Socialist Countries: Communicative Format of Data Recording on Magnetic Tape

MARTIN BLOCH, JIRI BUDIL,* IVAN GANICKY, and MARTIN SVOBODA

Central Office for Scientific, Technical and Economical Information, 113 57 Prague, Czechoslovakia

Received March 15, 1976

The achievement of socialist countries in standardization of record structure on magnetic tape as a base for automation in information interchange and the generalized types of utilizable record structures are described. Consequences of record standardization to software unification and objections to ISO 2709-1973 are discussed.

ISSTI is the International System of Scientific and Technical Information situated in Moscow. Member states of ISSTI are Bulgaria, Cuba, Czechoslovakia, German Democratic Republic, Hungary, Mongolia, Poland, Roumania, and USSR. The following propositions have been put forth by ISSTI:

1. A communicative record format standard on magnetic tape is the first step toward automation in information science.
2. A third generation concept should be preferred.
3. Modern computer information systems have to be studied.
4. Standard nine-track, 800 bpi, NRZI tapes should be preferred in ISSTI member states at the present time.

We have examined a number of record structures[1-9] which are used or defined by important world information producers and sponsors (Table I). The objective of this effort was to identify general features of record structures used for information interchange and processing. Therefore, little or no attention was paid to highly restricted or specially tailored structures, as, for example, MEDLARS, DERWENT, API, IPC Appleton, etc.,[10-13] or to those not connected with prospective considerable information production or sponsorship, e.g., interesting but minor software systems. Some terminology follows.

Record is a collection of items (data elements) related to one object of description and organized in some manner.

Data element has two facets: its meaning (e.g., author) and the value (e.g., Smith, F. D.). The meaning can be inferred from the fixed position and the length of the data element within each record, or some denomination of the meaning (*data element name* or *tag*) can be stored in the record along with the data element value. Data elements of the first type are called *positional data elements*; the second type are called *named data elements*.

Data element names are often stored separately from the data element values in the portion of a record called *directory area*. Data element values are then stored in another portion of a record called *value area*. One *directory entry* contains data element name and control information necessary to access the data element value (i.e., its position and length). This technique is implemented in the record types L and M (Table II and Figure 1b,c).

The head of the record contains control information necessary to process the record; this portion of the record is called *record description area*. Another possibility is to store the data element name adjacent (usually prefixed) to the data element value and enter the common length of this aggregate. This technique is implemented in the record type K (Table II and Figure 1a). We examined many other structures, but their common disadvantage was the lack of generality imposed by tough presumptions concerning the number, length, or representation of data elements.

The major disadvantage of the K type is the absence of a directory. This may bring better economy in space, but excludes more refined and faster methods of access to the data element names (e.g., binary search).

Table I

Type	Used/defined by
K	ERIC, *Excerpta Medica*
L	CAS, ISSTI
M	LC/MARC-II, INIS, AGRIS, INSPEC, FSTA, Engineering Index
	Standards: ANSI Z39.2, BS 4748, ISO 2709

Table II

Type	IBM syst concept	Record format[a]	Control information	Directory oriented	Positionally oriented
K	360/370	V	Binary	No	Yes
L	360/370	V	Binary	Yes	Yes
M	1401/1410	U	Decimal Characters	Yes	No

[a] According to ISO R 1001[14]: U = undefined record format. Its length is variable and is *not* explicitly expressed elsewhere in the record. V = variable record format. Its length is variable. The first 4 bytes express the length (bytes 0 and 1) and spanning indication (bytes 2 and 3) in binary form.

The disadvantage of the M type is the fact that system concept represents second generation and thus is not suitable for internal processing on third generation systems. For instance, INIS does not use its M type structure for its own internal processing because it cannot provide features necessary for effective work of computer systems; they use a structure very similar to the L type instead.[15] Weak points of the M type (as defined by ISO, ANSI, etc.) enable the implementers to differ in certain important features which cause troubles in practical applications.

The problems we experienced in starting up processing of three magnetic tape services (INSPEC, INIS, COMPENDEX) concurrently were due to the different record formats used despite the fact that all claimed to adhere to the same ISO 2709, yet each adopted its own "buts" to this standard. This resulted from the vague definition of the record structure in ISO 2709 which led to the hidden pseudo-standardization. Although ISO 2709 has been in effect since 1973, the setting up of the committee ISO/TC46/SC4/WG4 as early as in May 1974 to provide eventual revision of ISO 2709 clearly manifests users' dissatisfaction with this standard.

Czechoslovakia's disapproval of ISO 2709 in 1972 initiated an extensive discussion on its effective applicability in ISSTI. Having succeeded in convincing ISSTI that ISO 2709 format is unsuitable both for information interchange and for internal processing, Czechoslovakia submitted another draft. This draft has been studied and modified by an ISSTI working group and finally was officially declared as an ISSTI standard in October 1974. All bibliographic data interchange between ISSTI member states will be accomplished using this format. The biggest producer of such data, Soviet VINITI, is soon expected to start supplying bulk data concerning chemistry and electronics in the format according to this standard. We

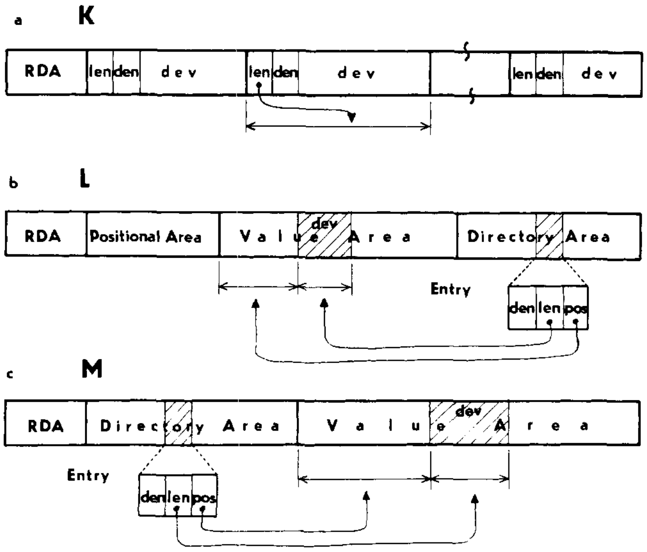

Figure 1. Record types K, L, and M: den = data element name, dev = data element value, len = length of the data element value, pos = position within value area, and RDA = record description area.

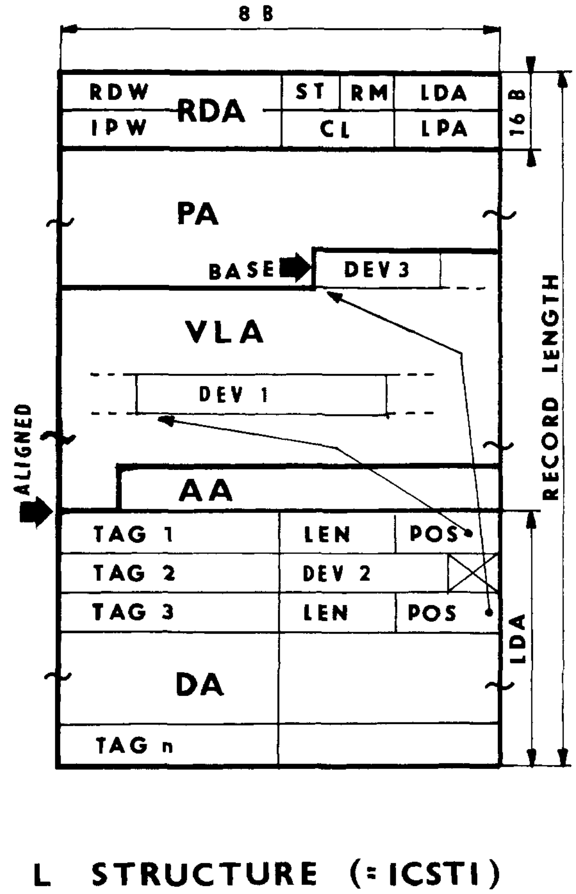

Figure 2. The L-type structure layout. RDA = record description area, PA = positional area, VLA = value (long) area, AA = auxilliary area, DA = directory area, RDW = record descriptor word, ST = service token, RM = record modifier, LDA = length of DA, IPW = internal processing word, CL = common length = LPA + length of VLA, LPA = length of PA, DEV = data element value, LEN = length of DEV, POS = position (starting), and TAG = tag.

believe that the information explosion can be faced using large computers; therefore a standard record structure was designed especially to suit their requirements.

The properties of the L type enable this structure to be used not only for information interchange but also for effective internal processing on third generation systems utilizing the byte concept. The most important design features of the L type (see Figure 2) are: this technique has full compatibility with standard data management systems (e.g., IBM 360/370 OS/DOS); all control information is expressed as aligned binary numbers (its processing is thus as fast as possible); the data can be stored as named or positional data elements; the named data element can only be accessed via its entry in the directory area, which provides for necessary variability in data element number, length, and representation (e.g., code); the positional data element (or elements) is located in the positional area and can be accessed directly through its fixed position and length. Access to the positional data elements is very fast and moreover allows for application of "classical" data processing techniques.

These interesting properties of the L type intentionally serve as a base for software unification. The software package called USS (Unified Software System) which is based on the L-type structure has been developed and is further being improved in our institute. USS operates presently in batch mode under IBM/360 OS and is written in Assembly Language and PL/1. USS is a small but comprehensive package which is capable of covering approximately 80% of the functions required in any typical processing in information science.

ACKNOWLEDGMENT

We express our appreciation to Mr. Barrie T. Stern, Director of Excerpta Medica, Database Division, for his assistance.

APPENDIX

UDC 002.532.13:681.3

INTERNATIONAL CENTER OF SCIENTIFIC AND TECHNICAL INFORMATION

ICSTI STANDARD

COMMUNICATIVE FORMAT OF DATA RECORDING ON MAGNETIC TAPE

S 2-74

Accepted by the Governing Board of Member States of ICSTI October 1974

This standard applies to information interchange on magnetic tape within International System of Scientific and Technical Information and states the structure of data recording and formal rules for displacement of data on the medium.

1. BASIC TERMS

1.1. Data element—basic unit of description of any object. Each data element involves its value and its determination. There are two types of data elements: positional and named.

1.1.1. Positional data element—an element whose value can be accessed by means of its position in the record.

1.1.2. Named data element—an element whose value can be accessed by means of its tag.

1.1.3. Tag of the data element—a set of six hexadecimal digits identifying the named data element.

1.2. Record (logical record)—a logical unit of processing involving an ordered set of data elements describing one object and mutually related from viewpoint of programming technique.

1.3. Block (physical record)—physical unit of processing of a set of records.

2. FORMAT STRUCTURE AND ITS TECHNICAL REQUIREMENTS

2.1. Block.

2.1.1. The block consists of the block control word and a group of complete records (see Figure 3). Block length (in

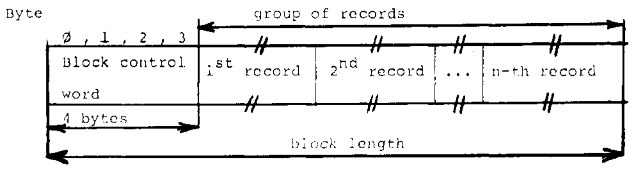

Figure 3.

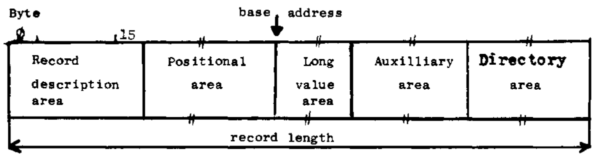

Figure 4.

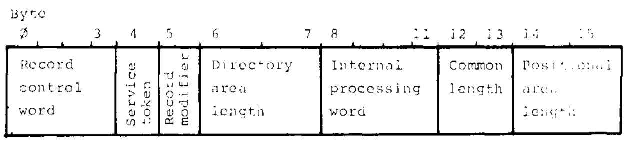

Figure 5.

bytes) must lie within interval ⟨2Ø,32756⟩.

Note: If not agreed otherwise, it is recommended that, for technical reasons, block length does not exceed 4Ø92 bytes.

2.1.2. Block control word containing basic control information is located in bytes Ø–3 of the block. Bytes Ø, 1 contain a fixed point binary integer determining the block length including the length of the block control word. Bytes 2 and 3 are reserved for the operating system and must contain binary zeros.

2.1.3. The group of records must contain at least one record.

2.2. Record.

2.2.1. A record is a variable length record and consists of areas ordered in the following sequence (see Figure 4):

record description area
positional area
area of long values
auxilliary area
directory area

All areas, except for record description area, are variable or null length areas. Record length is evenly divisible by eight. Maximum record length is 32 752 bytes.

Note: If not agreed otherwise, it is recommended that, for technical reasons, record length does not exceed 4Ø88 bytes.

2.2.2. Record description area is fixed length area (16 bytes) and is located at the beginning of the record in bytes Ø–15. This area consists of the field ordered in the following sequence (see Figure 5):

record control word
service token
record modifier
directory area length
internal processing word
common length
positional area length

2.2.2.1. Record control word—field containing basic control information. It is located in bytes Ø–3 of record description area. Bytes Ø, 1 contain a fixed point binary integer determining the record length including length of record control word. Bytes 2 and 3 are reserved for operating system and must contain binary zeros.

2.2.2.2. Service token—field used as an indicator of coding method or of the structure used for representation of all the data elements in the record. It is also used as an indicator of availability of the record. Service token is located in byte 4 of record description area (see Table III).

Table III[a]

Contents	Meaning
ØØ	Method of coding of named element value is specified by the contents of the field"code mode"
Ø1–ØF	Reserved
1*	Reserved
2*	EBCDIC
3*	Reserved
4*	Packed decimal number
5*	Reserved
6*	Fixed point binary integer in 4 bytes
7*	ISO code
8*	Floating point binary in 4 bytes
9*	Bit string
AØ–FE	Reserved
FF	Record not available

[a] The symbol * represents any hexadecimal digit. It is suggested that all values of data elements be coded in EBCDIC (or corresponding code CMEA). In the interval Ø1–FE the contents of the field "code mode" is ignored.

Table IV

Contents	Meaning
ØØ	Reserved
Ø1	Data element tags in any sequence
Ø2	Data element tags in ascending sequence
Ø3	Data element tags in descending sequence
Ø4–FF	Reserved

2.2.2.3. Record modifier—field determining organization of directory entries (see Table IV). It is located in byte 5 of the record description area, and it is given by two hexadecimal digits. Ordering of directory entries is related to the beginning of the directory.

2.2.2.4. Directory area length—field containing the length of the directory area in bytes in the form of binary fixed point integer. It is located in bytes 6 and 7 of the record description area.

2.2.2.5. Internal processing word—field containing zero in the form of binary fixed point integer. It is located in bytes 8 through 11 of the record description area.

2.2.2.6. Common length—field containing the sum of positional area and long values area lengths in the form of fixed point binary integer. It is located in bytes 12 and 13 of the record description area.

2.2.2.7. Positional area length—field containing the length of the positional area in the form of the fixed point binary integer. It is located in bytes 14 and 15 of the record description area.

2.2.3. Positional area—area containing values of positional data elements. It follows the record description area.

2.2.4. Long value area—area containing the values of the named data elements in the long form. It follows the positional area.

2.2.5. Auxilliary area—area designed to fulfill the requirement that the address of the beginning of the directory related to the beginning of the record is the minimum number evenly divisible by eight. It is permissible to lengthen this area by a number of bytes which is evenly divisible by eight. Any additional usage of the auxilliary area is possible and is to be agreed on especially.

2.2.6. Directory area—area containing variable number of fixed length directory entries. This area is located at the end of the record. The address of the beginning of the directory area related to the beginning of the record should be evenly divisible by eight.

2.2.6.1. There are two types of directory entry structures, whose length is always 8 bytes: structure of entry for the short form of the named data element and structure of entry for the

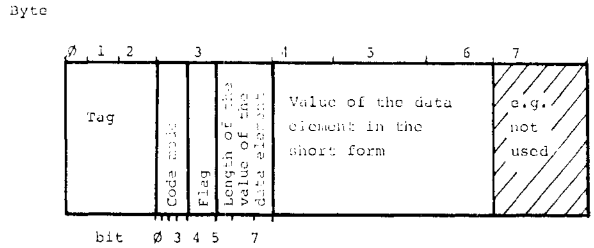

Figure 6.

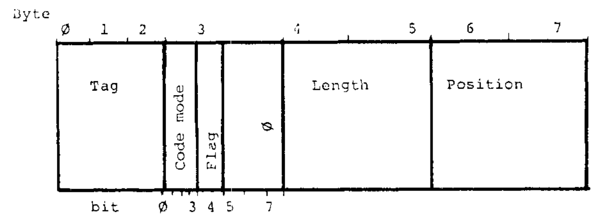

Figure 7.

long form of the named data element. In the first case the entry refers to proper value of the data element which is placed within this entry. In the second case the entry refers to proper value of the data element which is placed in the long value area.

2.2.6.2. Entry for the short form contains fields ordered in the following sequence (see Figure 6):

tag
code mode
flag
length of the value of the element in the short form
value of the data element in the short form.

2.2.6.3. Entry for the long form contains fields ordered in the following sequence (see Figure 7):

tag
code mode
flag
field containing binary zero
length
position

2.2.6.4. Tag—field located in bytes Ø–2 of each entry. The structure of the tag is to be agreed on especially.

2.2.6.5. Code mode—field indicating the coding method or the structure in which the value of corresponding named data element is represented. This field is located in bits Ø–3 of byte 3 of each entry. The coding method or the structure is expressed by one hexadecimal digit (see Table V). The contents of code mode field are ignored if the value of service token field is not equal to zero.

2.2.6.6. Flag—field reserved for internal processing. It contains zero in case of information interchange. It is located in bit 4 of byte 3 of each entry.

2.2.6.7. Length of the value of the named data element in the short form—field containing the length (in bytes) of the value of the data element which is placed directly in the directory entry for the short form. The length is expressed

Table V

Contents	Meaning
Ø	All coding methods not corresponding to the values 1 through *F*
1	Reserved
2	EBCDIC
3	Reserved
4	Packed decimal number
5	Reserved
6	Fixed point binary integer in 4 bytes
7	ISO code
8	Floating point binary integer in 4 bytes
9	Bit string
A–E	Reserved for national usage
F	Value of data element in long form has the same structure as the record described by this standard

as a binary integer from the interval $\langle 1, 4 \rangle$ in bits 5–7 of byte 3 of the entry. For the long form of a named data element this field contains a binary zero.

2.2.6.8. Value of data element in the short form—field containing the value of a data element in the short form. It starts in byte 4 of the entry for the short form. The length of the field is 1 to 4 bytes. If the length of the field is less than 4 bytes, the remaining bytes of the entry are not used.

2.2.6.9. Length (see Figure 7)—field containing the length (in bytes) of the named data element located in the long value area. This field is located in bytes 4 and 5 of the entry for the long form and contains a fixed point binary integer.

2.2.6.10. Position (see Figure 7)—field containing a starting position of the value of the named data element in the long value area related to the base address. This field is located in bytes 6 and 7 of the entry for the long form and contains a nonnegative fixed point binary integer.

2.2.6.11. Base address—the reference point for addressing the values of long data elements in the record. This point corresponds to the beginning of the long value area.

LITERATURE CITED

(1) Specification Manual for Computer-Readable Files in Standard Distribution Format, Vol. 1, Chemical Abstract Service, Columbus, Ohio.
(2) ISO 2709 Documentation—Format for Bibliographic Information Interchange on Magnetic Tape, 1973.
(3) ANSI Z39.2, American National Standard for Bibliographic Information Interchange on Magnetic Tape, 1971.
(4) EUSIDIC Information Interchange Format, internal document, Nov 1970.
(5) INIS Magnetic Tape Specifications and Record Format. IAEA-INIS-9 (Rev. 0), Vienna, Jan 1970.
(6) INSPEC Tape Services Manual, Part 2, Reference Manual, INSPEC/IEE, London.
(7) COMPENDEX Technical Description and Specifications Manual, Engineering Index, Inc., New York, N.Y., 1976.
(8) ERIC Magnetic Tape Service. Technical Specification. Education and Research Information Centre.
(9) Specification of Magnetic Tapes for *Excerpta Medica* Tape Services, INFONET, Amsterdam, Holland, Nov 1975.
(10) MEDLARS. Internal documentation.
(11) DERWENT Central Patent Index, Tapes and Programs, Instruction Manual No. 5, 1974.
(12) API Magnetic Tape Service. Internal documentation.
(13) EIC Magnetic Tape Service. Internal documentation.
(14) ISO/R 1001 Magnetic Tape Labeling and File Structure for Information Interchange, 1969.
(15) INIS: Description of Computer Programs. IAEA-INIS-14 (Rev. 1), Vienna, 1974.