ON THE RECURSIVE DEFINITION OF A FORMAT FOR COMMUNICATION

Leonid N. SUMAROKOV: Head, Research Department, International Center for Scientific Information, Moscow, USSR

A recursive presentation of a communication format is discussed and a form of pertinent notation proposed. Recursive notation permits presentation of an interchange format in more general terms than heretofore published, and expands application possibilities.

The development of the forms of exchange of information among documentation systems, and particularly the development of the technique of recording machine readable bibliographic data on magnetic tape, has led to the requirement for the adoption of an agreement on a standard for a format for communication. Thus, the problem of a format for communication reflects the existing tendency toward ensuring compatibility among formats.

At the present time the greatest impact on world information practice has been caused by the American National Standard Institute (ANSI) *Standard for Bibliographic Information Interchange on Magnetic Tape* (1) and the several implementations of that standard: MARC, INIS, COSATI and others. It should be noted that, despite numerous existing peculiarities, in principle there is no difference in structure among the formats.

One of the most important requisites for a communication format is universality. The practice of processing large quantities of information has emphasized the flexibility of the above-mentioned formats; their use has permitted identification of huge numbers of documentary materials in

various forms, thereby creating the impression that the structure of the format has been developed to such an extent that it can be canonized for any application.

It must be said that support or rejection of this impression can be based only upon future experience in the application of a communication format. Nevertheless, it appears expedient to generalize about the structure of a communication format by making a few preliminary remarks and thereby contributing toward expanding the sphere of its application.

The remarks deal with the following. In the existing systems for interchanging information on magnetic tape, the document is the object of identification. With the development of data banks the characteristics of the objects to be identified may prove to be so varied, even though presented in the proper documentary form, that their uniform presentation will cause difficulties. (Actually, examples can be given of data banks in which data appear in the capacity of objects: information regarding firms, rivers, information about products of the electrical engineering industry, etc.). Furthermore, even if it is possible to identify in principle a certain object with the aid of the format, one must distinguish between the question of possible identification in principle, and that of the optimal (or rational) form of identification in view of the limitations of a certain system.

The recursive notation of a communication format is presented below. Certain definitions and ideas in general are used as source material for such a notation, using the American Standard for Bibliographic Information Interchange on Magnetic Tape (1). It must be conceded that the use of one term or another for defining individual elements of a notation, as well as the general structure of the entire notation, are not the principal subject of discussion here; this means that any change, either in definition or, to a certain extent, in the structure of the notation, will not affect the proposed form of the notation. Consequently, this article does not pretend to describe a certain universal structure for a communication format. It has a different purpose, *viz.*, to point out wider perspectives that will unfold by applying the recursive presentation of notations in formats at the expense of an object with any hierarchical depth.

For the following symbols explanations can be found in the ANSI *Standard* (1):

R=record
L=leader
Dr=directory
T=tag
D=data, or data elements
FT=field terminator, or field separator
RT=record terminator, or record separator

The concept TT used below, and standing for tag terminator, is analogous to FT and RT. So also is the concept SF, meaning specific fields for de-

fining contents that did not appear in the proposed notation although utilized in actual formats. The following symbols are also used:

TG=tag generalized
F=field
DF=data field
BF=bibliographic fields

Utilization of special notation in brackets (analogous to the form used in algorithmic languages) enables R to be defined in the form of the following consecutive structure:

1) $R=[L][Dr][SF][BF]$

The symbols written in brackets after the equal sign maintain the relationship of priority.

Further, the recursive universal tag TG is defined as follows:

2) $TG=[T;TT]$

Such a notation indicates that the expression in brackets is T or TT. The recursiveness of the notation indicates that it is possible that

TG is $T_1 T_2 \ldots T_p:TT$

where p is any whole number, a larger or an equal unit. (Obviously p defines the depth of the hierarchic description in accordance with the given characteristic.)

Finally

3) $F=:[TG][D]$;
4) $DF=:[F;FT]$;
5) $BF=:[DF;RT]$.

Thus, the general notation of the format is expressed by 1), in which the element BF, which constitutes the basic part of the so-called alternate fields, is expressed recursively with the aid of the system 2)-5). As is evident, the quantity F in DF, and DF in BF, as well as in the case of the subscripts TG, can arbitrarily be a whole number, changing from notation to notation.

REFERENCE

1. "USA Standard for a Format for Bibliographic Information Interchange on Magnetic Tape," *Journal of Library Automation*, 2 (June 1969), 53-65.