DOCUMENT RESUME

ED 071 678 LI 004 073

AUTHOR Lauterbach, Guy
TITLE A Cataloguing System for Machine Readable Data Bases.
PUB DATE 7 Jun 71
NOTE 15p.; (2 References)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Cataloging; Catalogs; Computers; Computer Science; *Data Bases; Electronic Data Processing
IDENTIFIERS *Machine Readable Cataloging

ABSTRACT

With the fantastic growth in computerized data processing and management, there arises a great need for improved techniques in cataloging of machine readable data bases. The purpose of this report is to define a system by which computerized data bases may be cataloged for easy reference and availability. Developed from a computer scientist's viewpoint, emphasis was placed on identification of what information should be included in or excluded from such a catalog. A glossary is also included to provide a standard reference base. The objective of the proposed cataloging system is to provide the potential user with information which would help him decide whether or not he would want to use a particular data base. (Author)

A CATALOGUING SYSTEM FOR MACHINE READABLE DATA BASES

by
Guy Lauterbach
[Student, Oregon State University]

June 7, 1971

Table of Contents

I. Introduction

With the fantastic growth in computerized data processing and management, there arises a great need for improved techniques in cataloguing of machine-readable data bases. The purpose of this report is to define a system by which computerized data bases may be catalogued for easy reference and availability. Developed from a computer scientist's viewpoint, emphasis was placed on identification of what information should be included in or excluded from such a catalogue. A glossary is also included to provide a standard reference base.

The objective of the proposed cataloguing system is to provide the potential user with information which would help him decide whether or not he would want to use a particular data base.

It is hoped that this report may serve as a first step in development of cataloguing procedures for use by libraries and other agencies.

II. Included Information

The proposed cataloguing system should include eight specific items.

1) Owner Name and Location
2) Author
3) File Name
4) Subject
5) Date
6) Record Count
7) Fields per Record
8) Security

1) Owner Name and Location

Owner Name and Location is, of course, a necessity to a potential user since the user must be able to contact the owner, whether for further information regarding the data base, or for actual arrangements for usage.

2) Author

As in the case of published material, it is necessary for an individual to receive credit for his work. However, inclusion of an author in a catalogue of data bases is even more important as it would in many cases provide a potential user with the name of a specific person to contact regarding use of the data base. In instances where the creation of the data base had no single author, this field may indicate a corporate authorship or if unknown, could be left blank in the catalogue.

3) File Name

The file name serves to identify an individual data base. This field is actually the key field in the catalogue as each particular data base entry would have a different file name within any one owner's library.

4) Subject

The subject field allows for a few sentences to describe the data content of the file. It is this field that a user would be interested in and a keyword or subject system should be established here to allow more rapid location of desired information. Even if a keyword system were employed, it would still be advantageous to include a description of the contents, use, significance, or other pertinent information about the data base which would serve to enlighten the potential user as to whether he would want to use the file or not.

5) Date

The date category is broken down into smaller subfields depending on whether the information is of a periodic nature or not. In the case of time-related information there are four sub-fields.

a) Time Span
b) Volume Period
c) Release Date
d) Retention

a) Time Span - This sub-field indicates the exact time period covered by the information within the data base.

b) Volume Period - This section indicates the time interval between new issues or volumes of the file.

c) Release Date - The release date serves to indicate the point in time that a new volume of a series would be made available for use.

d) Retention - Unlike published material, machine readable data bases are normally obsoleted after a certain length of time. It is important to include this in a catalogue to provide the user with a definite date when he can no longer obtain information. Archival policy may override this date.

In the case of material which is not time-related, only two sub-fields are required.

a) Creation Date
b) Retention

a) Creation Date - The creation date is the date the file was originally released in machine-readable form.

b) Retention - (same as above).

6) Record Count

The record count serves to describe the size of the data base, much as the number of pages describes the size of a book. It would certainly help a user in determining the amount of information available to him.

7) Fields per Record

The number of fields within a record describes the size of each record. A field is a unique piece of information within the record, thus, this measure would help the user in determining the complexity or extent of detail of the information within the data base. It may specify that the records contain a variable number of fields, with indicated typical and maximum counts. Also, there are cases wherein this field has no significant meaning.

8) Security

It is anticipated that most, if not all, data bases could and would be catalogued in libraries, and since many data bases would be considered "classified" by their owners, it would be useful to have a security description. This would prevent interested users from many needless inquiries into the availability of classified data and still allow maximum operation of the cataloguing system. Possible catagories could be as simple as "public", "private", and "semi-public". Public files would be open to all interested users, private files being completely closed, and semi-public indicating that some information would be available, leaving it to the user to contact the supplier for more details.

III. Excluded Information

The proposed cataloguing system would exclude the following four items.

1) Medium and Mode Description
2) Field Description
3) Extent of Usage
4) Miscellaneous Dates

Each was given careful consideration before being dropped, and rejected because it was felt that the amount of useful information given was not worth the extra space and time required to include them in the catalogue.

1) Medium and Mode Description

With the large number of storage devices on the market today and the many varied code formats, it is virtually impossible to include all necessary information in a catalogue. Also, if a user was definitely interested in using a data base, he would perhaps need it converted to a form acceptable to his equipment. It is anticipated that most data would need to be converted for each different user. Therefore, it is not really worthwhile to include detailed hardware description or recording formats.

2) Field Description

Keeping in mind the need for a brief catalogue, it is hardly practical to present a detailed field description. Many data bases will have scores of fields, and to include a description or even a name for each would require much more room than a catalogue could allow.

3) Extent of Usage

It is often helpful to know the significance and scope of use of a data base simply for a better understanding of the data base itself. However, it would be difficult to lay down rigid guidelines for description of such nebulous terms as significance and scope. Therefore, any information along this line could be included in the subject area if so-desired.

4) Miscellaneous Dates

Many key dates were considered before finally arriving at those included in the catalogue. Dates ommitted included the following:

a) Time of data collection
b) Time of assembling data into machine-readable format
c) Date of output into current logical format
d) Date of output into current physical format
e) Dates of supplementary files

The preceding dates are informative but fairly useless in providing information concerning the contents of the data base.

IV. Conclusion

Design of the cataloguing system was influenced by the need for maximum ease of usage and data description relevant to the needs of most users. Thus the most important goal of a cataloguing system should be to provide information on a maximum number of data sources, and provide the potential user with an adequate description so that he may then decide if he is interested in using a particular data base. This approach also reduces overhead involved in more complex cataloguing systems and allows for uniform cataloguing of virtually all types of data bases. Finer details of the data file and its availability would then be left for the supplier and user to arrange themselves.

Examples of the proposed system follow in section V.

V. Examples

Owner: Office of Fiscal Affairs, Oregon State University, Corvallis, Ore.
Author: O.S.U. Computer Center
Name: UPDATE II
Subject: Personnel data for non-classified employees of Oregon State University, including information on appointments, degrees, FTE, Tenure, and other related information.
Date: Time Span = 1966 to date
Volume Period = 1 academic year
Release Date = July 1
Retention = 10 years
Record Count: 3200 records
Fields/Record: 48
Security: Semi-public

Owner: Library, Oregon State University, Corvallis, Ore.
Authors: Jennings, Michael A.; Spigai, Frances; Mahan, Thomas
Name: LOLITA
Subject: Book ordering and fund accounting system handling purchases, receipts, gifts, etc. and all necessary accounting and report printing.
Date: Creation Date = March 1970 (daily changes)
Retention = Indefinite
Record Count: 6000
Fields/Record: approx. 31
Security: Semi-public

VI. Glossary

Character - Primary element of information storage. For instance, letters of the alphabet, numbers, and special characters. An important means of determining data storage or transmission capacity.

Data Base - Reservoir of data, collection of facts which may or may not be related structurally, but must be available to the facilities of the system where used. Thus, a machine-readable data base must be directly available to the computer or data-processing equipment.

Field - A grouping of one or more characters which is treated as a whole for the purpose of representing a particular catagory of data within a record.

File - A collection of related records treated as a unit. Any collection of informational items similar in purpose, form, and content, and structurally related.

Medium - The equipment or material within which data is stored, e.g., tape, cards, disk, etc.

Mode - The method of structuring or coding information on a file, e.g., binary, BCD, ASCII, etc.

Record - A group of one or more consecutive fields on a related subject, a group of related facts or fields of information treated as a unit; a unique data entry.

Record Count - The number of records within a file, number of unique data entries.

Release Date - Date on which a dated file or data-base is released or made available for general use, following the closing of one volume time period.

Retention - Time period a data file is held before being disposed of or made available for change.

Security - Refers to the availability of data for general use. Having to do with the classification of information...private, semi-public, public.

Bibliography

Lawrence, Ben; Minter, John: Management Information Systems: Their Development and Use in the Administration of Higher Education, Boulder, Colo., Western Interstate Commission for Higher Education, 1969, pp. 10-11.

Spigai, Frances G.; Mahan, Thomas: "On-Line Aquisitions by LOLITA," Journal of Library Automation, 3 (Dec. 1970), 276-94.