8.1—OSI seven levels and examples.

7	Application	FTAM, X.400, STEP
6	Presentation	ISO 8822, 8823
5	Session	Session Protocol 8327
4	Transport	ISO 8073
3	Network	X.25
2	Datalink	LAN logical link Control, LAP-B
1	Physical	Token Bus, Token Ring, Slotted Ring

Materials data transfer is, by definition in OSI, concerned only with the highest level, Layer 7, the application layer. The problem with many transfer standards for materials properties and other data is that they do not generally adhere to the eminently sensible OSI scheme. This means that the transfer formats are more complex than they need be since OSI separates functionality into different levels in order to enable software at lower levels to be safely and effectively reused. Thus, it is advantageous if a purely seventh-level application, such as materials data transfer, restricts itself to purely seventh level implementation details and uses the ability of other software to supply the underlying support (such as error-checking). This simplifies the format and reduces complexity.

If datafiles are to be exchanged on a potentially noisy system, such as by sending floppy disks through the letter post, it makes sense to use a separate commercial or public domain software to compress the file and to calculate a redundancy check (CRC), which can be checked by the recipient on decompressing to ensure that uncorrupted files are received. If transmitted over a communications system then the appropriate lower-level protocols should be used to provide underlying error-checking.

All the standards that define OSI are written in a formal language, Abstract Syntax Notation One [*20*], which helps to ensure that the standards are precise and unambiguous.

OSI Profiles

Because there are several alternative OSI standards in each "level" or "layer," for example, token-ring, token-bus, ethernet alternatives for local area networks, the total number of combinations is enormous. Thus purchasing and procurement organizations have introduced simplifications for their own use. These take the form of "profiles" or "stacks," which are "vertical" sections through the OSI 7-layer structure, and which include only one or two standards at each layer. Currently both the United Kingdom and United States governments have each defined a Government OSI Profile (GOSIP).

MAP/TOP and miniMAP

The Manufacturing Automation Protocols and Technical and Office Protocols (MAP/TOP) are all based on OSI principles and adopt many of the same standards. They also add a few more options at Levels 1, 5, and 7. MAP was devised by General Motors and TOP by Boeing, but both are now maintained by public bodies.

---

[4] The materials schema draft is available to all interested parties from the Chairman of the STEP Materials Committee, currently J. R. Rumble Jr., NIST. Up-to-date entire STEP drafts in machine-readable form are freely available, automatically, to anyone with access to electronic mail. Send a message consisting only of the words SEND HELP, and then, on a new line, SEND INDEX to Internet address nptserver@cme.nist.gov.

A cut-down version of MAP, called miniMAP, offers faster response for the factory shop floor by compacting the 7-layer system into 3 layers (which renders it non-OSI compatible). As far as top level (level 7) applications programs are concerned both OSI and MAP/TOP "look" the same so any materials transfer format that works with OSI will work with MAP/TOP [*5*].

Open Distributed Processing

One problem with the OSI view of the world is that a very large variety of functions and requirements seem to be classed as Level 7. This is quite natural. It is because there is no "upper limit" to the functions that people would like to embody in computer systems. OSI is only concerned with the transmission of data, not information. It does not address any of the issues that are involved in distributed systems, where the processing of the information also moves over the system and changes location. Architectures for distributed processing are now under development (an architecture in this context makes knowledge about a subject systematic through frameworks of abstractions, design templates, design guidelines, and recipes), and will define where OSI "stops."

The United Kingdom Alvey research program set up a project to study distributed systems called Advanced Networked Systems Architecture (ANSA), and this work is now taken up by ISO (SC21/WG7: Open Distributed Processing), by CCITT (Distributed Application Framework), and by ECMA (Standard 127 of TC32: ODP support Environment). Some international standards already exist that are properly better thought of as part of ODP rather than OSI, for example, CCITT's X.500 standard for directory services.

Open distributed processing standards will be important for "active" materials data interchange and vital to materials index interchange. The notion of setting up "conversations" is software in which information can be traded about which databases contain the materials data relevant to a query falls firmly in the province of ODP. Future materials information systems should ensure that they stay consistent with the developing international standards in this area.

TRANSFER FORMAT ISSUES

A passive communication of some materials property data must be able to express two things:

(1) the relationship between names of properties ("fieldnames") and their values, and

(2) the interrelationship or associativity between sets of these names and values.

The associativity is necessary in order to express the relationship between the data and that data that describe or modify the simple data. Examples of associations are those between the maximum and minimum values of a measurement, the ranges of conditions over which some other measurement is valid, or the existence of a functional dependence of a measurement on some other property (for example, hardness depends strongly on the heat-treatment history whereas density does not). The design of materials systems requires a larger number of different categories, typically several hundred, with complex interrelationships. The expression of the degrees of associativity permitted in a format must, however, always be part of the format definition since it is a direct function of the syntax.

Material	Tube	Bar	Strip	Wire
'Ti-6Al-4V'	No	Yes	Yes	No
'Al 6061-T6'	Yes	No	Yes	No
'Ag/Y-Ba-Cu-O'	No	No	No	Yes

versus:

Material	Form
'Ti-6Al-4V'	'Bar'
'Ti-6Al-4V'	'Strip'
'Al 6061 T6'	'Tube'
'Al 6061 T6'	'Strip'
'Ag/Y-Ba-Cu-O'	'Wire'

FIG. 8.3—Different ways of expressing names and values.

In the past, these two aspects of materials property data, names, and associations, have often been confused.

Terminology

The definition of names and the interpretation to be placed on them (their "meaning") can be defined within the specification of a transfer format, or can be excluded and defined elsewhere in some other reference document such as a data thesaurus [*21*]. The naming problem is general and independent of the structure of the data transfer format. All types of transfer format require that terminology be agreed; but the precision and detail of definition for software systems is much greater than that typically found in technical glossaries (for example, the CEC's Common Reference Vocabulary [CRV] or ASTM's on-going terminology projects).

A basic problem became apparent during the development of the CRV. Software developers require sets of definitions precisely as they are used within the databases that refer to them even if these definitions are more restricted or divergent from those generally accepted. This is highly confusing to general users who are used to "standard" definitions from pre-existing glossaries, which may have many, slightly variant meanings. An example from the CRV is "paintability," which can be used to denote the ability to be covered with paint, or the ability to be used as a paint.

It is easy to say that database developers should only use vocabulary that has already been precisely standardized, but the reality is that many databases built with nonstandard terminology now exist, and in any case standardization efforts will always seriously lag behind the need for terms defined with the precision required by data dictionaries.

Names and Values

It is not always obvious what should be a name of a fieldtype and what should be a data value. For numeric data it is usually straightforward: a fieldname of "modulus" and a value of "42.1" is unambiguous, but there is also the problem of different ways of expressing the same thing, even using the same names (Fig. 8.3).

Here the terms "bar," "tube," and so forth, are fieldnames in the first example, but data in the second. This also demonstrates that very similar databases, such as those represented by these two tables (Fig. 8.3), may find it impossible to communicate unless there is a defined means to convert from one database logical structure (schema) to the other. This currently cannot be done automatically even for the simple case illustrated here. This thus implies personal involvement of human users in materials data transfer for some time to come. The only alternative is the widespread adoption of STEP-standard database schemas and a wide use of EXPRESS to define them, although it should be noted that the two tables shown here would have different EXPRESS descriptions that would be incompatible.

Capability for General Expression

In devising a data transfer format to be used between databases it is necessary to ensure that the format is at least as capable of expressing relationships as are any of the participating databases. Many formats have been proposed for materials data, but the specification of the degree of this capability, of the degree of associativity, is often not well defined, In some cases an arbitrary degree of "nesting" is permitted, and it is naively assumed that this is generally adequate. What is required is some more formal argument that will convince potential users that the format's capabilities have been precisely specified.

Functional Dependency

Functional dependency is the dependence of the value of a field in a row on the values of other fields in the same row. This is a property of the real world, which is being described by the database and cannot even in principle be observed from only the data. Any dependence is an unchanging aspect of the database, like the names of the fields and the tables, not an ephemeral, such as the values in the tables at any one instant of time.

Table 8.2 shows a functional dependency of hardness of degree of plastic strain (such as might be applied by repeated cold-rolling). There is a dependence of hardness on plastic strain that is fairly obvious from the data alone. There is a multifunctional dependence in Table 8.3, which is not obvious, and in fact cannot be deduced from the data alone, that is, the mutual and complex dependence of hardness on yield stress where the proportionality of the dependence also depends on the plastic strain.

Thus a simple transmission of just a set of data tables is inadequate because the functional dependencies are not represented. However a set of data tables together with a set of catalogue tables can represent the necessary associativity.

Materials Identification

In materials databases there is usually only one commonly assumed dependency, the dependence of a material's properties on its designation or identification. Because all dependencies are forced into this one mold, the definitions of what are needed in a materials identification become more and more complex, including such things as materials supplier, heat-treatment schedules, chemical composition, age and so on, when many of these "designatory" properties could be more clearly represented as metadata for a particular property measurement. Thus material subjected to different degrees of plastic strain (Table 8.2) would usually be given a different designation for each strain to remove all functional dependency between nonkey fields, but this trick would not work for Table 8.3 where there is multifunction dependency.

TABLE 8.2—Example of functional dependency.

Material	Hardness, MPa	Plastic Strain, %
Copper	424	0
Copper	752	25
Copper	1334	48

What is needed for materials data transfer is a way of representing the dependencies explicitly in addition to the names and values. (Another problem with materials data is that in many cases, although the data are known and measured to some degree of accuracy, the dependencies are unknown, for example, does polymer creep-rate depend on humidity or not? Should polymers with different water contents be identified as different materials?)

TABLE-BASED FORMATS

A relational database consisting of several tables and integrity constraints is capable of representing any logical set of relationships between data that is describable by the relational calculus [*12*]. This is a very wide capability indeed. It follows that a set of tables of data, communicated as a data transfer format, has the same representational capability. However there are many different ways of representing the same data in tables, so simple table formats are always multivalent (as defined earlier). Another difficulty is that the relationships remain implicit in the tables. There is no standard way of representing dependencies even if the tables are normalized up to 5th Normal Form [*12*]. Thus complex relationships have to be documented to accompany any multitabular data transfer format (an attempt is described at the end of this chapter).

Single tables are very important in practice, they are the single most commonly used basic medium of data transfer. Thus sometimes it is appropriate to denormalize an existing relational database and to generate from several tables a sin-

TABLE 8.3—Example of multi-functional dependency.

Material	Hardness, MPa	Plastic Strain, %	Yield Stress, MPa
Copper	424	0	60
Copper	752	23	134
Copper	1334	48	476

gle table containing many duplicated fields, which is then transferred to another system. This is potentially dangerous because functional dependency can become implicit. This is an example of a general case: a transfer format can often use repetition of data as a strategy to avoid complexity in the format itself.

SAE Aerospace Standard 4159

This aerospace standard has been submitted to ANSI and proposes a standard way of encoding a single table of materials data [*22*]. This imprecise document also defines a new type of datafile, a "table file," which contains a normalized table of numerical data (1st Normal Form) packed into 80 character lines (card images) with initial label fields in accordance with MIL-STD-1840A (labels that merely describe the source and destination of the data). After the labels, this table file has a header containing a list of fieldnames and then the numerical data.

It imposes the requirement that the table be in "1st Normal Form" [*12*], so it resembles a single xBase file (see below) in its capability for representing relationships. There is no defined method for associating several tables. It also defines syntax for the addition of footnotes, which means that it is really only suitable for information oriented directly at human beings who can read and interpret the footnotes. Table 8.4 is an example taken from AS 4159 [*22*].

xBase and Equivalent Tabular Data Transfer Formats

Relational databases can always be thought of as (and are designed to be viewed as) a set of tables with labeled column headings and unlabeled rows together with a set of integrity constraints. Normalized tables always permit only a single value at each location in the table, and column headings (fieldnames) have no units or other attributes. They are just labels (1st Normal Form [*12*]). All other information must be stored in other associated normalized tables.

xBase

The xBase database file format is simply a means of encoding a single table as a single file of bytes, together with a large number of implementation-related detailed restrictions. The format (the ".dbf" file) is now the subject of standardization activity by an industry consortium, but it originated with a proprietary software system and was originally defined for files only on one type of personal computer. However, it is equally valid for any computer that regards files as a simple sequence of bytes. A set of xBase files, in which common fieldnames make cross-references, can contain data with any degree of associativity expressible in the relational calculus. The xBase language however does not provide integrity and consistency support, and functional dependency can only be expressed through hand-built catalogue files. xBase is typical of a great number of proprietary database file formats, and all have similar types of facilities and restrictions. Since xBase is soon to be a standard, and since it is so typical, it is worthwhile describing some of its restrictions in some detail.

xBase limits the total size of any file, the total number of fields, and the total size of each record. It imposes on the user the need to classify each column ("field") precisely in terms of the type of value to be stored and its maximum length (in bytes). Values must be classified as character, fixed-numeric, logical, date, or "memo." Character strings must be less than 254 bytes long and must not contain ASCII "NULL" characters. Numeric values must be less than 19 digits and exponential format is not allowed. Integers and floating point fields are not interchangeable. Free-form text of up to 4000 bytes can be stored "in" a memo field, but the text is actually kept in a separate file; the datafile contains a pointer to it. Most important for materials data, xBase does not permit any kind of NULL value for numeric fields.

Criteria for Tabular Formats

The criteria for a useful plain-text tabular format [*8*] are as follows

(1) The simplest useful format.
(2) A flat-file table-like representation.
(3) Convertible to and from xBase .dbf files.
(4) Plain-text to aid editing and word processing.
(5) Possible to include as an external reference in an SGML document.
(6) Compatible with being made conformant to MIL-STD-1840A.
(7) Designed for machine-readability, not just a way to represent tables of data for presentation to people.
(8) Extensible to arbitrary complexity.

The commercially supported database format (xBase) is already being used by many materials properties database managers to upload test-data into their databases. The main disadvantage of the xBase format is that it is defined in binary, and the contents of xBase files are not easily viewed or edited, especially not if the files are transferred to minicomputers or mainframes.

The argument for supporting a simplest possible useful format is that, whatever happens, people will use a simplest-possible format because using formal international standards has very high overheads in programmer-training and software cost. If such formats are going to be used anyway, it makes sense to design one in such a way that it can be

TABLE 8.4—Example of AS4159.

UNS	ASTM	Form	min_D	max_D	X-Area	Tensile	Yield	Elong
M11311	AZ31B	bars	0	6.32	all	241	145	7
M11311	AZ31B	bars	6.35	38.07	all	241	152	7
M11311	AZ31B	bars	38.10	63.47	all	234	152	7

(1) used as part of existing standards schemes (for example, SGML: ISO 8879-1986) and (2) extended smoothly to the higher functionality required by other users.

Simplicity and User-Editing

The requirement to be simple for users to edit by hand has several implications. First, there must be no arbitrary "counts," such as numbers of records or lengths of strings to be typed in. It is awkward to have to re-edit a number at the beginning of the file whenever a spelling mistake is corrected near the end. The user should not have to make unnecessary decisions about the data that are not relevant to his own work, such as how long to allow for the length of text fields or the number of places of decimals required for numerics.

There should be no artificial and arbitrary limits on what can be typed so long as it obviously makes sense. Software is quite capable of counting the number of fields, of distinguishing between strings and numbers, of measuring the length of the longest string, and of recognizing a wide variety of number formats; so it makes no sense to impose these tasks on the user.

In a plain-text translation of xBase data files, data consist of a number of "tuples" (rows or records) where each tuple contains one value for each field name. The world "tuple" is used because a table's row structure may not be evident in the data transfer structure. The general lesson here is that a minimum of formal formatting should be required.

The numeric and string types can be automatically recognized by the values that appear. There is a slight danger that a single typing error, such as O ("oh") for 0 ("zero"), may cause an entire set of numbers to be classified as strings, but the translation software should be aware of this possibility and produce appropriate warning messages [*23*].

Restrictions

We must apply a number of restrictions if we are to maintain free interconversion with xBase files [*8*]. The most notable is possibly that which limits fieldnames to being only 10 characters long. Some of these restrictions are awkward and unpleasant, but rather than extend them immediately, it makes sense to define a base format and a separate, upwardly compatible extended format that removes the restrictions. A great deal may be gained by having a freely xBase-interconvertible format because of its industrial prevalence.

xBase does not permit NULL values for numeric (or date) fields, although strings can contain the empty string, and "logicals" can be "unset" (neither True not False). This lack of NULLS is a severe disadvantage for materials property data where absent values are very common. With xBase the only way to represent NULL is to associate every numeric field with another field that contains a value determining whether the numeric is NULL or not (which is how databases that do support NULLS actually do it). An alternative "workaround" is to use special numbers which will "never" appear in real data such as 0.0, −999, and so on, but this is dangerous.[5]

The general lesson is to be aware of the NULL value problem when using xBase-like formats and to try to use such formats in a simple, unsophisticated way because this will aid format conversion later.

Multiple Tuples

In relational databases it is essential for tuples (rows, lines) to be distinct. Multiple tuples have no meaning, and even in cases where they can be entered into a database they cannot then be retrieved. Unlike relational databases, tabular formats permit several tuples to be the same. This is actually a common occurrence in experimental data and only causes problems if the data communication is going to be immediately and automatically loaded into a relational database on receipt. If it does cause problems then the cure is simple: add a field to all tuples that consist of a sequence number.

Alternative Presentations

The freedom to change the layout independently of the content results from permitting a variety of separator characters and of making multiple separators mean the same as a single separator.

Multitabular Format

There are many drawbacks of the simple, single tabular exchange format even without considering the restrictions imposed by xBase compatibility. These are as follows:

(1) Repeated data are repeated in the transmission.
(2) Single-point data require a whole file to itself.
(3) The available associativity is only very simple [*7,23*].

These can be alleviated by using multiple data files to describe the same set of data, a technique known as "normalization." If all the data are sent (or stored) in a single table as defined above then it is "simply normalized" or "in 1st normal form," further normalization removes redundancy without losing information. The principle involved is very simple, that is, any one item of information should only be represented once.

The simplest extension to the simple tabular format is to add an extra definition file to a set of single-table files encoded according to the simple tabular form. No meaning should be ascribed to the order in which the filenames or the tables appear [*8*].

Integrity Restriction

Using multiple tables to describe a single set of data gives more capabilities but also introduces a few more possibilities of error. The tables are linked by having some fields in common. Therefore every table must have at least one field in common with another table, though not necessarily the same field. If a table has no common fields then it is really describing a distinct set of data and should not be referenced by the definition file.

---

[5] Many materials in some databases acquire melting points of 32°F because unavailable values have been set to zero degrees, interpreted as Celsius by default.

Extended Format

An extended format, with the requirement for xBase compatibility removed, should have the following additional capabilities:

(1) Fieldnames can be arbitrarily long, can be in upper and lower case (but "Strength" is not distinct from "strength"), and can contain any characters apart from quotes.
(2) String values can be arbitrarily long.
(3) Tuples can contain any number of fields.
(4) There is no limit on the number of tuples in a data file.
(5) Numerics can contain any reasonable, unambiguous representation of numbers, for example, ".1e-0005."
(6) Numerics can be of any magnitude, that is, less than 1e-17, and more than 1e19.
(7) Comments can be inserted in the datafiles wherever whitespace is permitted. Comments can be nested.
(8) NULLS are permitted and can be interpreted as a valid string (distinct from the empty string) and as a valid numeric (distinct from zero).
(9) A title and description can be included in the data file.
(10) Units can be described more straightforwardly.

This is a general list of requirements that cannot be met by most common xBase-like formats. If an extended format is required, then this presents reasonable criteria for development.

Since the extended format can describe data not expressible in xBase format, information could be lost if data files of this type were translated into xBase. The extended format is therefore intended as a transfer format in its own right, to be used between software and databases that contain direct translators. This implies that the range of software available that could use this more powerful format would be smaller than the range available for the xBase compatible version.

It is awkward to have to set up a distinct table to contain the units information. So in the extended format, a "unit list" might be developed. Such a structure would contain the same number of strings as the "field list," with each string containing the unit information for the respective field name.

Multitabular capabilities can be added to the extended format by adding a structure for collections of tables in a definition file. Extensions to the definition file format itself consist only of the ability to include comments. Note that comments should be avoided if it is at all possible to put the same information in properly defined fields. This is particularly noteworthy because comments are often used to store informal quality information (such as "this data no good"), which really should be represented explicitly.

A new restriction is necessary when combining the multitable facility with the possibility of NULL values. For the fields common to more than one table it is important that no values in any of the tables are NULL. A moment's thought will show that merely because two distinct sets of data (tuples) are lacking a data point (that is, have a NULL value somewhere) does not mean that they should be linked, as they would be on any other common value [*12,13*]. This is an example where two simple extensions to a transfer format have an unpleasant synergy.

Conclusions on Table-Based Formats

The dominance of xBase-like binary formats in personal computer systems has prompted an analysis of their restrictions for materials data transfer. Text versions of these formats are entirely feasible and remove many of the difficulties. There are a number of simple extensions to single-table formats, such as using multiple tables and implementing NULL values, which while innocent and useful in themselves lead to problems in combination.

A general lesson is that while very simple formats are possible, the simpler the format the greater the load placed on software to produce it and, especially, interpret it. Developing software (even if it is just spreadsheet macros) is expensive. Therefore simple formats have hidden costs.

ITEM-BASED FORMATS

The basic principle that distinguishes item-based formats is that they start by associating one name with one value and then extend the concept to lists of values and sometimes lists of names by making additions to the permitted syntax. Item-based formats are thus based on the NAME = VALUE system where the VALUE can, in some formats, consist of lists and other name/value pairs, that is, NAME=(VALUE, VALUE, VALUE. .) or NAME=(NAME=VALUE, NAME=VALUE, . .). If only value-lists are permitted then the structure of the associations can only be tree-like (branching from a common root) although several distinct trees in any set of data usually permitted (one per "record").

Modern object-oriented data modeling techniques are item-based. They add extra operators, such as inheritance, which significantly complicate the semantics and greatly increase the required preciseness of definitions. They will be required for future generations of materials concept-ontologies on which interoperating "active" data transfer systems will be based [*16*], but currently they are research tools only.

Possible Associativities

Item formats, since they resemble linear strings, can be thought of as being one-dimensional (1D) compared with the clearly two-dimensional (2D) nature of tabular formats. The reason for the emphasis on item formats is that a simple item format is simpler than a simple tabular format. However, such simple formats have inadequate expressiveness for the associativity required for materials information. What is required is some way of generalizing from 1D format more dimensions without losing the simplicity of the original 1D.

More general associativities, such as those possible with relational databases, can be thought of as multiple-connected lattices rather than trees. This capability can be achieved in a list-oriented (item-based) transfer format or data model only if "reference names" can be used that allow cross-reference between distinct trees. This corresponds to a transfer format where a name is permitted anywhere a value is required. There are other common cases that even this technique cannot handle.

The above discussion of trees, lattices, and tables is a re-

TABLE 8.5—Material designation guidelines.

Characterization
  compositional detail = ...
  phase composition = ...
  elemental symbols = ...
  measured weight % = ...
  minimum weight % = ...
  maximum weight % = ...
Source
  manufacturer = ...
  country of origin = ...

capitulation of the historical comparisons of hierarchical, network, and relational databases, but from the point of view of their descriptive power for materials information. The database aspects are discussed more fully in textbooks [*12,13*].

Partial Identity

Partially synonymous terms are always a problem in materials databases. For example, "elastic limit" and "fracture" are identical for ceramics, but distinct for metals and polymers (because they are plastic and not brittle). Any materials database system or transfer system based only on globally-defined names cannot handle this problem unless the syntax allows some way for the meaning or interpretation of the name to be modified under certain circumstances. Relational databases, because the tables are treated as mathematical sets, must consider all names to be defined "at the same level."

This is an area where a strictly tree-structured transfer format (or database) has no problems. The meaning of a term can be easily redefined to apply to all subtrees from that point. However as soon as cross references are added, then two policies for interpretation are possible: Is the meaning of a term (1) that which derives from the access route (via any cross reference) or (2) that which derives from its position as declared in the tree? Both policies have sufficient drawbacks so that most practical databases find both to be useless.

Item Formats

The simplest item format is one where there is a single list of fieldnames, each of which is permitted precisely one value. Usually, however, only one fieldname need have a single value (the "key" in relational database terms), and multiple values are both useful and sensible. A good example are the fields from ASTM E-39's generic guidelines for materials designation as shown in Table 8.5.

TABLE 8.6—ASCII data recording format.

```
SPEC
"AMS", "4975", " "
SPEC
"MIL", "F-83142", " "
COMP
"Al", 5.5, 6.5
COMP
"Sn", 1.8, 2.2
COMP
"Zr", 3.6, 4.4
FORM
"Wrought"
```

TABLE 8.7—Implicitly nested data (retabulated for clarity).

```
PROP
"CmpYldStr", "MPa", "ksi", 0, 0.145038
  COND
  20, "air"
  VALU
  1075, "t", "AA"

  COND
  205, "air"
  VALU
  800, "t", "AA"

  COND
  371, " "
  VALU
  695, "t", "AA"
```

Table 8.6 shows a fragment showing a NAME=(VALUE, VALUE. .) type of format used by ASM. Here multiple values require repetition of the fieldname, and values are separated by newline characters, not by = symbols [*24*]. The fieldnames, SPEC and COMP (specification and composition), are built-in to the format itself. This is typical of most item-formats in that they only allow predefined names entities to have values, where the values must conform to a predefined implicit structure.

This same format also has NAME=(NAME=VALUE, NAME=VALUE, . .) features in its representation of experimental data plots. A sequence of numeric data is associated with the most recent declaration of a property: PROP=(COND=20, Cmp YldStr=1075, . .) (Table 8.7).

This format also has open-world capability, in addition to the closed-world nature of the earlier lists of predefined names. It permits the declaration of a new property (and metadata) [*24*].

These types of item-based format are very capable for representing a particular database's archive data, but are almost useless for transferring data between different database systems with different schemas because the definition of the source database's schema is written into the format in an indelible and implicit manner. Not only is specific translation software required to transfer data from this format into a recipient database, but the format itself does not document its own implicit data dependencies. This makes the writing of translation software dependent on the informal documentation of the format (in English) rather than on a description written in a formal grammar. The situation would be improved if format originators could produce formal documentation of their format written in EXPRESS.

Problems of Complexity

It is possible to extend an item-format to arbitrary levels of dimensionality by introducing series of upwards-compatible additions to the format definition, but every level of capability adds increasing complexity to such a degree, in both concepts and extra syntax, that it becomes harder and harder to understand and to use. On the other hand, tabular (relational) formats have the advantage of a decade or more experience with relational databases [7].

Conclusions on Item-Based Formats

Since item-based formats start by associating one name with one value and then extend the concept to allow lists of values, they are well-suited to situations where the data consist of a few, individual values. The linear or tree-like structure, especially if extended using cross-references to subtrees, is perfectly adequate for many types of simple data.

Thus unless the designer of a general transfer format is prepared at the outset to handle the complexity of extending one-dimensional to higher dimension [7], it is recommended that all item-based formats be avoided unless their requirements are (and will remain) extremely modest. Far better is to use EXPRESS to formally define the fields for the item-based format, and then rely on software produced for use with ISO 10303 compliant systems to handle the transformation into a physical file format, and retransformation back into a relational database.

CATALOGUE-BASED FORMATS

As we have seen with tabular formats there is a difficulty in representing functional dependencies. In both item-based and tabular formats there is a difficulty in agreeing fieldname definition. This section outlines an example of a table-based, textual method for transferring materials property information including data structure and definitions. The principle is that there are a very small number of pre-defined tables with predefined columns (fields) and that these represent at a very abstract level the concepts that the data are based on [*25,26*]. These defined tables are used to define more specific fieldnames, which then are used in tables which both define the numerical data and the relationships between data and metadata.

First, and perhaps surprisingly, we need to be sure that we all mean the same things when we use the same names and labels. Hence the first thing here is a table that relates material identifications to a textual description, followed by a table that does the same thing for properties (including defining the units to be used throughout a set of data). These are catalogue tables (Tables 8.8 and 8.9). They are given unique names of the form #TABLE-NAME# (the descriptive text would of course be much longer in practice).

Note that there is no mention of numbers of places of decimals, precision, the differences between character and numeric data here. The software that interprets this data and loads it in to the receiving database has to do that in any case. This is all ASCII text.

Table 8.10 shows some actual data (abstracted from several undergraduate textbooks). Real data will be incomplete, that is, sometimes only a lower bound given or sometimes only a typical estimate. This table also shows several methods for representing precision. Note that this way of representing properties, as values in a table rather than as fieldnames, has several advantages, which increase flexibility. It is the abstract closed-world method.

Table 8.11 shows how inheritance information can be represented using tables: narrower and broader terms as in *McCarthy's Data Thesaurus* [*21,27*]. The tables here are very rudimentary, a third catalogue table (in addition to properties and materials) should be used to define the relationships properly.

TABLE 8.8—Material catalogue tables.

#Material-Catalogue#

Short-Name	Long-Name
metal	all pure metals and alloys, but not MMCs
cu-alloys	brasses, bronzes, etc.
brasses	copper and zinc alloys
st-steels	stainless steels (austenitic)
al-alloys	aluminum alloys
mld-steel	mild steels
PE	polythene: low and high density
LDPE	low density polythene

Sets of catalogue tables are also possible to represent as follows:

- two or more related properties for one material,
- one property that depends on more than one material (for example, frictional coefficient),
- a function of an independent variable, and
- experimental data taking into account the relationships between specimen, test, material, and sample.

It is very likely that tables will be managed by systems that use the relational algebra so fieldnames must be unique within a table. If we have a table that relates two properties then either we "normalize the tables" and generate many little fields whose only job is to kink up ones we already have [*12*], or synonyms have to be declared.

Now it should be possible to see the need for a super-catalogue of everything that appears in the set of data (Table 8.12). It is so important that as well as having a unique table name, its fields are also predefined. The intention here is that everything of the form #. . .# describes the structure of the set of data and is used identically by everyone. The idea is that any field that is not named in the #. . .# form could be entirely different from database to database, but they would still be able to transfer data.

EXPRESS

The data modeling here is done entirely in terms of tables that represent the data, the metadata, and the relationships and functional dependences between them. Since they are

TABLE 8.9—Property catalogue tables.

#Property-Catalogue#

Property	Long-Property	Units
price	bulk selling, semi-finished, 1985	sterling/tonne
yield	yield point or fracture	MPa
ductility	ductility from yield to fracture	dimensionless
density	Archimedes method	tonne/m^3
stiffness	Youngs modulus in compression	GPa
E-11	orthotropic in-plane stiffness	GPa
E-12	orthotropic in-plane stiffness	GPa
E-22	orthotropic in-plane stiffness	GPa

TABLE 8.10—A data table using defined terms.

#Data-Table#

Short-Name	Property	Typical	Low-Bound	High-Bound	± Range	*/Factor
st-steels	price	1245	1100	1400	150	1.128
brasses	price	899	750	1062	156	1.190
st-steels	yield	386	286	500	107	1.322
brasses	yield	350	60	960	450	4.000
st-steels	density	7.8	7.5	8.1	0.3	1.039
brasses	density	8.1	7.2	9.0	0.9	1.118
st-steels	stiffness	195	190	200	5	1.026
brasses	stiffness	135	120	150	15	1.118
st-steels	ductility	0.55	0.45	0.65	0.10	1.202
brasses	ductility	0.14	0.01	0.55	0.27	7.416

just tables, any multitabular transfer format would be adequate for their communication, including sets of spreadsheets and xBase files. However, in planning the structure of the relationship tables a more abstract data modeling tool is really required, one that explicitly represents functional relationships. EXPRESS and its graphical design notation, EXPRESS-G, are appropriate tools for future generation of catalogue-based formats.

CONCLUSIONS ON DATA TRANSFER

Item, Table, and Catalogue Formats

It has been shown that it is easier to demonstrate the expressiveness of tabular formats compared with the expressiveness of item-based formats, but an item-based format with the right kind of formally specified grammar can be capable of all required associativities.

Both item and table types of format can theoretically handle the associativity required (though individual formats always have limitations). However, only the most complex of the existing proposals for item-based formats are able to duplicate the expressive power of the simplest multitabular format. The item-based formats can (theoretically) handle data dependencies directly whereas table-formats require explicit catalogues.

The catalogue-based format is an extension of the table format and relies on more tables to describe the information in, and the relationships between, the tables containing the data.

Because they derive from different needs, it is inevitable that if only one of these types (item-based or table-based) were to be adopted by a standards-making body, some user-communities would then develop the other type for their own use. It makes sense to forestall such a divergence of standards by trying to plan a system where both types can coexist and are inter-translatable from the start. This could only be achieved by looking for commonality at a "higher level" than that of the syntax, that is, the level of interpretation and meaning of the fieldnames, such as provided by catalogues. Thus a common system could be built only on a foundation of a catalogue containing a common reference glossary, data-dictionary, and structured data thesaurus. This could profitably use the ISO 10303 materials product definitions as a starting point and use EXPRESS as a means for definition and communication.

TABLE 8.11—A table defining material designation sub-terms.

#Material-Subclasses#

Short-Name	Special
al-alloys	al-6061-t6
al-alloys	al-1100
metal	al-alloys
metal	st-steels
metal	cu-alloys
cu-alloys	brasses
brasses	Cu-60/Zn-40
material	metal

General Conclusions

At present no standardized format for general materials data, index, or catalogue transfer exists. ISO 10303 and ED-

TABLE 8.12—Defined fieldnames.

#Field-Descriptions#

#field#	#description#
#field#	the fieldnames of all the columns of all the tables of data
#description#	description of a fieldname
#use-as#	similar to "special" but for field; needed to prevent duplicate fieldnames
short-name	material short identification
long-material-name	material long identification name
long-property-name	property long identification name
property	a property of a material, or several materials interacting
units	units for a material property; if none then "dimensionless"
typical	typical value for a material property
low-bound	low bound for a material property, exptl. or specified
high-bound	upper bound for a material property, exptl. or specified
± range	usual plus/minus range of typical values for a property
*/factor	usual mult./divide factor range for typical values for a property
special	specialization relationship between materials, requires #use-as# definition

IFACT are international standards that can handle some very restricted types of materials information transfer in some particular situations. ASTM E-49 is continuing to work in this area, and there are several industrial consortia in the United Kingdom (supported in part by the Department of Trade and Industry) and the United States (supported largely by defense contacts, notably the PAS-C consortium concentrating on polymer-matrix composite materials) developing materials information transfer techniques using EXPRESS and software developed to support ISO 10303.

On a smaller scale, the province of tabular spreadsheets and xBase-type software packages, there are a plethora of proprietary "uploading" and "downloading" formats specific to individual materials databases. This chapter has shown some pitfalls and opportunities in using such simple systems.

APPENDIX

EDIFACT Code Sets

This Appendix is included because it is very difficult to discover precisely how EDIFACT functions without access to EDIFACT standard working documents. Without this kind of information it is impossible to make a rational evaluation of EDIFACT in comparison with other formats.

Segments are denoted by a 3-letter code and elements by a 1-letter and 3-digit code. This example shows initially one of the groups of segments from the draft "Quality Data Message" for describing goods, items or services, either directly to a product or to a batch of delivered items. This Group is then followed by an expansion of one of these segments (MEA Measurements) in terms of its component elements:

---	Segment Group 5 --- (100 repetitions permitted)
MEA	Measurements
DTM	Date/Time References
RFF	References

Now the description of the MEA segment. Details are from the EDIFACT Segments Directory issue 88.1. "an. .3" means an alphanumeric code of between 1 and 3 characters long and "n. .15" means a series of between 1 and 15 digits.

MEA	Measurements	
6310	Measurement Specification Identifier	an. .3
6312	Measured Dimension Identifier	an. .3
C174	VALUE/RANGE	
6410	Measure Unit Specifier	an. .3
6314	Measurement Value	n. .15
6162	Range Minimum	n. .15
6152	Range Maximum	n. .15
6320	Measurement Significant Code	an. .2
6154	Measurement Attribute Code	an. .2

Now the list of the allowable values for these 2 and 3-letter codes is given for the elements 6310 (only first few elements given), 6312, 6410, 6320, and 6154 taken from the EDIFACT Code Sets Directory issue 88.1. Note that there is duplication in some of the meanings in the values allowed for these last two elements; this could be because the information is taken from a draft of the directory. This is a very small fraction of the EDIFACT directories, but it should give a flavor of how the standard works.

6310	Measurement specification identifier
BL	Bundle Limitation
BZ	Batten size
CH	Chemistry
CN	Core Notch Dimensions
DT	Dimensional Tolerance
LM	Layer of Multi-layer Product
PC	Parting Cut (Sawcut)
PD	Physical Dimensions (Product Ordered)
RL	Receiving Facility Limitations
SH	Shipping Tolerance
SR	Surface Roughness
...	......

6312	Measured Dimension Identifier
01	Unit Net Weight
02	Unit Gross Weight
03	Total Net Weight
04	Total Gross Weight
05	Net Net Weight
A	Consolidated Weight
AZ	Arbor Size
B	Billed Weight
BO	Lateral Bow (Camber)
C	Actual New Repeated for Combination

6410	Measure Unit Specifier
BAG	Bags
BL	Barrel
BLI	Blister
BLK	Block
BOT	Bottle
BOX	Box
BRD	Board
CAN	Can
CAR	Carboy
CAS	Case

6320	Measurement significance code
04	Approximately
05	Equal to
06	Greater than or equal to
07	Greater than
08	Less than
09	Less than or equal to
10	Observed value
11	Trace
12	True value
13	Not equal to

6154	Measurement attribute code
01	Clear
02	Hazy
03	Excess

04	Some
05	Undetectable
06	Trace
10	Present
16	Nil
23	Absent
30	Less than
31	Greater than

REFERENCES

[1] Westbrook, J. H. and Grattidge, W., "The Role of Metadata in the Design and Operation of a Materials Database," *Computerization and Networking of Materials Databases: Second Volume, STP 1106*, J. G. Kaufman and J. S. Glazman, Eds., American Society for Testing and Materials, Philadelphia, 1991, pp. 84–102.

[2] Sidhu, G. S., Andrews, R. F., and Oppenheimer, A. B., *Inside AppleTalk®*, 2nd ed., Addision-Wesley Publishing Co. Inc., ISBN 0-201-55021-0, 1990.

[3] Buttner, P. and Kröckel, H. *Materials Databanks*, Report of a VAMAS Workshop on Standards for Materials Databanks, CEC Joint Research Centre, Petten, The Netherlands, 15–17 Nov., 1988. VAMAS Technical Report 4, 1989.

[4] Kröckel, H., Reynard, K., and Steven, G., Eds., *Factual Material Databanks, Proceedings of the CEC Workshop at JRC Petten, Nov. 1984*, Netherlands Report EUR 9768EN Publishing Office of Official Publications of the European Communities, Luxembourg, 1985, pp. 109–111.

[5] Westbrook, J. H., Behrens, H., Dathe, G., and Iwata, S., Eds., *Material Data Systems for Engineering, Proceedings of the CODATA Workshop*, Schluschsee, Germany ISBN 3-88127-100-7, 1986, pp. 78–80, 110.

[6] Kröckel, H., Reynard, K., and Rumble, J., Eds., *Factual Material Databanks: The Need for Standards*, Report of VAMAS Technical Working Area 10, Publ., National Institute for Science and Technology, Gaithersburg, MD 20899, 1987; VAMAS, Versailles Project on Advanced Materials and Standards, secretary Dr. B. Steiner, National Institute for Science and Technology, Gaithersburg, MD 20899.

[7] Sargent, P. M., *Materials Information for CAD/CAM*, Butterworth-Heinemann Publishers, Oxford, United Kingdom, ISBN 0-7506-0277-5; and Sargent, P. M., *Data Quality in Materials Information Systems Computer Aided Design Journal*, Vol. 24, 1992, No. 9, pp. 477–490.

[8] Sargent, P. M, "Materials Data Interchange for Component Manufacture, Engineering with Computers," Vol. 6, No. 4, 1990, pp. 237–247; Sargent, P. M., "Technical Data Interchange Using Tabular Formats," *Journal of Chemical Information and Computer Sciences*, Vol. 31, 1990, pp. 297–300.

[9] Colton, J. S., "The Design and Implementation of a Relational Materials Property Data Base," *Engineering with Computers*, Vol. **4**, Nos. 1, 2, 1988, pp. 87–97 (Springer-Verlag).

[10] Bamkin, R. J. and MacRae, S. C. F., "Experience of Designing a Relational Materials Database," *Computerization and Networking of Materials Database, Third Volume, STP 1140*, T. I. Barry and K. W. Reynard, Eds., American Society for Testing and Materials, Philadelphia, 1992.

[11] Date, C. J., *A Guide to the SQL Standard*, Addison-Wesley Publishing Company, Reading, MA, ISBN 0-201-05777-8.

[12] Date, C. J., *An Introduction to Database Systems*, 5th ed., Vols I 1990; and II 1988, Addison-Wesley Publ Co., Reading, MA, ISBN 0-201-51381-1, 1988.

[13] Rumble, J. R. and Smith, F. J., *Database Systems in Science and Engineering*, Adam Hilger, 1990, ISBN 0-7503-0048-5.

[14] Nirenberg, S., Monarch, I., Kaufmann, T., Nirenberg, I., and Carbonell, J., *Acquisition of Very Large Databases: Methodology, Tools and Applications*, Carnegie Mellon University Center for Machine Translation technical report CMU-CMT-88-108, 1988.

[15] Parsaye, K., Chignell, M., Khosafian, S., and Wong, H., *Intelligent Databases: Object-Oriented, Deductive and Hypermedia Technologies*, John Wiley and Sons, New York, ISBN 0-471-50345-2, 1989.

[16] Sargent, P. M., Downs, M., Greene, R., Rishel, D., and Subrahmanian, E., "Materials Information and Conceptual Data Modelling," *Computerization and Networking of Materials Databases: Third Volume, ASTM STP 1140*, Thomas I. Barry and Keith W. Reynard, Eds., American Society for Testing and Materials, Philadelphia, 1992.

[17] *Common Reference Vocabulary*, European Materials Demonstrator Programme, copies available from Mr. G. Heine, DG XIII/B/2, Batiment Jean Monnet, Plateau du Kirchberg L-2920, Luxembourg, 1989.

[18] Kelly, A., "The Techno-Economics of New Materials," *Materials Architecture, Proceedings of the 10th Risø International Symposium on Metallurgy and Materials Science*, J. B. Bilde-Sørensen, N. Hansen, D. Juul Jensen, T. Leffers, H. Lilholt and O. B. Pederson, Eds., Risø National Laboratory, Roskilde, Denmark, ISBN 87-550-1551-4, 1989.

[19] "Making a Case for CASE," Product Focus Section, *Byte*, Dec. 1989, pp. 155–171.

[20] "Basic Encoding Rules for Abstract Syntax Notation One," ISO/TC97/SC21 N25 July 1984, and CCITT "Data Communication Networks: Message Handling Systems, Recommendations X.400-X.430," Vol. VIII-Fascicle VIII.7, Eighth Plenary Assembly, Malaga-Torremolinos, Oct. 1984. Geneva, ISBN 92-61-02361-4, 1985, pp. 62–93.

[21] McCarthy, J. L., "Information Systems Design for Materials Property Data," in *Computerization and Networking of Materials DataBases, STP 1017*, J. S. Glazman and J. R. Rumble Jr., Eds., American Society for Testing and Materials, Philadelphia, 1989, pp. 135–150.

[22] "Specification for an Automated Interchange of Standards Data," SAE Aerospace Standard AS 4159, issued 12 April 1988, submitted by SAE to ANSI.

[23] Sargent, P. M., *Associativity in Materials Property Data*, Cambridge University Engineering Dept. Technical Reports, Cambridge, England, 1989.

[24] Gall, T. L., Heller, M. E., and Cverna, F., "The Development of a Computerized Materials Properties Database Management System," *Computerization and Networking of Materials Databases: Second Volume, STP 1106*, J. G. Kaufman and J. S. Glazman, Eds., American Society for Testing and Materials, Philadelphia, 1991, pp. 161–167.

[25] Sargent, P. M., "Data Models and Data Dictionaries for Materials Information," *Proceedings of CAMSE'90, 1st International Conference Computer Applications to Materials Science and Engineering: Computer Aided Innovation of New Materials*, 28–31 Aug. 1990, Ikebukuro, Tokyo, Japan, 1990.

[26] Sargent, P. M., "Use of Abstraction in Creating Data Dictionaries for Materials Databanks," *Computerization and Networking of Materials Databases: Second Volume, STP 1106*, J. G. Kaufman and J. S. Glazman, Eds., American Society for Testing and Materials, Philadelphia, 1991, pp. 114–131.

[27] McCarthy, J. L., "The Automated Data Thesaurus: A New Tool for Scientific Information," 11th CODATA Conference, Karlsruhe, Sept. 26–29th, 1988.

9

Building a Model Database: EXPRESS Example

Edward Stanton[1]

OVERVIEW

To build a good database requires reliable data relevant to the user's application and a data architecture that efficiently loads the information into a database system. In this chapter we focus on the actual database building process and recommend other chapters of this book for additional important information on computerizing materials data. To allow this focus, we build a model database from test data from an automotive composite structure without reviewing the design requirements that led to this particular test matrix for automotive materials. In this model development we must deal with a complication not usually found in building numeric databases, namely, technical data requiring extensive metadata for safe use.

As Rumble and Smith [*1*] observe, material databases contain by their very nature heterogeneous property units, heterogeneous test metadata (usually ASTM designations), and technical footnotes all needed to clearly define conditions under which the properties data are valid. Even tensile strength has over ten ASTM tests defined for different material types and application environments. Westbrook and Grattidge argue convincingly that metadata are the most important part of any database, especially a materials database [*2*].

Defining the Application

The design application largely determines the relevant properties, the acceptable data sources, and the maintenance and quality control procedures that are appropriate for a particular material database. Sargent [*3*] provides an excellent review of these issues for material selection in design environments.

Here we take these to be properties and environments required by the Automotive Composites Consortium for structural design [*4*]. This lets us focus on building a model database recognizing that many other property sets and related metadata could be defined for other automotive composite applications.

Estimating Costs and Benefits

Several billion dollars are spent annually generating material data worldwide [*5*]. Organizing this expensive resource in computerized databases for engineering applications has many benefits and significant costs to consider. A computerized edition of MIL-HDBK-5, for example, with over 1300 tables and 7500 graphs took several years to build and to verify computer data entry working from the paper edition. Parameters significant to the cost of building a database include:

- The source data type, paper or electronic, and how the data are organized.
- The number of data tables and the number of properties, units, metadata and footnotes per table.
- The number of graphs and the number of graphical data types and footnotes per figure.
- The complexity of the database schema: how many entities and attributes per material, how are the entities related, and how closely does the schema match the organization of the source data.

In the case of the MIL-HDBK-5 example, about one man-hour per table and two man-hours per figure were required to develop and complete acceptance testing of the database from paper sources.

The benefits from computerized material database use include better access to relevant data, reduced engineering man-hours, reduced prototype testing, improved producibility and reduced maintenance costs for manufactured products. A benefit difficult to measure but of considerable significance is improvement in the quality of the material selection process, which is biased toward older more familiar materials when relevant property data for newer materials are not easily accessed.

DATABASE SOFTWARE AND HARDWARE

Rumble and Westbrook [*5*] noted at the Fairfield Glade Workshop in 1982 that software was the primary shortcoming in technology for computerizing materials data. Development of the EXPRESS language for modeling product data [*6*] and advances in the SQL language have improved things, but the statement is probably still true. Many databases are large "flatfiles" meaning there is no inheritance of

---

[1] Vice-president, PDA Engineering, 2975 Redhill Ave., Costa Mesa, CA 92626.

Copyright© 1993 by ASTM International www.astm.org

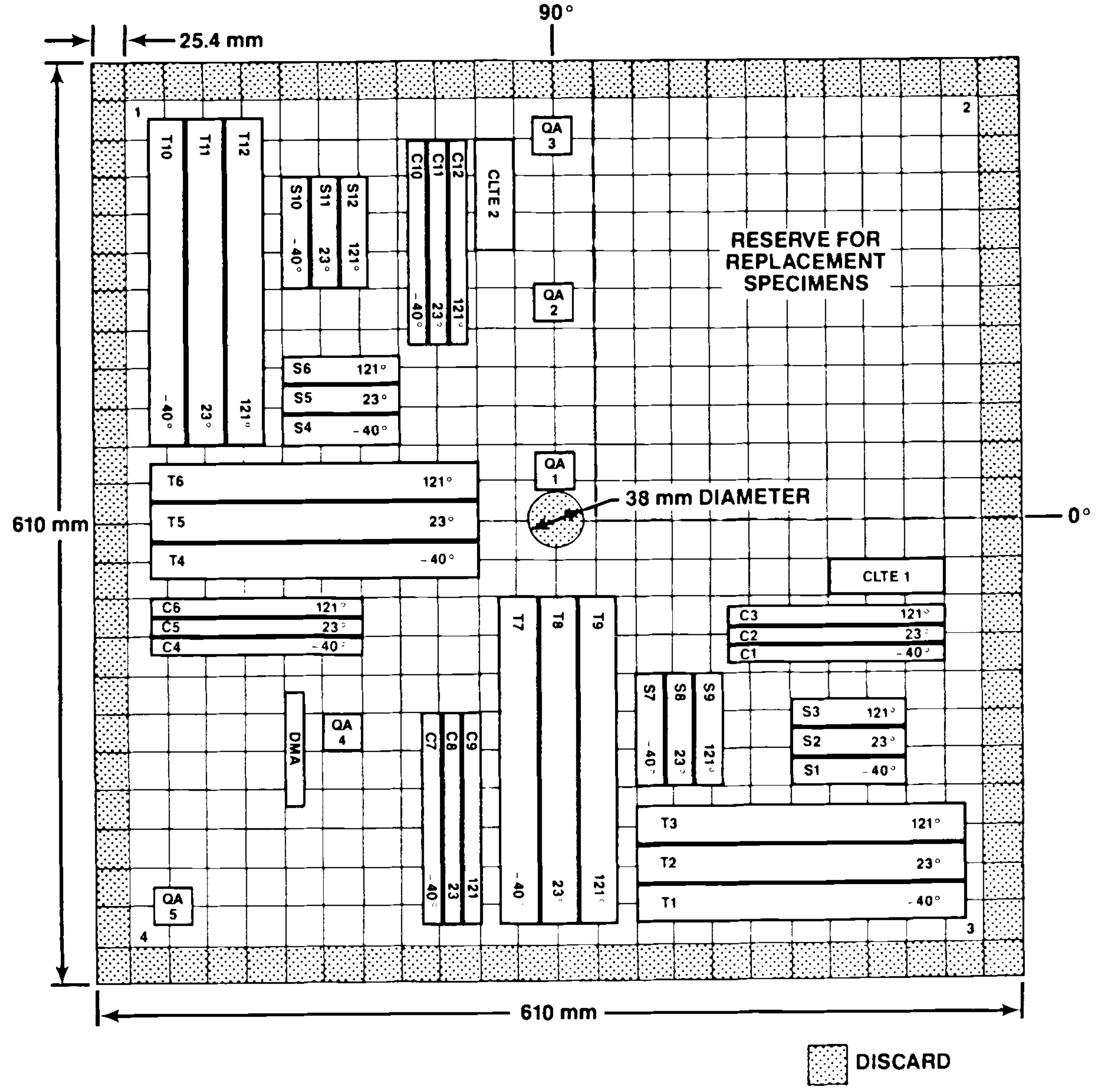


FIG. 9.1–ACC specimen location template.

attributes among entities. The introduction of object-oriented features with EXPRESS and similar languages made it easier to model material data and metadata efficiently. These features will be used in the model database architecture where, for example, attributes at the material product form level are inherited by specimens.

Hardware Systems

What criteria should your project or organization use in selecting hardware for material database applications? A difficult question. First, let us note that the hardware used in the building process very often will not be the hardware used in applications. In a remote access application a PC or even a smart graphics device will work if the database and query system are remote. If the database is remote and the query system is local but not required to interact with a computer-aided design (CAD) or computer-aided engineering (CAE) system then a PC or low end workstation is adequate. When the application is strongly interactive with CAD or CAE systems, or both, then hardware/software compatibility will dictate the choice.

Older flatfile database systems typically require a lot of memory and raw CPU power to load large handbook databases. These are "mainframe" computer systems with user terminals connected to the mainframe via a network file server. The MPD Network is a good example of this type hardware system [7]. Newer PC-based systems range from collections of product datasheet information to very sophisticated National Institute of Science and Technology (NIST) physical properties data collections loaded using relational database systems. Today most CAE work is done on workstations connected to a central file server or "disk farm," and material databases as large as mainframe handbook data-

TABLE 9.1—ACC Composite Material/Process Data Sheet.[a]

ASTM Name	Value	Units	ACC Name
Material Reference . . .	Dow 411-C50/CT U750	-0-	Composite Designation
*Material Class	Composite	-0-	-0-
*Matrix Class	Polymer	-0-	-0-
*Reinforcement Class	Fiber	-0-	-0-
Structural Detail	RTM Laminate	-0-	-0-
*Prec. Type	Dry Mat	-0-	-0-
*Prec. Name	U750	-0-	Product Code
*Prec. Manufacturer	Certainteed	-0-	Manufacturer
Prec. Matrix Gel Cond.	121	Deg C	Glass Transition Temperature (Tg)
Prec. Matrix Viscosity	100	cp	Viscosity
*Matrix Subclass	Thermoset	-0-	-0-
*Matrix Chemical . . .	Vinyl Ester	-0-	Resin Composition
*Matrix Commercial . . .	DOW 411-C50	-0-	Manufacturer
Matrix Density	1.12	g/cc	Specific Gravity (cured)
Matrix Strength	79.2	MPa	Neat Resin Tensile Strength (D638)
Matrix Modulus	3.30	GPa	Neat Resin Tensile Modulus (D638)
Matrix Strain at Failure	1.00	%	Neat Resin Elongation (D638)
*Reinf. Subclass	Continuous	mm	Fiber Length
*Reinf. Common Name	E Glass	-0-	Fiber Material Type
*Reinf. Chemical . . .	Borosilicate Glass	-0-	-0-
*Reinf. Form	Random Mat	-0-	-0-
Reinf. Density	2.60	g/cc	Specific Gravity
Reinf. Manufacturer	Certainteed	-0-	Manufacturer
Reinf. Yield	6633	g/km	Roving/Yarn Yield
Reinf. . . .	2500	-0-	Bundle Size/Splits
Reinf. Diameter	16.0	micro m	Filament Diameter
Reinf. Sizing	Siline KE6N850501	-0-	Chemical Size Description
Reinf. Binder	Thermal Plastic Polyester	-0-	Binder Description
Reinf. . . .	1.50	oz.CSM	Fiber Product Form
Process Stage Spec	Molding	-0-	Process
*Process Stage Type	RTM	-0-	Molding Process
*Processor	ACC	-0-	-0-
Process Date	19900210	YYYYMMDD	Molding Date
Process Equip. Type	Epoxy Mold	-0-	Mold Composition
Process Condition_1	2.1	kg/min	Resin Injection Rate
Process Condition 2	275	MPa	Resin Injection Pressure
Process Condition_3	20.	Deg C	Mold Temperature
Process Condition 4	689.	MPa	Mold Pressure
Process Condition_5	90.	sec	Fill Time
Process Condition 6	1200.	sec	Cure Time
Process Condition_7	123.	Deg C	Postcure Temperature
Process Condition_8	3.	hrs	Postcure Time
*Part Form	Plate	-0-	Plaque
Part Dimension_1	610.	mm	Length
Part Dimension_2	610.	mm	Width
Part Dimension 3	3.0	mm	Thickness (Average)
Part Dimension_3_SD	0.1	mm	Thickness (Standard Deviation)
Part Reinf. Content	39.9	% wt	Fiber Content (Average)
Part Reinf. Content_SD	2.0	% wt	Fiber Content (Standard Deviation)

[a] The symbol -0- indicates a null value in this paper.
*Essential field for identification of composite material.

bases will load on these machines. A large database can take several hours to load on a workstation and several days on a PC. However, once loaded most database servers can retrieve data in a few seconds in response to a query if the schema is well designed.

The Automotive Composites Consortium (ACC) anticipated the need to organize large volumes of test data from suppliers using their test procedures manual [*4*], and they provide an IBM PC-compatible program [*8*], for collecting standardized format reduced test data and test metadata. We transferred that data to a SUN workstation for building the model database.

Software Systems

When should a project use existing off the shelf database software and when should it consider building a special purpose system for material database purposes? Rumble [*1*] notes the difficulty in building a new system and relates experiences from a NIST project of several years ago. The ex-

TABLE 9.2—Typical ACC structural property data set; test environment: temperature = −40°C.

Property	Value	Units	ACC Property Name
CTE11	27.9	micro m/m °C	Coefficient of Linear Thermal Expansion (Avg D696)
CTE11 SD	1.6	micro m/m °C	Coefficient of Linear Thermal Expansion (Std. Dev.)
US11T	152.8	MPa	Tensile Strength (Average D3039)
US11T SD	21.6	MPa	Tensile Strength (Standard Deviation)
E11T	10.26	GPa	Tensile Modulus (Average D3039)
E11T SD	0.97	GPa	Tensile Modulus (Standard Deviation)
NU12	0.33	-0-	Poisson's Ratio (Average D3039)
NU12 SD	0.03	-0-	Poisson's Ratio (Standard Deviation)
UE11T	2.54	%	Tensile Failure Strain (Average D3039)
UE11T SD	0.28	%	Tensile Failure Strain (Standard Deviation)
ER11T	1.95	kJ/m^3	Tensile Failure Energy (Average)
ER11T SD	0.38	kJ/m^3	Tensile Failure Energy (Standard Deviation)
US11C	275.6	MPa	Compression Strength (Average D3410)
US11C SD	22.2	MPa	Compression Strength (Standard Deviation)
E11C	10.16	GPa	Compression Modulus (Average D3410)
E11C SD	0.81	GPa	Compression Modulus (Standard Deviation)
UE11C	3.46	%	Compression Failure Strain (Average D3410)
UE11C SD	0.24	%	Compression Failure Strain (Standard Deviation)
ER11C	4.78	kJ/m^3	Compression Failure Energy (Average)
ER11C SD	0.62	kJ/m^3	Compression Failure Energy (Standard Deviation)
US12	108.1	MPa	Shear Strength (Average ACC Direct Shear)
US12 SD	9.25	MPa	Shear Strength (Standard Deviation)
G12	3.40	GPa	Shear Modulus (Average ACC Direct Shear)
G12 SD	0.85	GPa	Shear Modulus (Standard Deviation)
CTE22	25.3	micro m/m °C	Coefficient of Linear Thermal Expansion (Avg D696)
CTE22 SD	3.6	micro m/m °C	Coefficient of Linear Thermal Expansion (Std. Dev.)
US22T	162.7	MPa	Tensile Strength (Average D3039)
US22T SD	20.9	MPa	Tensile Strength (Standard Deviation)
E22T	10.33	GPa	Tensile Modulus (Average D3039)
E22T SD	0.78	GPa	Tensile Modulus (Standard Deviation)
NU21	0.34	-0-	Poisson's Ratio (Average D3039)
NU21 SD	0.04	-0-	Poisson's Ratio (Standard Deviation)
UE22T	2.64	%	Tensile Failure Strain (Average D3039)
UE22T SD	0.26	%	Tensile Failure Strain (Standard Deviation)
ER22T	2.15	kJ/m^3	Tensile Failure Energy (Average)
ER22T SD	0.41	kJ/m^3	Tensile Failure Energy (Standard Deviation)
US22C	299.4	MPa	Compression Strength (Average D3410)
US22C SD	23.9	MPa	Compression Strength (Standard Deviation)
E22C	11.37	GPa	Compression Modulus (Average D3410)
E22C SD	1.21	GPa	Compression Modulus (Standard Deviation)
UE22C	3.20	%	Compression Failure Strain (Average D3410)
UE22C SD	0.44	%	Compression Failure Strain (Standard Deviation)
ER22C	4.80	kJ/m^3	Compression Failure Energy (Average)
ER22C SD	0.92	kJ/m^3	Compression Failure Energy (Standard Deviation)

pense in terms of people, direct costs, and schedule costs is very high, which effectively limits this option to large institutions with very special needs. However, most projects will have their own schema and data dictionary with, hopefully, a thesaurus to provide synonyms for the casual user.

Two database languages are used to model materials information in the present application, EXPRESS and SQL. The structured query language (SQL) allows relational database query and other operations as described in the American National Standards Institute (ANSI) standard for database languages. The basic building block is the Table, a two-dimensional array of standard data types that include character strings and numbers. It is the older language and very widely used in commercial software products. The EXPRESS language is used to model information describing any product. It is being used to write PDES/STEP Material Product standards for different data environments. The basic building block is the entity, and the language can be used to define object-oriented databases for product information of any type. SQL data types are available in EXPRESS, and various government agencies are committed to the continued development of this language. NIST maintains a national PDES/STEP Testbed [*9*], which includes software that will translate a certain class of EXPRESS Schema into SQL Tables.

Our model database was built using a commercial product [*10*] that has an interface to the ACC test data file. This particular product allows object-oriented input data models to be loaded into a relational database with special attention given to units and metadata for materials.

ELEMENTS OF DATABASE DESIGN

We begin the model database development by first collecting relevant data from the source, here ACC, using ASTM

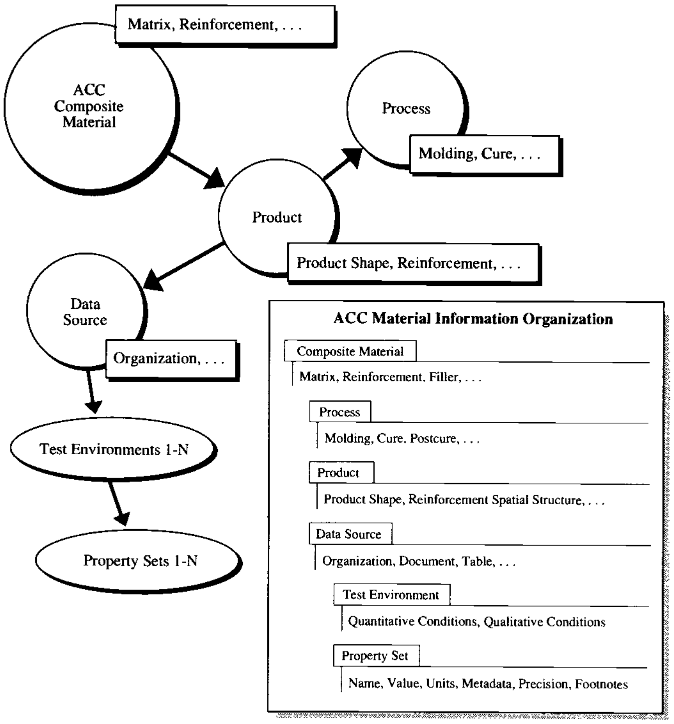


FIG. 9.2—ACC material information organization.

E-49 draft standards to organize like attributes. This model represents a commercially important class of materials, and it is instructive to see what constitutes relevant properties for automotive structural composites. This is the basis for understanding which data architectures will produce efficient databases for this application.

Creating an Object-Oriented Model

The ACC defines an ensemble of test specimens shown in Fig. 9.1. and a collection of material product attributes shown in Table 9.1 that define the raw material and process information important to manufacturers of automotive composite structures. The ASTM composite standards consulted in preparing Table 9.1 were ASTM Guide for the Identification of Composite Materials in Computerized Material Properties Databases (E 1309) and ASTM Guide for Identification of Fibers, Fillers, and Core Materials in Computerized Material Property Databases (E 1471) as well as ASTM Guide for the Identification of Polymers (excludes Thermoset Elastomers) in Computerized Material Property Databases (E 1308). It is interesting to note the large number of process parameters needed to identify a composite material for automotive design. In fact, the ACC source contains data on cure initiators and mold release agents not included here for lack of space.

To these data we now add a property data set for structural applications as shown in Table 9.2 for one test environment. These are the data produced by the specimens in Fig. 9.1 from a series of tests at −40.0°C and the ACC source has data for two other test temperatures. We now have all the relevant data for one material tested at one environment. Note that relevant can be and usually is redefined several times over the life of any product. Table 9.2 contains only test results with our model database missing important specimen and test procedure information not normally consulted by a designer. ASTM Guide for Development of Standard Data Records for Computerization of Mechanical Test Data for High Modulus Fiber-Reinforced Composite Materials (E 1434) for composite mechanical test data should be consulted for a lab database.

We now organize the data into like attributes as illustrated in Fig. 9.2. The entity names follow in general ASTM E-49 terminology that could be given synonyms for individual databases. What this figure suggests is an inheritance of material attributes by all the specimens cut from each panel, where for statistical coverage the ACC requires 27 panels per material. If we design our database using this architecture, we save storing all the material attributes at the property data set level. There are many ways to model materials information in EXPRESS. We next describe one for the ACC material and then illustrate how to create a relational database schema from the EXPRESS Schema (see Fig. 9.3).

It is at this point traditional ASTM activities and ideas are recast in the EXPRESS computer modeling language selected by the International Standards Organization (ISO) to define standards for the exchange of product model data. A specialist in information modeling may be required to produce an EXPRESS Schema. The one we present here is very basic and meant only to illustrate the process, not to define a standard. The Schema is a map that tells a database program how we want to model information about automotive composite material products. It is up to the database program to actually use the map to load data into our model database, and there are many ways this can be done.

Note the use of a WHERE rule to check input values. This is an important aspect of a Schema that we only highlight here. Very sophisticated integrity constraints can be defined by the database builder to ensure all data loaded using a Schema meet these constraints. There are compilers for the EXPRESS language that check a model for syntax, cross-references, and redundancies. There are also utility programs for generating a table of contents, schema index, and EXPRESS_G diagrams of a model. These are important tools in the design of a database so it can be updated and maintained without anomalies creeping in through convoluted functional dependencies. This process is called normalizing the database schema, and Colton [*11*] describes the process in detail for aircraft composites. There are very few database programs today that can use an EXPRESS Schema to load data directly. Stanton and Rahmann [*12*] have loaded composite data from a schema where all the entities but one were explicit attributes, and these data then were exported as an EXPRESS physical file. Please note that EXPRESS modeling and STEP data exchange are emerging technologies and not widely available in 1992 but will be in the future. The computer industry is moving to support PDES/STEP rapidly. However, today most database programs first convert or approximate EXPRESS models as an SQL relational database schema. Some [*3*] would argue that today only relational database systems offer a rational basis for material databases, with object oriented systems likely to be the choice for future systems.

```
SCHEMA ACC_Material_Model ;

    ENTITY ACC_Composite_Material
      SUPERTYPE OF (ONEOF ( ACC_Matrix, ACC_Reinforcement,
                            ACC_Process, ACC_Material_Product)) ;
           Material_Reference    : STRING ;
           Material_Class        : OPTIONAL material_class_list ;
           Matrix_Class          : matrix_class_list ;
           Reinforcement_Class   : reinforcement_class_list ;
           Structural_Detail     : OPTIONAL STRING ;
           Ply_Type              : precursor_type_list ;
           Ply_Name              : STRING ;
           Ply_Manufacturer      : STRING ;

    END_ENTITY ;

    ENTITY ACC_Matrix
      SUBTYPE OF (ACC_Composite_Material) ;
           Matrix_Subclass       : OPTIONAL STRING ;
           M_Chemical_Name       : STRING ;
           M_Commercial_Name     : STRING ;
           M_Specific_Gravity    : value_unit_metadata ;
           M_Glass_Trans_Temp    : value_unit_metadata ;
           M_Viscosity           : value_unit_metadata ;
           M_Neat_Resin_UTS      : value_unit_metadata ;
           M_Neat_Resin_E        : value_unit_metadata ;
           M_Neat_Resin_UTE      : value_unit_metadata ;

    END_ENTITY ;

    ENTITY ACC_Reinforcement
      SUBTYPE OF (ACC_Composite_Material) ;
           Reinforcement_Subclass  : STRING ;
           R_Chemical_Name         : OPTIONAL STRING ;
           R_Form                  : OPTIONAL STRING ;
           R_Product_Form          : value_unit_metadata ;
           R_Specific_Gravity      : value_unit_metadata ;
           R_Manufacturer          : STRING ;
           R_Length                : value_unit_metadata ;
           R_Yield                 : value_unit_metadata ;
           R_Bundle_Size           : value_unit_metadata ;
           R_Filament_Diameter     : value_unit_metadata ;
           R_Chemical_Sizing       : STRING ;
           R_Binder                : STRING ;

    END_ENTITY ;

    ENTITY ACC_Process
      SUBTYPE OF (ACC_Composite_Material) ;
           Process_Designation     : STRING ;
           Process_Stage_Type      : process_stage_list ;
           Processor               : STRING ;
           Process_Date            : STRING ;
           Process_Condition_1     : value_unit_metadata ;
           Process_Condition_N     : value_unit_metadata ;

    END_ENTITY ;

    ENTITY ACC_Material_Product
      SUBTYPE OF (ACC_Composite_Material) ;
           Part_Form                  : STRING ;
           Part_Dimension_1           : value_unit_metadata ;
           Part_Dimension_2           : value_unit_metadata ;
           Part_Dimension_3           : value_unit_metadata ;
           Part_Dimension_3_SD        : value_unit_metadata ;
           Part_Reinforce_Content     : value_unit_metadata ;
           Part_Reinforce_Content_SD  : value_unit_metadata ;

    END_ENTITY ;

    ENTITY Material_Test_Environment ;

           Material_Tested            : ACC_Composite_Material ;
           Test_Temperature           : value_unit_metadata ;
           Test_Humidity              : value_unit_metadata ;
           Test_Results               : material_property_set ;

    END_ENTITY ;

    ENTITY material_property_set ;

           CTE11         : value_unit_metadata ;
           CTE11_SD      : value_unit_metadata ;
           US11T         : value_unit_metadata ;
           US11T_SD      : value_unit_metadata ;
           E11T          : value_unit_metadata ;
           E11T_SD       : value_unit_metadata ;
           NU12          : value_unit_metadata ;
           NU12_SD       : value_unit_metadata ;
           UE11T         : value_unit_metadata ;
           UE11T_SD      : value_unit_metadata ;

           (* The property set can be completed
           using Table 2 *)

    END_ENTITY ;
```

FIG. 9.3—Example of how to model materials information using EXPRESS.

```
    ENTITY value_unit_metadata ;

           property_value    : REAL ;
           unit              : STRING ;
           metadata          : STRING ;
           precision         : REAL ;

           (* Here we illustrate a WHERE rule
           used to check input values *)

    WHERE
           R1 : precision > 0.0 ; -- Range Check

    END_ENTITY ;
    TYPE  metal               = STRING ; END_TYPE ;
    TYPE  polymer             = STRING ; END_TYPE ;
    TYPE  ceramic             = STRING ; END_TYPE ;
    TYPE  carbon              = STRING ; END_TYPE ;
    TYPE  composite           = STRING ; END_TYPE ;
    TYPE  user_defined        = STRING ; END_TYPE ;
    TYPE  fiber               = STRING ; END_TYPE ;
    TYPE  filler              = STRING ; END_TYPE ;
    TYPE  core                = STRING ; END_TYPE ;
    TYPE  prepreg             = STRING ; END_TYPE ;
    TYPE  prelam              = STRING ; END_TYPE ;
    TYPE  tow                 = STRING ; END_TYPE ;
    TYPE  BMC                 = STRING ; END_TYPE ;
    TYPE  XMC                 = STRING ; END_TYPE ;
    TYPE  SMC                 = STRING ; END_TYPE ;
    TYPE  preforming          = STRING ; END_TYPE ;
    TYPE  laminating          = STRING ; END_TYPE ;
    TYPE  cure                = STRING ; END_TYPE ;
    TYPE  post_cure           = STRING ; END_TYPE ;

    TYPE  resin_transfer_molding = STRING ; END_TYPE ;
    TYPE  injection_molding      = STRING ; END_TYPE ;

    TYPE
           material_class_list = SELECT
                                 ( metal,      polymer,
                                   ceramic,    composite,
                                   user_defined_name ) ;
    END_TYPE ;

    TYPE
           matrix_class_list = SELECT
                               ( metal_matrix, polymer_matrix,
                                 carbon_matrix, ceramic_matrix ) ;
    END_TYPE ;

    TYPE
           reinforcement_class_list = SELECT
                                      ( fiber,    filler,
                                        core ) ;
    END_TYPE ;

    TYPE
           precursor_type_list = SELECT
                                 ( prepreg,    prelam,
                                   tow,        BMC,
                                   SMC,        XMC,
                                   user_defined_precursor ) ;
    END_TYPE ;

    TYPE
           process_stage_list = SELECT
                                ( preforming,    laminating,
                                  cure,          post_cure,
                                  resin_transfer_molding,
                                  injection_molding,
                                  user_defined_process ) ;
    END_TYPE ;
END_SCHEMA ;
```

FIG. 9.3—(Continued).

Creating a Relational Database Schema

We could model the ACC material information directly using SQL or we could translate our EXPRESS Entities into SQL Tables as described in the monograph by Morris [*13*]. If we limit the discussion to explicit attributes, it is possible to map an EXPRESS entity into an SQL Table. The attributes become columns in a Table, and the data types for explicit attributes map directly. We illustrate this for the Entity ACC_Composite_Material as follows:

```
CREATE TABLE ACC_Composite_Material (Material_Ref
                              CHAR (40) NOT NULL
                              PRIMARY KEY,
Material_Class                CHAR(40),
Reinforcement_Class           CHAR(40) NOT NULL,
Matrix_Class                  CHAR(40) NOT NULL,
Structural_Detail             CHAR(40),
Ply_Type                      CHAR(40) REFERENCES
                              P_Type_List,
Ply_Name                      CHAR(40) NOT NULL,
Ply_Manf                      CHAR(40) NOT NULL, );
```

Note that attribute names have been abbreviated for economy of screen space following the practice in fedex_sql [*12*], and the use of KEYs to help the SQL schema define constraints and functional dependencies. Even very simple material models in this language require the database designer to manage low level details. This has led most commercial database systems to offer "enhanced query languages" to simplify what can be a tedious and labor intensive process.

The job of translating EXPRESS schemas into SQL tables can get very complicated, and in general may require changes that are not one-to-one. When the attributes are entity data types it is not possible to directly map to an SQL table without using pseudo columns or adding extra columns. We illustrate this for the entity material_property_set:

```
CREATE TABLE MATERIAL_PROPERTY_SET (
  CTE11                FLOAT(10), -- value
  CTE11_units          CHAR(40), -- units
  CTE11_metadata       CHAR(40), -- metadata
  CTE11_precision      FLOAT(10) CHECK
                         (CET11_precision
                         > 0.0), );
```

where every property attribute, here CTE11, has its units, metadata, and precision entered explicitly as a column in the table definition.

There are "enhanced" relational database systems that support loading units and metadata only once per property attribute [*10*], and we illustrate that feature in the following schema for a single column specification,

```
METADATA
    CTE11 = ASTM D696
  DESIGN SCHEMA
    CTE11   REAL   1   1   "micro m/m deg C",
    "Coefficient of Linear Thermal Expansion in 0-deg,
    Direction, Avg", , 0.1
```

Here the units for coefficient of linear thermal expansion and the metadata describing the test, ASTM Test Method for Coefficient of Linear Thermal Expansion of Plastics (D 696), are part of the schema. They will be inherited by every table loaded using material_property_set. What this means to the material database builder is that, in general, SQL schemas will require more resources than "enhanced" SQL schemas with features designed to support technical data with extensive units and metadata.

There are other features, like precision in the sense of the ASTM Practice for Use of the International System of Units (SI) (The Modernized Metric System) (E 380), that need to be supported by the database server. The SQL data type FLOAT(10), for example, indicates a binary precision of ten (10), and this must be preserved in units conversion. In many systems, precision is an implicit attribute not easily available for operations on data. Here, of course, we have made precision an explicit attribute of the CTE11 value.

The schema illustrated in Fig. 9.2, which shows the first ENTITY, ACC_Composite_Material, loaded at the top level and the last ENTITY, material_property_set, loaded at the lowest level, follows a natural hierarchy for this application. It allows all the ACC_Process attributes, for example, to be inherited by the property sets without having to load that data as columns in every property table. This can be a very substantial savings.

One final note about the model database schema is that the tables in this model contain only numeric property data. In general, a material database will also require graphical data for nonlinear properties. Temperature dependent property data and fatigue life data are typical nonlinear material property sets where an analytic entity (X(Y),Y: Y1<Y<Y2) or parametric entity (X(t), Y(t): T1<t<T2) is required. Most relational database servers treat graphics as a post-processing feature, but it is possible to store graphical entities in a relational database system if the data type is supported (Fig. 9.4).

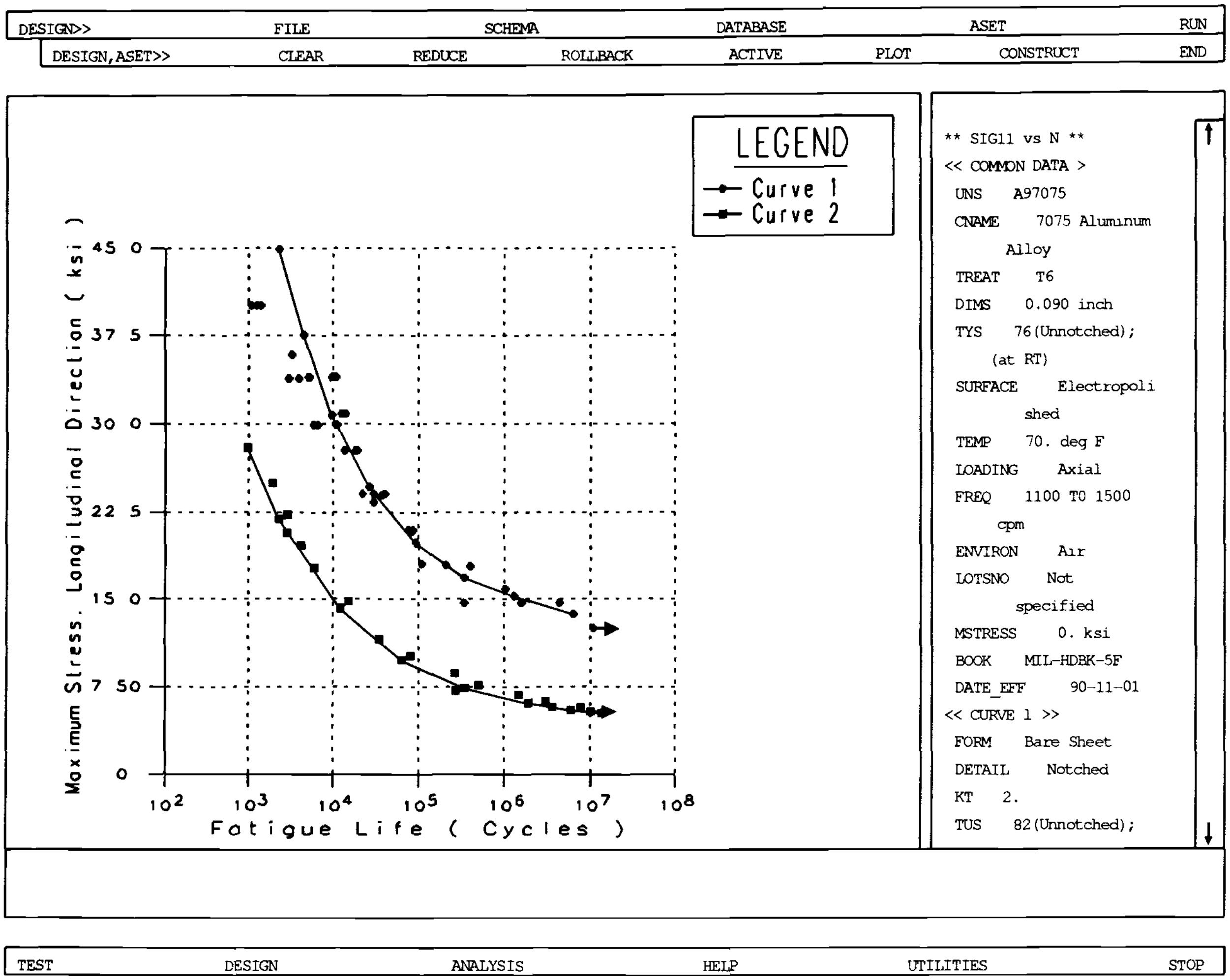
CONSTRUCTING A DATABASE

We now have data and a database design (schema) for loading material properties for automotive composite design applications. Usually data from a test series are available in spreadsheet form from the source. The ACC provides a PC program, the Supplemental Data Reporting Package, for an automotive design application per ACC test procedures [*4*]. These data were made available to ASTM E-49 for a round-robin test. The goal is a paperless transfer of information from the raw data stage through data reduction, database construction, and ultimately CAD/CAM/CAE applications as illustrated in Fig. 9.5. In general there will be a number of databases serving different data environments, and we describe the steps in constructing one such database.

The intended use of our database is the design and analysis of structural composite parts. Members of the ACC use the finite-element analysis method for this application [*8*] and the data we are about to load contain material properties for FEA codes to use in structural analyses. There is also enough process information to uniquely identify the composite for other purposes such as molding simulation.

Building Schema Based Physical Files

The physical file containing data organized using the Schema is called the instantiated model. The EXPRESS modeling language has an exchange file syntax [*14*] that may become a standard for this function, but today most commercial database programs have their own physical file format. We are using one designed for materials data and refer the reader to the product literature for a more comprehen-



```
** SIG11 vs N **
<< COMMON DATA >
 UNS    A97075
 CNAME     7075 Aluminum
        Alloy
 TREAT    T6
 DIMS    0.090 inch
 TYS    76(Unnotched);
     (at RT)
 SURFACE     Electropoli
           shed
 TEMP    70. deg F
 LOADING     Axial
 FREQ    1100 TO 1500
       cpm
 ENVIRON     Air
 LOTSNO     Not
          specified
 MSTRESS     0. ksi
 BOOK    MIL-HDBK-5F
 DATE_EFF      90-11-01
<< CURVE 1 >>
 FORM    Bare Sheet
 DETAIL     Notched
 KT   2.
 TUS    82(Unnotched);
```

FIG. 9.4—A graphical material entity from MIL-HDBK-5 database.

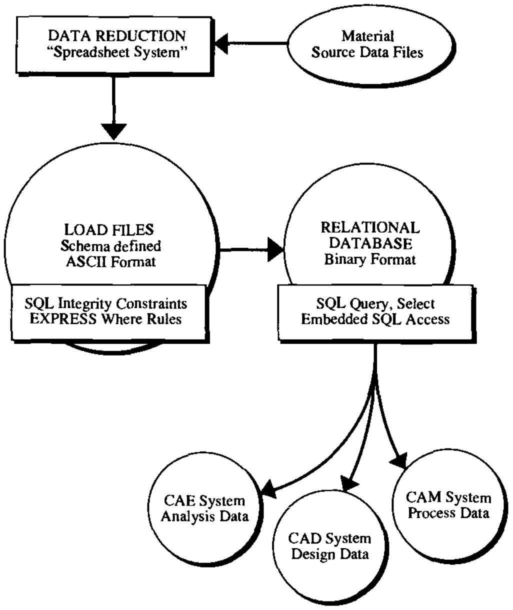


FIG. 9.5—Database development steps and application interfaces.

sive description of that system's syntax [10]. Figure 9.6 is the ACC model database physical file used to input, that is, load data based on the EXPRESS Schema. It was necessary to map the Schema into the syntax of that program, which is similar to the SQL table schema.

Note that metadata and units are not re-entered for each Test_Environment, only the property data are input. If data for another form of DOW 411-C50/CT U750 were available, for example, a thicker part, then another ACC_Material_Product and material_property_set under the same ACC_Composite_Material designation would appear next in the load file. After all data for this particular material are loaded, then a second material data set would be loaded and so on until all the source data were in the load file.

Validating Data Entry

A data model can introduce integrity constraints for use in validating data input by specifying <unique>, <referential>, and <check> SQL Table constraints applied after the execution of each SQL statement. Domain constraints are available in EXPRESS for performing the same function on the load or input file as we illustrated earlier using a WHERE

ACC Physical File for M/VISION input:

```
METADATA
  E11T   = ASTM D3039
  E22T   = ASTM D3039
  NU12   = ASTM D3039
  NU21   = ASTM D3039
  US11T  = ASTM D3039
  US22T  = ASTM D3039
  UE11T  = ASTM D3039
  UE22T  = ASTM D3039
  ER11T  = Fail_Energy_3039
  ER22T  = Fail_Energy_3039
  E11C   = ASTM D3410
  E22C   = ASTM D3410
  US11C  = ASTM D3410
  US22C  = ASTM D3410
  UE11C  = ASTM D3410
  UE22C  = ASTM D3410
  ER11C  = Fail_Energy_3410
  ER22C  = Fail_Energy_3410
  G12    = ACC_DS1
  US12SA = ACC_DS1
  G21    = ACC_DS1
  US21SA = ACC_DS1
  CTE11  = ASTM D696
  CTE22  = ASTM D696
  M_TG   = ASTM D4065
  M_UST  = ASTM D638
  M_ET   = ASTM D638
  M_UET  = ASTM D638
ACC_Material
   CMP_NAME  = Fiberglass/Vinyl Ester
   ACC_DESIG = Dow 411-C50/CT U750
   PMC       = M22GL100UP0320
   M_CLASS   = Polymer
   R_CLASS   = Fiber
   PLY_TYPE  = Mat
   PLY_NAME  = CT U750
   PLY_MANF  = Certainteed
   R_TYPE    = E Glass
   R_FORM    = 1.5
   R_WPCT    = 40.00
   R_SG      = 2.60
   R_MANF    = Certainteed
   R_PROD    = U750
   R_LENG    = continuous
   R_YLD     = 6633
   R_SZS     = 2500
   R_FDIA    = 16.00
   R_CHSZ    = Siline (KE6N850501)
   R_BIND    = thermal plastic polyester
   M_TYPE    = 411-C-50
   M_WPCT    = 60.00
   M_SG      = 1.12
   M_MANF    = Dow Chemical U.S.A.
   M_CODE    = 411-C-50
   M_TG      = 121.00
   M_VISC    = 100
   M_COMP    = Vinyl Ester
   M_UST     = 79.20

   M_ET      = 3.30
   M_UET     = 1.00
ACC_Process
   INI_TYPE = MEKP
   INI_CON  = 2.0
   PRM_TYPE = CoNap
   PRM_CON  = 0.3
   ACL_TYPE = DMA
   ACL_CON  = 0.2
   OC1_TYPE = Chemtrend 2005
   OC1_FNC  = Mold release
   OC1_MANF = Excel
   OC1_CODE = 2005
   OC1_INFO = Mold release for RT RTM
   PRC_DATE = 1990/02/10
   PRC_TYPE = RTM
   MOL_COMP = epoxy
   INJ_RATE = 2.1
   INJ_PRES = 275
   MOL_TEMP = 20
   MOL_PRES = 689
   FILLTIME = 90
   CURETIME = 1200
   POS_TEMP = 123
   POS_TIME = 3
ACC_Product
   PRD_R_WPCT    = 39.9
   PRD_R_WPCT_SD = 2.0
   PRD_M_WPCT    = 60.1
   PRD_M_WPCT_SD = 2.0
   PRD_SG        = 1.5
   PRD_SG_SD     = 0.0
   PRD_TG        = 112.0
   PRD_TG_SD     = 2.0
   PRD_TH        = 3.0
   PRD_TH_SD     = 0.1
Source
     TABLE_NAME = Sample Test Data Summary
     TEST_ENGR  = Peterson, Johnson, Hagerman
     TEST_ORGN  = Automotive Composites Consortium
     DATA_USE   = Structural Design (ACC Model Database
                  Testing)
Test_Environment
   TEMP  = -40.
   HUMID = -0-
```

FIG. 9.6—ACC models database physical file used to input.

```
property_set
   CTE11    = 27.9
   CTE11_SD = 1.6
   US11T    = 152.8
   US11T_SD = 21.6
   E11T     = 10.26
   E11T_SD  = 0.97
   NU12     = 0.33
   NU12_SD  = 0.03

   (* The remaining forty properties for -40 deg C go here *)

END
   (* Test_Environment 23.0 deg C property data *)
END
   (* Test_Environment 121.0 deg C property data *)
END
```

FIG. 9.6—(Continued).

rule for an Entity. These can be very helpful in screening clerical and basic functional dependency errors. Each database has its own property set and data environment that need to be considered in developing these constraints. At the moment there is no ASTM Recommended Practices Guide at this level. A structural material, for example, should be checked for an elastic modulus of less than a 1000 GPa and a weight density of less than 100 Mg/m^3.

We can build on this idea and imagine views of our data like the ones used to aid in material selection by Ashby et al. If we plot the elastic modulus versus density for all the materials in our database (Fig. 9.7), they should lie in very well known envelopes by material type, and we can present a graphic of that integrity constraint. This particular view is only one of many; another common view is elastic modulus versus coefficient of linear thermal expansion, which has a "1/X" shape. The ISO 9000 standard for software [*15*] requires testing and validation procedures for acceptance that can be met at least in part using integrity constraints like these on the input data. We end this heuristic discussion of data quality by referring the reader to Chapter 6 for a comprehensive discussion of data evaluation, validation, and quality.

CAD/CAM/CAE Access to Material Databases

The investment in developing quality materials data in electronic formats benefits the design process most when it

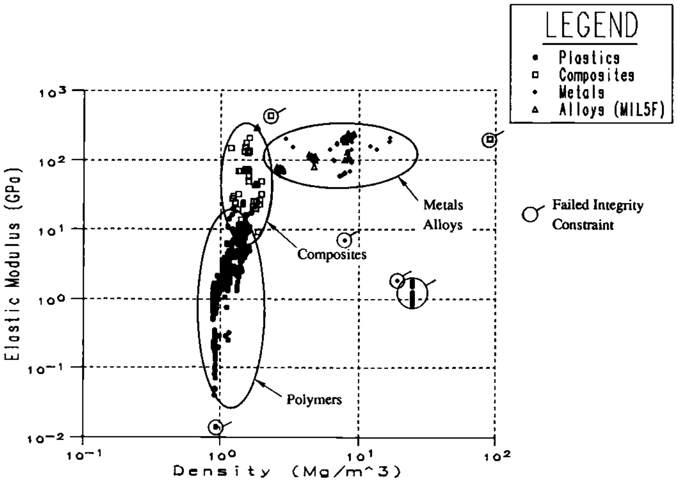
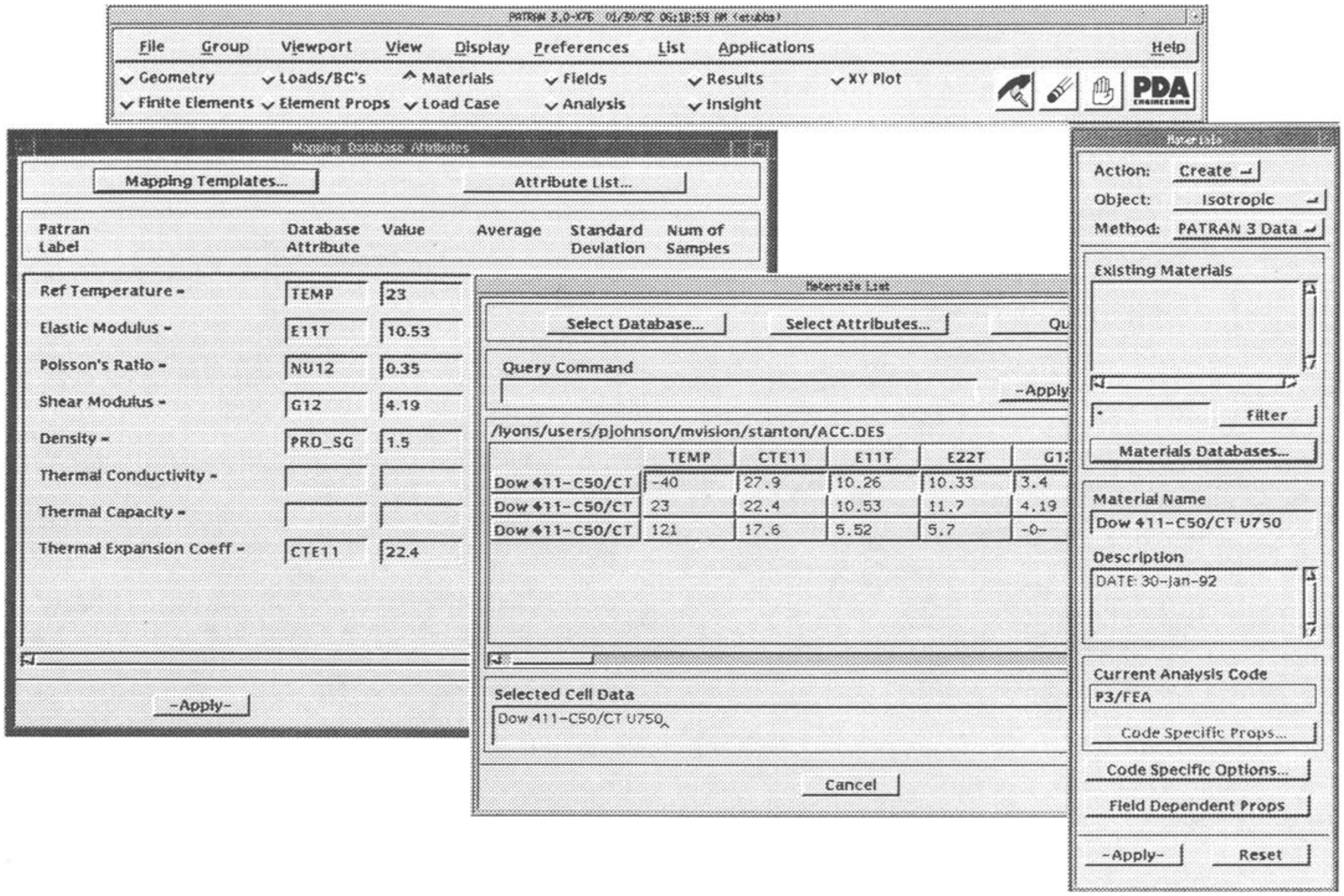


FIG. 9.7—Data entry validation test graphic.



Patran Label	Database Attribute	Value	Average	Standard Deviation	Num of Samples
Ref Temperature =	TEMP	23			
Elastic Modulus =	E11T	10.53			
Poisson's Ratio =	NU12	0.35			
Shear Modulus =	G12	4.19			
Density =	PRD_SG	1.5			
Thermal Conductivity =					
Thermal Capacity =					
Thermal Expansion Coeff =	CTE11	22.4			

/lyons/users/pjohnson/mvision/stanton/ACC.DES

	TEMP	CTE11	E11T	E22T	G1
Dow 411-C50/CT	-40	27.9	10.26	10.33	3.4
Dow 411-C50/CT	23	22.4	10.53	11.7	4.19
Dow 411-C50/CT	121	17.6	5.52	5.7	-0-

Selected Cell Data: Dow 411-C50/CT U750

Action: Create

Object: Isotropic

Method: PATRAN 3 Data

Material Name: Dow 411-C50/CT U750

Description: DATE: 30-Jan-92

Current Analysis Code: P3/FEA

FIG. 9.8—CAE material database application.

is used in CAD/CAM/CAE applications. This integration function needs the PDES/STEP standards for product data exchange that are just now emerging for materials information. The IGES standard has very limited FEA materials data namely basic linear coefficient data like initial elastic modulus and coefficients of linear thermal expansion. The system used to load our ACC model database has an export feature to PATRAN and to IGES neutral files. Access to various application codes for process simulation or structural performance analyses is then possible from these neutral files.

The problems with property data exchange include missing properties required for a specific analysis and differences between the available test property parameters and the related analysis property parameters. There are also serious problems with material designation that Sargent [3] describes in some detail that we do not repeat here. An example of missing data might be a transverse modulus in a unidirectional ply material, and the CAE application requires the missing property to function. An example of property parameter differences actually occurs in our model databases; the elastic modulus in tension and compression are unequal while linear finite-element models require a single value. In both instances engineering judgment is required to complete the data exchange for the CAE application, and the judgment needs to be an informed decision. One approach uses graphical user interfaces to show the engineer the database attributes in native mode (Fig. 9.8) before deciding how to complete the property set for a CAE application. More sophisticated rule-based systems can be imagined, which seems an appropriate point to end the discussion on how to build a material database.

SUMMARY

In this brief model database development we have worked through the steps from test data source through the modeling process and illustrated concepts and procedures important to reaching the end user in a CAD/CAM/CAE application. Many intermediate steps have been left out, such as test data reduction, familiar topics for an ASTM reader, to concentrate on the less familiar computerization issues such as material data, metadata, and modeling languages.

The use of the database for structural analyses or any other application has not been covered. Also keep in mind that the model database is just that, a model to illustrate the database building process. Your application will likely have different property sets and less material process data if the material type is a metal alloy. Commodity or product information databases will be very different, having limited property sets but thousands of material products.

Experience in building and using material databases has led to an appreciation of the importance of a good design (schema) for efficient use and the importance of standards for efficient exchange of information among CAD/CAM/CAE applications. The latter are just now entering the round-robin stage and need the support of the entire engineering community, not just materials and process engineers.

REFERENCES

[*1*] Rumble, J. R. and Smith, F. J., *Database Systems in Science and Engineering*, Adam Hilger Publisher, ISBN 0-7503-0048-5, 1990.

[*2*] Westbrook, J. H. and Grattidge, W., "The Role of Metadata in the Design and Operation of a Materials Database," *Computerization and Networking of Material Databases, STP 1106*, Kaufman and Glazman, Eds., American Society for Testing and Materials, Philadelphia, 1991, pp. 84–102.

[*3*] Sargent, P., *Materials Information for CAD/CAM*, Butterworth-Heinemann Ltd., 1991, Chapter 5.

[*4*] *Test Procedures for Automotive Structural Composite Materials*, Automotive Composites Consortium, Troy, MI, 1990.

[*5*] Westbrook, J. H. and Rumble, J. R. *Computerized Materials Data Systems, Proceedings of a Workshop Devoted to Discussions of Problems Confronting Their Development*, Fairfield Glade, TN, 1982, p. 23.

[*6*] ISO CD 10303-11, Product Data Representation and Exchange—Part 11: EXPRESS Language.

[*7*] Kaufman, J. G., "Increasing Data Systems Responsiveness to End-User Expectations," *Computerization and Networking of Material Databases, STP 1106*, Kaufman and Glazman, Eds., American Society for Testing and Materials, Philadelphia, 1991, pp. 103–111.

[*8*] Dearlove, T. J., et al., "Standardization of Test Methods for an Automotive Database," *Proceedings of the Sixth Annual ASM/ESD Advanced Composites Conference*, 1990, pp. 101–112.

[*9*] Clark, S. N., *An Introduction to the NIST PDES Toolkit*, NISTIR 4336, U.S. Department of Commerce, 1990.

[*10*] *M/VISION User Manual*, Publication No. 2190011, PDA Engineering, 1990.

[*11*] Colton, J. S., "The Design and Implementation of a Relational Materials Property Data Base," *Engineering with Computers*, Vol. 4, 1988, pp. 87–97.

[*12*] Stanton, E. L. and Rahmann, S. E., "Constructing Material Databases for CAE Systems," presented at the Sixth U.S. Japan Conference on Composite Materials, Lake Buena Vista, FL, 1992.

[*13*] Morris, K. C., *Translating Express to SQL: A User's Guide*, NISTIR 4341, U.S. Department of Commerce, 1990.

[*14*] ISO CD 10303-21, Product Data Representation and Exchange—Part 21: Clear Test Encoding of the Exchange Structure.

[*15*] ISO 9000 International Standards for Quality Management, Part 3: Guidelines for the Application of ISO 9001 to the Development, Supply and Maintenance of Software, ISBN 92-67-10165-X, 1991, pp. 21–47.

Subject Index

Copyright© 1993 by ASTM International

About the Editor

DR. CRYSTAL H. NEWTON

CURRENTLY a project engineer for Materials Sciences Corporation in Fort Washington, Pennsylvania, Dr. Crystal Newton has served there for seven years. Her areas of interest include data management and standardization, including development of composite material databases, the interaction between expert systems and material property databases, and analysis of hybrid composite materials.

From 1984 to 1986, Dr. Newton was a research engineer in the Materials Research Center at Lehigh University, performing research on near-threshold fatigue of metals, fatigue and fracture of polymers and composites, and the viscoelastic properties of elastomers.

The editor holds a B.S. degree in mechanical engineering from Carnegie-Mellon University, and an M.S. in applied mechanics and a Ph.D. in mechanical engineering from Lehigh University.

Dr. Newton is an active member of ASTM's Committees E-49 on Computerization of Mechanical and Chemical Property Data and D-30 on High-Modulus Fibers and Their Composites and participates on several other committees. She is also a member of Sigma Xi and Pi Tau Sigma.

ISBN 0-8031-2052-4