AN INFORMATION FILING AND RETRIEVAL SYSTEM FOR THE ENGINEERING AND MANAGEMENT RECORDS OF A LARGE-SCALE COMPUTER DEVELOPMENT PROJECT

*G. A. BARNARD, III, and LOUIS FEIN**

Introduction

This paper describes the organization of a file and retrieval system developed for use on a large-scale engineering project—the development of the ERMA Mark I. ERMA (Electronic Recording Machine Accounting), which was built by Stanford Research Institute for the Bank of America, is a large-scale computer and data-processing system designed to process bank checks automatically.

Although the authors have had no experience in the field of documentation as such, a useful and potentially efficient paper system was produced. The work on the file system was motivated by a respect and appreciation for the importance of a good retrieval system—not only in terms of the cost and efficiency of a research project, but also for its success as well.

The principles developed and utilized in creating this system should be useful to technical and research management and to others concerned with projects that can be structuralized. Some research projects, in which only the broad outlines of the problem are initially understood, will not meet the specification of prior structuralization. However, most engineering and research projects do meet this specification.

General Background

Crash programs are a significant characteristic of our industrial, commerical, and governmental operations. When such programs are hurriedly organized and executed, there is always the danger of inefficiencies arising from an inadequate information and record system.

The following is a partial list of situations and requirements that such an adequate system would serve to improve:

1. Training and orienting new personnel: Too often newly assigned people are thrown into the middle of a project without training or orientation—or they are given "some kind of literature" to read for a week or so and then are expected to be familiar with the project. A simple information system designed to provide for the trainee's needs for material and information already available will increase the efficiency of this orientation program.
2. Indispensable personnel: Frequently, vital information is carried in people's heads and, because it is not recorded, is often inconvenient and difficult to retrieve. The loss of such people due to illness or resignation may be catastrophic to the success of a project. An adequate information system produces a climate that discourages information hoarding.
3. Duplication of effort: It is wasteful, to have people working on one project duplicating the work of others on related projects, though this is sometimes unavoidable because of security or other reasons, but there is no legitimate excuse for duplication of effort within a project because of an inefficient information control system.
4. Action based on less information than is actually available: Often basic information has been generated within a project, but for some reason (usually because it has not been properly recorded and filed) it is not available to those requiring it. Under such

*G. A. Barnard, III, is an electronics engineer with the Stanford Research Institute, Menlo Park, California. Louis Fein, of Palo Alto, California, is a physicist who served as a consultant to the Institute on the ERMA project.

circumstances, a duplicate program may be initiated to obtain the desired information lest decisions be made on the basis of insufficient information.

5. Patent preparation: Innumerable man-hours are expended by patent attorneys in interviewing an inventor to obtain relevant information for purposes of disclosure and patent. Proper documentation and filing of patentable ideas would reduce this lost time.
6. Manufacturing information: Many "one-shot" projects include designers whose design paper is the "back of an envelope." One need scarcely emphasize that this is inadequate as a basis for translation to production specifications. Obviously, even this insufficient information is never filed.
7. Dissemination of information: Various types of information distribution systems exist (functioning with varying degrees of success). An adequate record system, explicitly designed to coordinate with a good distribution system, will keep appropriate personnel promptly informed of current project developments.

Performance Specifications

In attempting to avoid these potentially dangerous and inefficient situations, and, of course, in trying to keep an adequate record of the ERMA project, a set of performance requirements for a filing and retrieval system was developed.* The performance requirements given below constituted a set of objectives which the actual system was designed to attain.

1. Any single question or statement about ERMA made by any interested observer should be keyed to a unique code number. The code number in turn should correspond to the position of the designated information in the file system.
2. The system should be closed. All major categories and category parameters characterizing the system should be identified and, at the outset should be exhaustive. Increased knowledge about the project should be incorporable within the originally designed file system by adding subcategories and subparameters.
3. Authors or their representatives should be responsible for assigning file code numbers to all papers generated or received by them. Thus, a key should be available and it should be easy to use.
4. Every conceivable kind of paper generated within the project or received from the outside, but relevant to the project, should have a unique place in the system—the system should thus be integrated and coordinated. The file should be a technical, as well as an administrative, file.
5. The whole system should be easy to use.

The Strategy

Before deciding on the details of the filing system, a general strategy was devised to be used as a guide for the design of the actual system. This was not formalized, nor were alternative strategies identified and evaluated so that an optimum choice could be made. Nevertheless, the following is the adopted strategy:

1. A multi-coordinate system with corresponding codes should be used.
2. An explanatory index (key) with simple explanations was assumed to be the best way to insure selection of appropriate file numbers for new material to be incorporated in the file.
3. To facilitate use, the physical construction of the record storage system should be modeled upon the structure of the multi-coordinate file system.
4. The entries on one coordinate of the file system should be the elements of the various levels (categories) of the hierarchical structure chosen to represent the ERMA Mark I system.
5. The entries on the second coordinate of the file system should be attributes characterizing the elements (categories) on each level of the hierarchy. These attributes should be *generic* to every category.
6. The entries on the third coordinate of the file system should be a listing by type of those activities relevant to the ERMA Mark I project.

*During the development of the file system, these performance requirements were not actually recorded in detail as they are in this paper. (Nor was the strategy actually detailed as it is here presented). With the benefit of hindsight, the authors recommend to readers intending to develop a similar file system, that performance requirements, strategy, and tactics be formalized and recorded in full detail during the course of the development.

The Special Aspects of the ERMA Mark I Project as they Affected the File and Retrieval System

ERMA Mark I was intended to be an engineering model, later to be product-designed and built in quantity. The engineering model was to be shipped to, installed, operated, and maintained at a pilot installation. The data-processor of the system was a wired-program digital computer with storage drums, magnetic tapes, keyboard entry machines, and a statement printer.

The ERMA project generated recorded information concerning, not only design, engineering, and production, but also shipping, installation, maintenance, and related activities. Of course, potential manufacturers, patent people, public relations people, and accountants, as well as designers, shop people, etc., made demands on the system for information of special interest to each.

The following is a partial list of the kinds of records generated in a project such as the Mark I ERMA project:

1. Correspondence
2. Logical design diagrams,. equations, and corresponding prose descriptions
3. Circuit diagrams, schematics, and/or corresponding prose descriptions
4. Specifications
5. Mechanical drawings
6. Wiring lists - diagrams
7. Instructions
8. Standards
9. Statistics
10. Photographs
11. Schedules
12. Logs

The following is a partial list of the types of individuals who were interested in using the recorded data:

1. Administrators - supervisors
2. Logical designers
3. Circuit designers
4. Wiremen
5. Technicians
6. Sponsors
7. Public Relations people
8. Patent attorneys
9. Potential manufacturers
10. Trainees
11. Accountants
12. Maintenance people
13. Spare parts people
14. Technical writers

The Adopted File System

The file system originally considered was three-dimensional. The entries on two dimensions were the same as those in the system actually adopted (*cf.* Figure 1). The third dimension comprised entries consisting of the user types enumerated above. To implement the three dimensional system, each type of user would have required separate records corresponding to the two dimensional entries in Figure 1. This procedure was considered to be too costly and, reluctantly the plan was abandoned.

Coordinate 1

A technical file system, such as this, is most efficient when it reflects the logical organizational structure of the equipment or system. The structure in turn should reflect the particular character of the equipment or system.

ERMA Mark I is a wired-program data processor with appropriate terminal equipment. The system and its functional organization may be described with the aid of the hierarchical structure shown in Figure 2. (The following paragraph should be read in conjunction with Figure 2.)

ERMA Mark I performs certain *tasks*. These *tasks* are performed by execution of some combination of a number of standard *routines*. A standard *routine* is executed with the aid of one of four *programmers*, several of the standard *service units*, *auxiliary equipment*, and *switch units*. These four kinds of equipments consist of some combination of standard *packages*. Each *package* consists of some combination of standard *components*. The *packages* and *components* reside in *racks* with the *wiring* constituting the connecting links.

Each italicised heading is designated as system *categories* with a two-digit code number ending in 0. Thus, ERMA is represented by 00, *tasks* by 10, *routines* by 20, etc. (*cf.* Figure 1). Actually, Column 1 entitled System Categories contains over 800 entries, that is 800 *subcategories* each related to one of the main *categories*. Figure 3 contains the identification and the key for a representative portion of the entire list of *Categories* and *Sub-Categories*. A numeric code is used throughout.

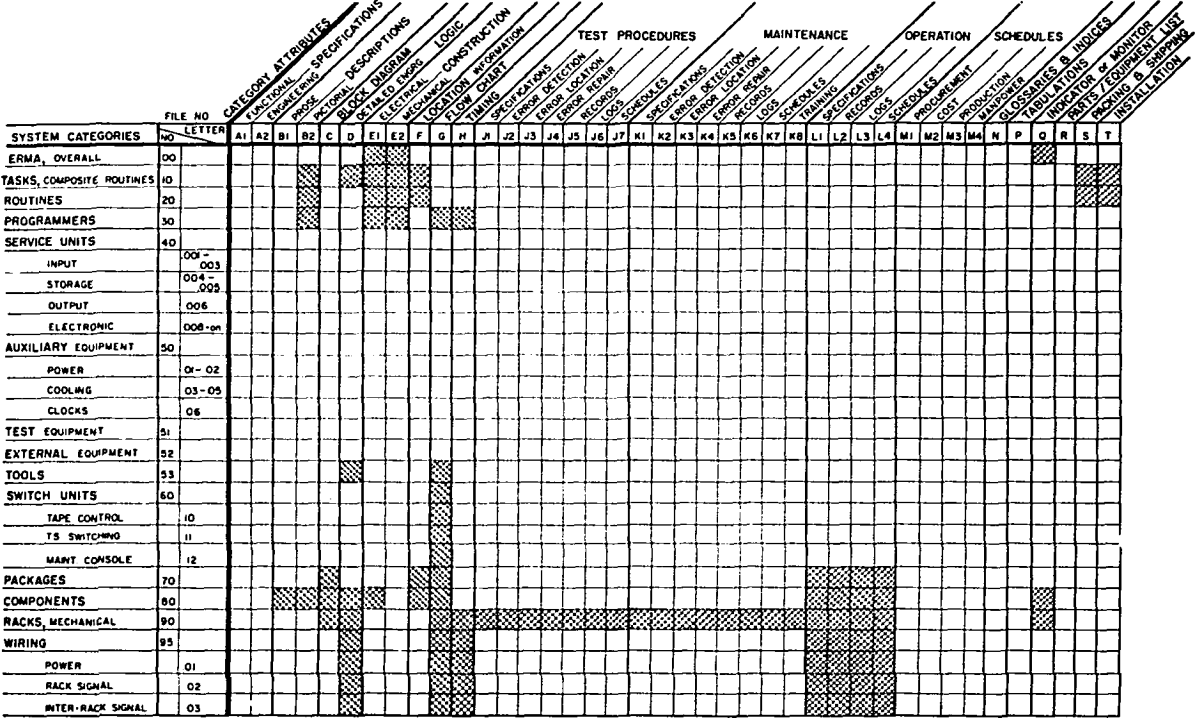

Fig. 1. The File System Framework

Coordinate 2

For this coordinate, names of attributes were sought which would be significant for all categories, where possible, and which would be sufficiently exhaustive to allow the filing of all generated and received documents. Such a list of category attributes is characteristic of the nature of a project or field of study. These attributes are listed on the horizontal coordinate. Success in selecting universally applicable attributes is measured by the number of shaded intersections in Figure 1. These shaded intersections correspond to attributes which do not apply to a particular system category. Attribute codes are alphanumeric.

All file numbers are composed of two parts: the first is the number designating the category or subcategory; the second is the letter code for the attribute; the two are connected by a hyphen. Thus 40.075-A1 is the file number for functional specifications of parity checkers, while 40.075.08-H is the file number for timing diagrams of parity checker #8 (*cf.* Figure 3).

Originally, code numbers for the system categories had two decimal digits. There were eleven numbers assigned to these categories; the first ten were 00, 10, 20, 30, 40, 50, 60, 70, 80, and 90. 95 was used for the eleventh. Entries 51, 52, and 53 were added later. Expansions of each category code were made decimally. A decimal point was placed after the category-code where subcategory codes were listed numerically to correspond to an alphabetical list of the subcategory names. Occasionally sub-subcategories were coded, also decimally. Thus service units called *Parity Checkers* are grouped under the subcategory code 40.075, with the ten parity checkers differentiated from one another by the use of sub-subcategory codes 40.075.01 through 40.075.10 (*cf.* Figure 3).

Because of their size, it is often inconvenient to file drawings with other records. Drawing transparencies create the additional problem of preventing creases during handling. Nevertheless, the category key was maintained for drawings, except that a prefix and a suffix were added to the key. The two notations added were: (1) a letter, A, B, C, D, or R preceding the file number indicates the standard paper sizes so that the drawings can be sorted by size

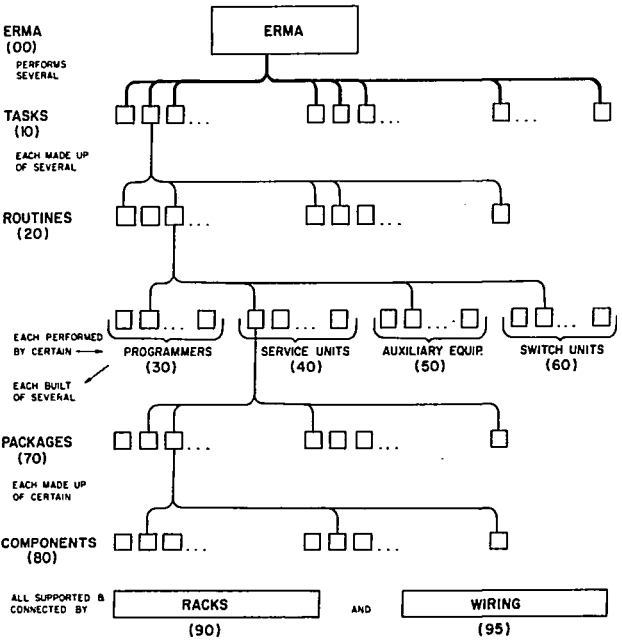

Fig. 2. The Erma Mark I Hierarchical Structure.

before being filed within each size group; (2) a lower case identification letter following the number (and separated from it by a hyphen) identifies the drawing from others that may be on the same subject. Thus the drawing number D 40.103.01-G-c is the third different flow chart drawn of the Temporary Storage sorter, Main Unit, and it is on "D" size paper. Revisions are not indicated by the file number.

In a large-scale computer project, various kinds of lists or tabulations are required and produced. In the ERMA project these invariably consisted of listings of one category or subcategory with respect to another. Examples are:

(1) A listing of the quantities of each standard package in ERMA was filed under 00-P(70), where the parenthesis contained the number corresponding to the type of item being tabulated. This was done only in the case of tabulations, that is, only after the attribute letter P. It is evident that without following parentheses the file number 00-P represents many different types of lists and tables about ERMA.

(2) A listing of the number and type of Resistors used in the Register package would bear the file number 70-P(80.26). Thus the first two digits of the file number represent the item of principal interest about which a descriptive list has been made; the item type being listed is represented by the number in parentheses.

Figure 4 is self-explanatory. It shows the arrangement of the records in the file folders.

This file system is "closed" in the sense that it is bounded by the decimally integral coding of categories from 00 to 100 and by an exhaustive listing of generic attributes and categories. However, it is expandable in the sense that, when necessary, new entries are permitted within the subcategories and parameters without resorting to modification of the original file structure. For example, categories *test equipment* - 51, *external equipment* - 52, and *tools* - 53, were added later in the only reasonable

40.067		NEW TAPE CLASSIFICATION GATES
40.068		NEW TAPE WRITE COUNTER AND CONTROLS
40.069		NODULE TIMING UNIT .3
40.070		NODULE TIMING UNIT .4
40.071		NON-AN FLEX WORD TRANSFER UNIT
40.072		(UNASSIGNED CATEGORY)
40.073		OLD TAPE READ COUNTER AND CONTROL
40.074		OVERFLOW SENSER
40.075		PARITY CHECKERS
	40.075.01	PARITY CHECKER #1
	40.075.02	PARITY CHECKER #2
	40.075.03	PARITY CHECKER #3
	40.075.04	PARITY CHECKER #4
	40.075.05	PARITY CHECKER #5
	40.075.06	PARITY CHECKER #6
	40.075.07	PARITY CHECKER #7
	40.075.08	PARITY CHECKER #8
	40.075.09	PARITY CHECKER #9
	40.075.10	PARITY CHECKER #10
40.076		PRE-ROUTINE SWITCHING UNIT TIMER
40.077		PROGRAMMER B CLOCK GATING SERVICE UNIT
40.078		PRINTER BIT AND COLUMN COUNTER
40.079		PRINTER CODE WHEEL STATISIZORS AND SERIALIZERS
40.080		PRINTER CONTROLS, MISCELLANEOUS
	40.080.01	CODE WHEEL DATA / COMPLEMENT
	40.080.02	FORMAT RELAY TIMER
	40.080.03	GATE BAR SENSER
	40.080.04	SET AND RESET

Fig. 3. A Representitive List of Categories

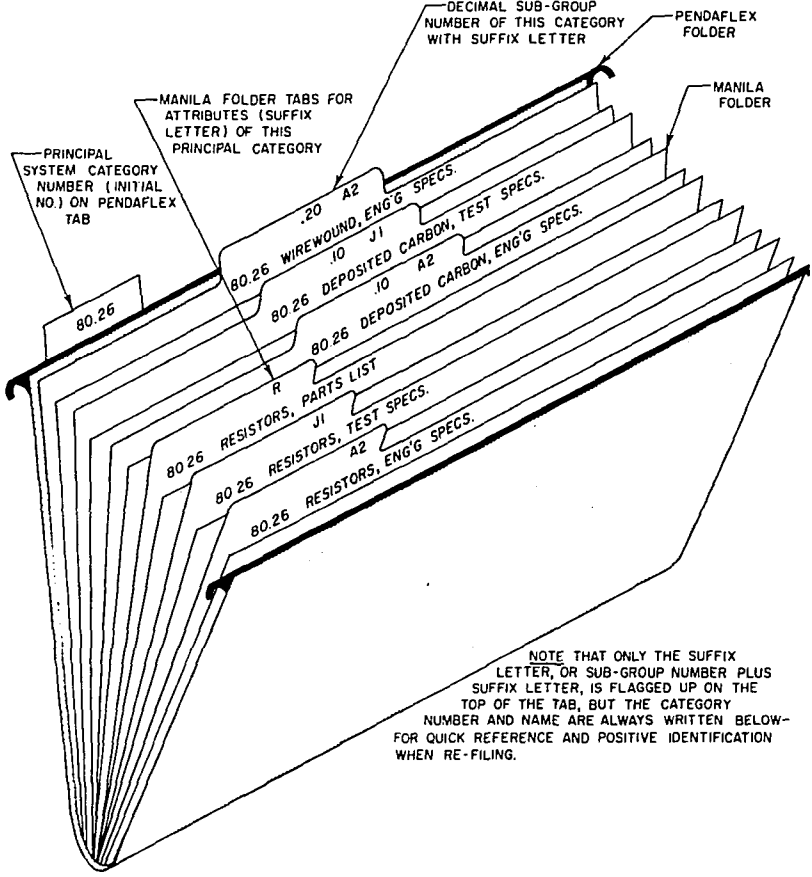

Fig. 4. Records Storage and Arrangement

location; in the 50 series, *auxiliary equipment*. This method of adding new categories to the file can be likened to the use of the familiar accordian file in which a suitable pocket can be stretched to fit new material, but the structure of the containing folds of the file and its major pockets remain unchanged. This is in contrast to a file of rigid construction in which the only recourse for addition is to append material or to make a new container.

Technical Management of Research and Development Projects

A filing system such as this is a most important tool for technical management. Recognition of this fact is obviously a prerequisite for the success of the system. As is the case with the introduction of any new method or procedure into any organization, management must give it acceptance, understanding, and support.

Supplementary advantages of this system are as follows:

(1) It provides an integrated, coordinated, and controlled record of project activity.
(2) It reveals the paper and hence the activity which needs recording and follow-up. It literally reveals "the holes" in the project.
(3) It compels structuralization of the project concepts. In order for system categories to be specified, structure specification is essential.

The authors believe that the principles of organization of this system are applicable to many technical projects.