NETWORK MANAGEMENT OF
MACHINE-READABLE DATA FILE RECORDS

BY

Marilyn Nasatir
Dublin, Ohio

Users of machine-readable data files (MRDF) bear the responsibility to share
information about data files as well as to share access to data files. Anyone needing
a particular data file would find it useful to be able to sit down at a terminal and
call up a record of that file by author or title or a variety of control numbers,
find out where that data file is and at the same time find out how to access the
file, wherever it happens to be. This presumes that someone has entered that record
into the database in order to be able to retrieve it.

The bibliographic networks such as OCLC (Online Computer Library Center), RLIN
(Research Libraries Information Network), WLN (Washington Library Network), UTLAS
(University of Toronto Library Automated Systems) and NOTIS (Northwestern Online
Total Information System) provide a mechanism for participants to share information
resources while the networks take care of storing, maintaining and managing machine-
readable records of them. At present, these records are bibliographic descriptions
of books, serials, audiovisual media, maps, manuscripts, music scores, and sound
recordings, making each item not only uniquely identifiable but also locatable for
loan or purchase by means of institutional holdings symbols. At least one biblio-
graphic network--OCLC--will be incorporating records of MRDF into its bibliographic
database to facilitate utilization of machine-readable data.

The two primary considerations in this important system enhancement are 1) the
applicability of existing standards for describing and accessing MRDF records and
2) the feasibility of integrating MRDF records with records of the traditional infor-
mation sources which comprise a bibliographic network database.

STANDARDS FOR DESCRIBING AND ACCESSING MRDF

The need for standards to enable libraries, archives, and other information
based agencies to cooperate in the description and access of machine-readable data
files already being collected by academic and other institutions was demonstrated
over a decade ago when the Descriptive Cataloging Committee of ALA's Cataloging and
Classification Section established the Subcommittee on Rules for MRDF. Over a period
of years the Subcommittee drafted position papers making recommendations on cataloging
MRDF which became the basis of chapter nine ("Machine-Readable Data Files") of the
second edition of Anglo-American Cataloging Rules (AACR2).[1] AACR2 attempts to inte-
grate cataloging of MRDF with all other types of materials.

The Subcommittee's recommendations supplemented by recommendations made by the
Joint Steering Committee of AACR2 were followed in 1978 by the National Conference
on Cataloging and Information Services for MRDF, organized to develop standards for

bibliographic control of MRDF. Two of the recommendations of the Conference were 1) the design of a MARC (Machine-Readable Cataloging) format for MRDF and 2) the definition of products and services to be derived from cataloging MRDF.[2]

In response to the first Conference recommendation, the Network Development Office of the Library of Congress (LC) produced Machine-Readable Data Files: A MARC Format (MRDF/MARC).[3] The US/MARC Formats are standards for representing biblio-graphic and authority information in machine-readable form. An individual MARC format is a set of conventions for encoding a particular type of machine-readable record. The structure of a US/MARC record is determined by the American National Standard for Information Interchange on Magnetic Tape (ANSI Z39.2 - 1979) and by Documentation - Format for Bibliographic Information Interchange on Magnetic Tape (ISO 2709 - 1973). The content of certain data elements in a MARC record is specified by the US/MARC Formats, while the content of traditional catalog elements is defined by other guidelines such as AACR2.[4]

In response to the second recommendation, the Inter-University Consortium for Political and Social Research (ICPSR) at the University of Michigan has created an automated cataloging system for MRDF from which such products as the guide to ICPSR's resources, facsimilies of catalog cards, bibliographic citations, indexes, and authority lists of authors and titles will be derived.[5] The system is still in the development stage and is not yet available to the public.

Other agencies have also derived products and services from cataloging MRDF. At the federal level, standards were developed for bibliographic citations and abstracts in the production of directories of MRDF. Two examples of applications of these standards are the Directory of Data Files issued by the Bureau of the Census and American National Standard for Computer Program Abstracts (ANSI X3.88 - 1981).

The existence of stanards thus demonstrated, the question of applicability must be addressed. In the intervening years since the adoption of AACR2, microcomputer and video game software have emerged which were not considered either in AACR2 or in MRDF/MARC. It seems more appropriate to extend the definition of MRDF to include microcomputer software than to violate input standards of the bibliographic networks by forcing these materials indiscriminately (and unretrievably) into the network formats for books or audiovisual media which were not constructed to handle them. Therefore, since input standards are rule-driven, catalogers and policy makers have been meeting to propose solutions to the problem of rules for physical description of microcomputer software. A special cataloging task force has been formed with the charge to develop a statement regarding microcomputer software description, to be published in LC's Cataloging Service Bulletin and to constitute national policy until AACR2 can be revised. This policy will be reflected in modifications to MRDF/MARC.

For assistance in the description of MRDF, catalogers and data librarians will have at their disposal Sue Dodd's new Cataloging Machine-Readable Data Files: An Interpretive Manual[6] as well as a manual of microcomputer software cataloging examples which is expected to be published by the Minnesota AACR2 Training Series.

The feasibility of integrating MRDF records into the database of a bibliographic network, which in netowrk terminology is called implementing a MRDF format, is dependent upon 1) the existence of data elements in a bibliographic record which are appropriate to MRDF users, 2) network systems and services already in place which meet the needs of MRDF users, 3) products which can be generated from these systems and services, and 4) a viable mechanism for extending treatment of traditional information sources to new formats.

Data Elements

Data elements present in a network bibliographic record needed by MRDF users are as follows: a data file description showing existence and source of data; a detailed abstract which includes the genesis and history of the file so as to link modified files; a keyword structure; links between data files and the software created to manage or operate them, including the presence and bibliographic citation of accompanying documentation, software compatibility, and linkage with other files or programs; and the applicability of the data to solving specific problems or analytic needs.

Dodd has grouped these data elements for MRDF into six categories, or levels. Level 1, Bibliographic Identity, includes such bibliographic elements as author, title, edition, distributor, notes, etc. Level 2, Data Abstract, is a descriptive summary, abstract, or subject analysis of the contents of the file. Level 3, Classification, provides the classification codes, indexing, or descriptors. Level 4, Technical Information for Access, spells out the physical characteristics needed to access the MRDF such as recording density or software compatibility. Level 5, Analysis or Use, gives a citation of documentation and related reports as well as such data collection information as how/when the data were collected, the unit of analysis, and sampling procedures. Finally, Level 6, Archiving or Maintaining, indicates records on processing, storage, use, and modifications to the MRDF.[7] While this may sound like more information than could be incorporated into a single record, a bibliographic record currently provided by OCLC, and presumably by the other networks, can accommodate over 4000 characters, considerably in excess of the approximately 2500 required, according to Dodd, for an expanded bibliographic MRDF record.

Systems and Services

The data elements just discussed, along with numerous others, form the basis of the systems and services already offered in different combinations by the bibliographic networks for managing, with appropriate variations, records of books, serials, audiovisual media, maps, manuscripts, music scores and sound recordings. Prominent among these systems and services are online union catalogs; shared cataloging; serials control; subject retrieval; patron access; interlibrary loan (ILL); authority control, services for converting existing bibliographic and location data into machine-readable form and adding them to the online union catalog; and acquisitions, ordering, and fund accounting which currently interface with a directory of libraries, publishers and vendors, soon to interface directly with the vendors themselves, and easily extendable to producers and distributors of MRDF. This powerful combination of

7

services will provide MRDF users with a systematic way of identifying and locating MRDF for in-house control as well as for acquiring and accessing MRDF from external sources.

Additional services not to be overlooked are the training programs, workshops, and materials provided both by the instructional coordinators of the bibliographic networks themselves and by the independent contracting network offices through which most institutions participate. It is interesting to note that the independent net-works that contract with OCLC to provide services to their member libraries are already getting requests for the MRDF format from a variety of sources. Government and research institutions are of course spurred on by the 1980 Census. Data archives are underutilized and eager to add their holdings to the OCLC database. "Silicon Valley" type organizations have a wide range of MRDF to control. Academic and public libraries, schools, and media centers, stimulated by the microcomputer craze, are asking to use OCLC for cataloging, ILL (especially for documentation), acquisitions, and collection development in the lively area of microcomputer software, games, and instructional materials as well as MRDF. In fact, DIALOG (available through OCLC's Affiliated Online Services) is adding the INTERNATIONAL SOFTWARE DIRECTORY and the MICROCOMPUTER INDEX to its long list of databases.

Products

A MRDF record unambiguously tagged in an online union catalog can serve as an organic record from which catalog cards, accessions lists, and institution-specific archive tapes can be generated just as they are for other types of materials. The potential exists for such additional products as union lists, bibliographic citations, catalog entries, data abstracts, card or tape distribution of MRDF records, and subject access either printed or on microform.[8]

Implementation of a MRDF Format

Although all of the bibliographic networks have the capability of implementing new formats in response to and support of LC policy and precedent, and despite Mary Magrega's paper describing the areas of activity which would be involved at UTLAS in implementing a MRDF format[9], OCLC is the only bibliographic network which has actually made a commitment to implement a MRDF format and is already involved in the analysis phase.

Analysis is the first of five stages of an OCLC Cataloging Maintenance Project, the developmental process decided upon as the most viable mechanism for expediting the MRDF implementation for two reasons. One is staff and resource allocation. The other is that MRDF/MARC will be distributed by LC as an Update to its MARC Formats for Bibliographic Data (MFBD) on which many Cataloging Maintenance changes are based.[10] The product of the analysis phase will be functional specifications written by a library systems analyst serving as Product Manager of a team of highly skilled catalogers, programmers, technical writers, quality control and testing staff.

The functional specifications will provide the basis for the four remaining stages of the Cataloging Maintenance process: development of software, preparation of user documentation, creation of scenarios for testing the changes, and finally installation of online system and related offline production changes. One of the

benefits of installing the MRDF format as a Cataloging Maintenance Project is that
subsequent modifications distributed by LC through MRBD Updates will be easily
accomodated.

CONCLUSION

A major goal of integrating MRDF into traditional library and information ser-
vices by incorporating MRDF records into a bibliographic network is to communicate
the availability of machine-readable materials and facilitate their utilization. It
has been shown that existing standards for the description and access of MRDF, with
minor modifications already in progress, are adequate for the bibliographic control
of a wide range of machine-readable materials currently being collected. It has also
been shown that data element descriptions, bibliographic network systems, services,
and products, and a network mechanism for implementing new formats all are firmly
established and easily applicable to MRDF ensuring the feasibility of the management
of MRDF records by the bibliographic networks. What remains to be demonstrated is
the MRDF users will avail themselves of this developing capability by contributing
and accessing MRDF records by means of the various systems, services and products
discussed. The vast numbers of inquiries already received suggest that they will.
It is anticipated that this will be a cooperative venture in which the time and money
spent by every user will be compensated for by benefits far outweighing the costs.

REFERENCES

1. American Library Association. Anglo-American Cataloguing Rules, 2d ed. Chicago:
   ALA, 1978.

2. Sue A. Dodd. "Toward Integration of Catalog Records on Social Science Machine-
   Readable Data Files Into Existing Bibliographic Utilities: A Commentary."
   Library Trends 30, n.3 (Winter 1982): 342.

3. Sue A. Dodd. Working Manual for Cataloging Machine-Readable Data Files (Chapel
   Hill: Social Science Data Library, University of North Carolina, 1976, mimeo).

4. "The US/MARC Formats: Underlying Principles." Library of Congress Information
   Bulletin (in press).

5. Carolyn Geda of ICPSR, private communication.

6. Sue A. Dodd. Cataloging Machine Readable Data Files: An Interpretive Manual.

7. Dodd, "Toward Integration," pp. 350-351.

8. Marilyn Nasatir. "Machine-Readable Data Files and Networks." Information
   Technology and Libraries (in press).

9. Mary K. Magrega. "A Bibliographic Utility's Implementation of the MARC Format for Machine-Readable Data Files." (Toronto: University of Toronto Library Automated Systems, 1981, mimeo).

10. United States. Library of Congress. Automated Systems Office. <u>MARC Formats for Bibliographic Data</u>. Washington: Library of Congress, 1980.

ICDBHSS/83

INTERNATIONAL CONFERENCE ON DATA BASES
IN THE HUMANITIES AND SOCIAL SCIENCES

The ICDBHSS/83 will be held in Rutgers, the State University, on the Douglas College campus, at New Brunswick, New Jersey, from June 10 to 12. It is intended that the conference be broad in scope and that there be a true exchange of ideas and information, not only between the social sciences and the humanities, but also between the various fields in each area.

Speakers from twenty-four countries will present papers concerning data bases, their structure, accessibility, and uses in the fields of archaeology, art, artificial intelligence, criminology, education, history, law, lexicology, linguistics, literature, music, philosophy, sociology, and religion, to mention only the broader areas. Most of the fields just mentioned extend over to other fields, e.g., art-history, socio-history, music-library, literary-linguistics.

The language of the conference will be English.

For additional information contact: Dr. Robert Allen
                                      Bishop House, Room 306A
                                      Rutgers University
                                      New Brunswick, N.J. 08903
                                      (201) 932-7335 or (201) 932-7505

INDEX OF PUBLIC OPINION POLL QUESTIONS COMPILED

Dennis Gilbert at the University of Louisville is compiling the AMERICAN PUBLIC OPINION INDEX, which will provide a topical listing of all questions asked in American public opinion polls, some information about where to get the results of the survey, and some information about the survey in which the questions were asked. The first volume will include questions asked in 1981. The second volume will cover 1982. By going beyond abstracts of polls to question level indexing, Gilbert hopes to provide researchers with the kind of detail they need to directly access the particular topic in which they are interested. Like any other index, APOI will not give researchers substantive information, but instead will tell where to find it. Consequently, polling agencies will retain control over their findings, especially those not yet published.

For further information, write to Dennis Gilbert, 501 Highwood Dr., Louisville, KY 40206 or call (502) 893-2527.