

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Annals of Economic and Social Measurement, Volume 4, number 3

Volume Author/Editor: NBER

Volume Publisher: NBER

Volume URL: <http://www.nber.org/books/aesm75-3>

Publication Date: July 1975

Chapter Title: A Standard Description for Magnetic Tape Files

Chapter Author: Harold King, Mitchell Krasny

Chapter URL: <http://www.nber.org/chapters/c10410>

Chapter pages in book: (p. 447 - 454)

COMPUTER NOTES

A STANDARD DESCRIPTION FOR MAGNETIC TAPE FILES

EDITOR'S NOTE

The following article is the direct result of a workshop on Documentation of Large Machine Readable Data Sets, sponsored by the NBER's Conference on the Computer in Economic and Social Measurement. The Conference, chaired by Charlotte Boschan, was held on April 18-20, 1974 at New York University. It was divided into six separate workshops, (1) Standards for Description of Storage Media, (2) Documentation for Interactive Use of Time Series Data Bases, (3) Bibliographic Aspects of Documentation, (4) Establishment and Management of Data Libraries, (5) Standards for Text Documentation of Social Science Data Bases, and (6) Potentials and Problems of Data Base Documentation in Machine Readable Form.

The standards for the description of magnetic tape files recommended in the present article were discussed in Workshop #1, developed by a subcommittee, and circulated among the members of the workshop for further suggestions.

Members of the Subcommittee were:

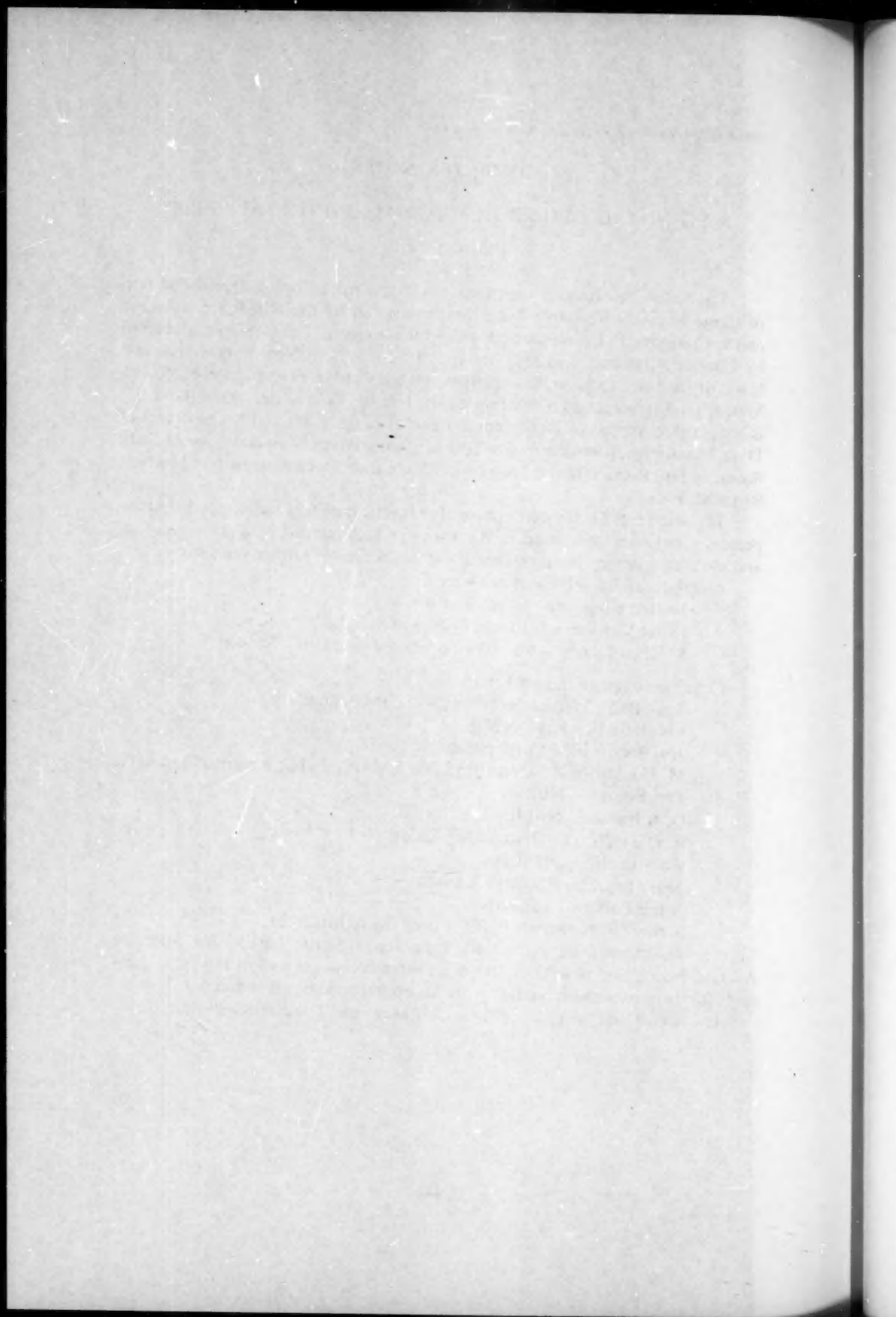
Harold King—The Urban Institute
Hazel McEwen—National Bureau of Standards
Mitchell Krasny—National Technical Information Service

Other members of the workshop:

Jerry Bell—Systems Software Div., Census Bureau
Charlotte Boschan—NBER
Joel Brest—IBM Corporation
M. E. Carroll—Machine Readable Archives, Public Archives of Canada
Teri Echaniz—NBER
Orin Hansen—NBER
R. D. Heilbron—Office of Education
Gary L. Hill—DUALabs
Mary Lennox—Statistics Canada
Bruce Phillips—Rapidata
Stewart Robinovitz—ICPR, University of Michigan

The recommended form is available from Helen Smith, NBER, 261 Madison Avenue, New York, New York 10016. Its extensive use, as well as further suggestions for its improvement would be greatly appreciated by all concerned.

It is hoped that the results of other workshops will be published soon.



A STANDARD DESCRIPTION FOR MAGNETIC TAPE FILES

BY HAROLD KING AND MITCHELL KRASNY

The National Bureau of Economic Research's Conference on the Computer in Economic and Social measurement sponsored a workshop on Documentation of Large Machine Readable Data Sets, April 18-20, 1974. This article describes a standard summary form for the description of files recorded on magnetic tape reels, which was developed at that workshop. The form covers a physical and magnetic description of the recorded data. It does not describe file content. This form is being considered for Federal use as part of the Federal Information Processing Standards (FIPS) Program of the National Bureau of Standards.

INTRODUCTION

During the last few years there has been a large increase in the number of machine readable data files being made available to the Social Science community. The medium for exchange, for the most part, continues to be via magnetic tape reels. One of the major shortcomings of these transactions has been the lack of a common descriptive summary which would tell the recipient what is on the tape in sufficient detail to allow him to process the file.

The problem is complicated by the continued use of non-standard practices for both hardware and software aspects. For example, a tape file generated on a UNIVAC 1108 using Uniservo IIIA tape drives and the FIELDATA character set is not directly readable on an IBM 360 using IBM 2400 tape drives, and the EBCDIC character set.¹ Two conversions must take place. To perform these conversions, the recipient of the file needs to know the facts contained on the attached form (Figure 1).

MAGNETIC TAPE FILE DESCRIPTION SUMMARY FORM

A general guideline which was used in designing the summary form was that it include only those items which were necessary to allow the recipient of a data file on magnetic tape to copy, dump or convert the file on a computer of his choice. The form was made applicable to multireel files, but is restricted to the description of a single file. If a reel has more than one file (multi-file reel), then a separate form should be completed for each one.

ADMINISTRATIVE DATA

The first eight items on the form (see Figure 1) are of an administrative nature. Item 1 is the date on which the summary form was prepared. Item 4 is the date on which the tape which is being transmitted with this form was recorded. This may not be the date on which the data was actually created unless one is sending his original source copy or duplicating it immediately after creation.

¹ The tape file is not directly readable even on a UNIVAC 3 with the same tape drives.

MAGNETIC TAPE FILE DESCRIPTION SUMMARY

01. SUMMARY DATE YR. MO. DAY			02. SUMMARY PREPARED BY (NAME AND PHONE)		03. EXTERNAL LABEL (REEL NO.)
04. RECORDING DATE YR. MO. DAY			05. FILE NAME		
07. SUBMITTING ORGANIZATION AND ADDRESS			06. SHORT TITLE (FILE ID)		
09. NUMBER OF TRACKS <input type="checkbox"/> 7 <input type="checkbox"/> 9 <input type="checkbox"/> OTHER (Specify)			10. RECORDING DENSITY		
11. NUMBER OF REELS			12. TAPE LENGTH (FT.)		
13. TAPE WIDTH <input type="checkbox"/> 1/2 IN. <input type="checkbox"/> OTHER (Specify)			14. RECORDING METHOD <input type="checkbox"/> NRZ <input type="checkbox"/> PE <input type="checkbox"/> OTHER (Specify)		
15. PHYSICAL RECORD LENGTH (Maximum) <input type="checkbox"/> CHAR. <input type="checkbox"/> BYTES <input type="checkbox"/> WDS. (____ BITS/WD.)			16. PHYSICAL RECORD TYPE <input type="checkbox"/> FIXED <input type="checkbox"/> VARIABLE <input type="checkbox"/> MIXED		
17. NUMBER OF PHYSICAL RECORDS (BLOCKS)			18. LOGICAL RECORD SIZE <input type="checkbox"/> CHAR. <input type="checkbox"/> BYTES <input type="checkbox"/> WDS. (____ BITS/WD.)		
19. LOGICAL RECORDS PER PHYSICAL RECORD			20. ERROR DETECTION TECHNIQUE <input type="checkbox"/> CRC <input type="checkbox"/> LRC <input type="checkbox"/> ODD PARITY <input type="checkbox"/> EVEN PARITY <input type="checkbox"/> OTHER		
21. CHARACTER SET <input type="checkbox"/> ASCII <input type="checkbox"/> BCD <input type="checkbox"/> FIELDATA <input type="checkbox"/> OTHER (Describe on separate page) <input type="checkbox"/> EBCDIC <input type="checkbox"/> X53					
22. RECORDED LABEL <input type="checkbox"/> NONE <input type="checkbox"/> STANDARD <input type="checkbox"/> OTHER (Describe)			23. CONTROL CHARACTERS <input type="checkbox"/> STANDARD <input type="checkbox"/> OTHER (Describe)		
24. LABEL ID					
25. RECORDING EQUIPMENT ID			26. I/O CONTROL SYSTEM		
MFR/CPU DESIGNATION		TAPE DRIVE	NAME/RELEASE		LEVEL/TYPE
27. FOR SUBMITTING ORGANIZATION USE					

Figure 1 Magnetic tape file description summary

Items 2, 7 and 8 ask for information about the organization and the people who were involved in preparing the form and recording the tape. Item 2 is the name of the person who prepared this form, and who can answer questions about the entries, if they arise. Item 7 is the name and address of the organization which is

sending the tape covered by this form. Item 8 is extremely important. It is the name and phone number of the individual who had the technical responsibility for recording the tape. This is the person who would be contacted if the computing center which was attempting to read the file ran into trouble of a technical nature. This person could also have technical knowledge about the subject-matter content of the data, if not, he would know the person to contact.

Items 3, 5 and 6 are labeling information about the file. Item 3 is the physical label identifier which appears on the tape reel (or cover) and which is used when referring to the reel. This number will be used to identify the reel containing the file for which this summary has been prepared. Item 5 is a descriptive title for the file such as "Current Population Survey—March 1973" and item 6 is a short title (CPS73).

PHYSICAL AND MAGNETIC DESCRIPTION

Items 9 thru 24 describe the physical and magnetic characteristics of the tape file. Some of these items might not be known to the person preparing the summary, but the information should be available from the technical contact. In some cases, the information may be available only from the staff of the originator's computer center. It is important that all of these items be determined and the proper information entered. Since some of these items remain constant for an installation, once they have been identified, they can then be used on any further forms which need to be completed.

Items 9 thru 13 are characteristics which are easy to determine. Items 9 and 10 describe magnetic criteria which are fairly standard. These are number of tracks which are usually either seven (7) or nine (9), and the recording density. These items may even be specified by the recipient if the originator can meet the options. Items 11, 12 and 13 are physical characteristics which can be determined by external examination of the reels. Tape length (item 12) may offer a slight problem. Some installations will recondition tapes by cutting off the beginning (which tends to be the only part which gets much wear), and recording on the remaining tape. In such cases, it is helpful to a computer operator to know that he does not have a reel of one of the standard lengths (600, 1200, 2400 or 3600 ft).

Items 14 and 20 represent information about the tape drive that most programmers (and some operators) are not aware. Item 14 is the recording method. A 9 track/1600 cpi tape drive uses the "Phase Encoded" (PE) recording method. Most other industry compatible tape drives (i.e., all 7 track drives and 9/track 800 cpi drives) use the "non-return-to-zero-interchange" (NRZI) recording method. The manufacturer's representative should be able to clarify this point, if someone in the operational environment doesn't know.

Item 20 is fairly straightforward, for industry compatible tape drives. Most 9 track tape drives use odd parity;² 7 track tape drives use either odd or even parity, although even is usually used unless binary zero characters can occur in the data. The "Longitudinal Redundancy Check" (LRC) is an extra frame added at the end of each tape block by the tape controller; it is used on all but 9 track/1600 cpi tape

² Honeywell has announced a 9 track drive for which "even" parity is an option. No further information was available at the time of this writing.

drives. The "Cyclical Redundancy Check" (CRC) appears only on 9 track/800 cpi tape, and is the byte immediately preceding the LRC. Since both these frames are generated by and used by the tape controller, it is not important that the user know any more about them. They are included in the summary form only to warn those involved in an interchange with non-industry compatible tape drives that problems could occur.

Items 15 thru 17 refer to physical records, as distinguished from logical records, which are described in items 18 and 19. Physical records are called blocks, and are defined to be the information recorded between "inter-record gaps".

Item 15 asks for physical record length. If the physical records are variable in length, the maximum length should be specified. The length can be specified in characters, bytes, or words. If words are used, enter the number of bits per word in the appropriate space. The distinction between characters and bytes is usually equivalent to the distinction between 7 track and 9 track tape, but it is conceivable that one could record 6 bit characters on 9 track tape, or 8 bit bytes on 7 track tape (UNIVAC's UNIVERVO IIIA tape drive records a 36 bit word in five 8 bit frames, so that density can be expressed as 1000 frames per inch or 1200 characters per inch; IBM 7 track tape drives on the System 360 or System 370 have an optional data conversion mode in which three 8 bit bytes are written as four 6 bit characters on tape).

Item 16 is the Physical Record Type. Normally, records are either fixed length or variable length. The term "mixed" was added after some internal debate. In a file with only fixed length physical records, it is not necessary to attach the physical record length to each record; in a file with blocked variable record lengths, it is necessary to have the record length specified. The situation can arise where some of the records in a file are fixed length and some are not. Use "mixed" in such cases.

Item 17 asks for the number of physical records or blocks. This number is usually provided by the utility program creating the copy. The program creating the original tape would normally count the number of logical records, which would have to be divided by the blocking factor (item 19) to develop the number of physical records. It is good practice to record this number on the physical label of the tape. If it is not known when the summary form is being filled out, it should be entered by the generating organization before the tape is forwarded to the recipient.

Item 18 is the logical record size, and is very similar to item 15, except that it refers to the logical organization of data rather than the physical organization. In general, logical records are "blocked" into physical records. For instance, one might block ten 80-character card images into one physical record. In this case, item 15 would contain "800 characters" and item 18 would contain "80 characters."

Item 19 asks for the number of logical records per physical record. In the above example, it would be 10. Normally, any two of items 15, 18 and 19 will enable one to derive the third. Item 19 is included both for error detection and to highlight the case when a logical record extends across two or more physical records.

Item 21 asks for information about the character set used to record the information on the tape. The first character set identifies the standard code approved by the American National Standards Institute (ANSI). ASCII (American Standard Code for Information Interchange) is a seven bit character code which allows for

the encoding of 128 graphic characters and controls. This is sufficient for all upper and lower case letters, ten digits, punctuation marks and many special control symbols.

Extended Binary Coded Decimal Interchange Code (EBCDIC) is IBM's own standard character set. It is an eight bit code and, hence, is capable of representing 256 different code groups.

Binary Coded Decimal (BCD), Excess-Three (XS-3) and FIELDATA are all 6 bit codes. XS-3 and FIELDATA are used on UNIVAC equipment.

If the data on the tape was recorded in a character set other than the above, the set should be described on a separate page. The best technique for describing the character set is to identify the punch card code and binary code for each character in the set.

A recorded magnetic tape may, or may not, contain a machine readable label. If a label is contained on the tape it may, or may not, be recorded in the manner approved by the American National Standards Institute (ANSI). Item 22 asks for this information. If the label is other than the ANSI standard, then it should be described in the space provided. This information is important in that many facilities are set up to process tapes with labels devised by their own staff. In other words, they cannot readily process tapes with "foreign" labels. Tapes with non-facility "standard" labels must usually be initialized by the facility prior to being processed.

If the tape has a label, then the label identification should be placed in item 24. This would be the recorded data set name, and could be the same as the file ID described in item 6. Since many facilities automatically check the label ID to insure that the proper file is being accessed, it is necessary that the name entered in item 24 be an exact representation of the name as it appears on the recorded label.

Item 23 asks for control characters. This could be extremely important when the tape is going from one manufacturer's computer to another. For example, the "word mark character" on a 1401 does not perform that function on any other computer which can read a tape generated by the 1401. Similarly, "end of file" on some computers is not recognized as such on others. In some systems, a character is recognized by the printer as "end of line." All such special function characters must be identified so that the recipient can determine whether his system will recognize them in the same way, or whether he needs special software to provide these functions.

A description of the equipment used to record the tape (item 25) is especially useful whenever the tape was written on a manufacturer's system which is different from the one on which it is to be read. This will allow the recipient to determine whether it is possible to read the tape on his system directly or whether he must create special software to read the tape. In some cases it may be necessary to go through an intermediate conversion on a third system before the file is usable on the recipient's system. If a system has been emulated, this fact must be noted here with any clarification required being entered in item 27.

Most computers have Input/Output (I/O) Control Systems, which handle tasks such as tape label processing, reading and writing physical records, and blocking or deblocking logical records. These control systems have characteristics

that must be duplicated by the tape recipient's system in order that the file can be retrieved from the tape. In some cases the effect of these systems can be bypassed. As an example, one system may create a recorded label that is not acceptable by another system but the second system may be able to ignore the label. On the other hand, tapes created by UNIVAC's LION cannot be deblocked by IBM's IOCS, and vice versa. To be able to determine what the system characteristics are, the tape recipient needs to know the Name and release, and level and type for the I/O control system used to record the tape. This information should be entered in item 26.

Item 27, For Submitting Organization Use, is to be used to clarify any information on the form which needs further description. This space should also be used for any handling restrictions which are applicable to the file. This would be such things as file classification and file retention. The space can also be used for any other information which the submitting organization feels would be useful to the recipient in processing the file.

ACKNOWLEDGEMENTS

The writers wish to acknowledge the efforts by Hazel McEwen of the National Bureau of Standards in assisting them as a member of the committee which developed the form described in this paper. The recommended form is available from Helen Smith, NBER, 261 Madison Avenue, New York, N.Y. 10016.