This is a preprint of a paper intended for publication in a journal or proceedings. Since changes may be made before publication, this preprint is made available with the understanding that it will not be cited or reproduced without the permission of the author.

UCRL - 74259
PREPRINT

CONF-730406--1

LAWRENCE LIVERMORE LABORATORY
University of California/Livermore, California

**UTILIZATION OF SCIENTIFIC DATA BASES AT LLL USING THE
"MASTER CONTROL" COMPUTER PROGRAM**

Viktor E. Hampel

John Wade

October 1972

This paper was prepared for presentation at the
AIAA Computer Network Conference
Huntsville, Alabama
April 16-18, 1973

NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Atomic Energy Commission, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

This redefinition of the program, during problem execution, may be employed indefinitely, thereby permitting progressive adaptation of the program and of data bases to contemporary needs. (Figure 1)

The command language is telegram-like in free form common-day English. An attempt was made not to include peculiar symbols and abbreviations, and to decrease the rigor of syntax wherever possible.

Data used in MASTER CONTROL may be text, Hollerith symbolic information, numeric, or intermixed. Data is defined by characteristic labels or acronyms, by which it is recognized, stored, retrieved, manipulated and displayed.

All 32 operations now available through MASTER CONTROL, Version-3, can be initiated either in batch processing or from teletype in the interactive mode. It is also possible to monitor the progress of batch runs from teletype.

Version-3 of the program is directly accessible on five CDC-6600/7600 computers as public file MCP, Master Control Program, and operates as a controllee in the time-sharing mode. Data bases are stored either in the IBM-1360 Photo-Digital storage system or on magnetic tape. The LLL General File Transport System is used for the transfer of files among storage media, and to and from the working computers.

B. Utilization of Bibliographic Data Bases

The reference group of the Technical Information Department at LLL routinely conducts searches of commercial, national and proprietary bibliographic data bases to alert the professional and administrative staff of the Laboratory about publications in special fields of programmatic interest.⁴ By employing one computer program for this task, the same query can be addressed to data bases of overlapping disciplines.

MASTER

UTILIZATION OF SCIENTIFIC DATA BASES AT LLL USING THE "MASTER CONTROL" COMPUTER PROGRAM *

Viktor E. Hampel and John Wade
Lawrence Livermore Laboratory
Livermore, California

October 1972

SUMMARY

MASTER CONTROL^{1,2} is a computer program designed to unify storage, manipulation, reorganization, retrieval, and display of information from dissimilar data bases. It is open-ended, user-oriented, and applicable to administrative and scientific purposes. The main characteristics of MASTER CONTROL are the free-form input of Hollerith and numerical data, packed memory allocation and a hierarchial order using pointers for random access, a telegram-like English command language, and options to interrogate and to manipulate data bases in batch-mode, or from remote consoles in a time-sharing mode.³

PROGRAM DESIGN PHILOSOPHY

A. General Comments

The design philosophy of MASTER CONTROL employs an originally undefined basic program, permitting the user to create his own data acquisition and data manipulation system. This adaptation to individual and changing requirements is accomplished directly, without programmer intervention, by naming and defining new data arrays, by introducing alternate and simultaneously active terms of the command language, by user control over data manipulation, and by user-specified data output formats and data list options.

*Work performed under the auspices of the U.S. Atomic Energy Commission.

Searches of this type match a specified number of terms, connected through Boolean logic, against the descriptions of computerized citations contained in such data elements like key-words, categories, titles and the text of abstracts. Authors, countries and any other data elements may be searched as well. The preparation of the search profile, also referred to as the interest profile, may contain stems and roots of words, whole words, phrases and numeric search criteria. In addition, all terms used may be qualified by suffixes or prefixes of known or unknown length, and may include unknown characters. Approximately 1000 interest profiles are processed at LLL each month against an influx of about 33,000 new bibliographic citations of the world literature, notably Nuclear Science Abstracts, Chemical Titles, Engineering Index and the American Institute of Physics SPIN tapes.

The ELF file transport system⁵ at LLL makes it possible for any computer user to access and to interrogate files in the public domain directly. Here it is recommended that preparation of a continuing interest profile be carried out with the assistance of the information specialist of the TID Reference Group. Their familiarity with various data bases and with different search strategies has proved to be an asset in the preparation of routine interest profiles that return citations with a high degree of relevancy. Established profiles can be committed to tape or disk and activated at the user's discretion, or are used retrospectively by TID for routine current awareness searches. A periodic mailer alerts the user of latest publications in his field of interest.

To facilitate the intercommunication and common interrogation of bibliographic data among originally disparate data bases, a standard interface was prepared for early implementation at LLL containing the smallest common set of data elements from some 16 major national and commercial abstracting services and from proprietary data bases originated at LLL.⁶

C. Creation and Utilization of Textual-Numeric, Scientific Data Bases

When a comprehensive survey of the world literature is needed in a particular field, containing many references, the user may request TID to employ MASTER CONTROL for the creation of a specialized computerized data base by direct transfer of records from different data bases and their progressive accumulation on magnetic tape or in the photo-store. This can be done concurrently with the mailed dissemination of current awareness notices and may include references identified by manual searching of the literature. (Figures 2,3)

Some or all data from data elements of specified or computer-retrieved records can be extracted. A comprehensive bibliography on plutonium and laser phenomena have been created in this manner.

Specialized data bases of this type can be cost-effectively subjected to various operations and extensions. Commentaries and notations are added by the user. The numeric content of reports is merged separately or with its bibliographic citation, and the data from selected measurements can be collated and extracted in user-specified formats, without programmer intervention, and serve as direct input into existing scientific computer programs.

Moreover, for some LLL activities the output of calculations is returned to the MASTER CONTROL data base, thus closing the cycle and establishing feedback where required for progressive evaluation and synthesis of numeric data. In this manner MASTER CONTROL offers the professional user a means to create an expandable scientific, textual and numeric, data base specifically adapted to his requirements.

In recent years, large technological and scientific numeric data bases have become available from commercial and governmental agencies. Some of these are now in use at LLL. (Tables 1 and 2) Here it is found that one of the most

promising areas of data utilization has come about from pattern recognition methods,^{7,8,9} operating in conjunction with data that is stored, filtered and manipulated by MASTER CONTROL. This has proven to be of practical advantage in decision making and solving of real-time mission oriented problems.

BASIC DESIGN PHILOSOPHY OF 'MASTER CONTROL'

Figure 1.

Figure 2.

SPECIALIZED SCIENTIFIC DATA BASES

Figure 3.

TABLE 1

MAJOR OPERATIONAL DATA BASES AT LLL
 "MASTER CONTROL"

October 1972

	<u>Approximate Number of Logical Records</u>
1. Nuclear Science Abstracts, 1968 -	300,000
2. Air Quality Data, California - 1970	171,540*
3. Chemical Titles, 1971 -	140,000
4. Engineering Index, July 1971	130,000
5. Air Quality Data, Basin-2 - 1963-1971	109,000*
6. American Institute of Physics, SPIN-tapes - 1971	90,000
7. Nuclear Safety Information, NSIC, 1971 -	35,000
8. Proprietary Data Bases	32,000
9. Mass Spectroscopy Data, Low Resolution	27,000
10. Proprietary Data Index	5,200
11. Nuclear Constants Bibliography	3,300
12. Plutonium Bibliography	3,100
13. Information Retrieval Bibliography	3,000
14. Technical Information Department Reference Group	2,500
15. Integral Neutron Experiments Bibliography	2,000
16. Reaction Rate Constants, Atmospheric Kinetic Data	2,000
17. Laser Bibliography	1,500
18. USAEC and LLL Patent Cases	800
19. Miscellaneous Personal and Administrative	

TOTAL

~ 950,000

*in the process of being merged

TABLE 2

MAJOR DATA BASES IN PREPARATION AT LLL
"MASTER CONTROL"

October 1972

	<u>Approximate Number of Logical Records</u>
1. Experimental Neutron Cross Section Data	750,000
2. Nuclear Science Abstracts, Subject Headings 1963 -	300,000
3. Chemical Titles, 1967-1970	300,000
4. Evaluated Neutron Cross Section Libraries	90,000
5. Engineering Data - Physical Properties	4,800
6. Radiation Shielding Information Center Data, RSIC	3,000
7. Physical Data of Isotopes	2,000
8. Equation of State Data	
9. Biological Abstracts	
10. National Earthquake Data	
<hr/>	
TOTAL	~ 1,500,000

REFERENCES

- ¹V.E. Hampel and J.A. Wade, *MASTER CONTROL - A Unifying Free-Form Data Storage and Data Retrieval System for Dissimilar Data Bases*, Proc. Am. Soc. Inform. Sci., 6 (1969). (Preprint UCRL-71686 (1968).
- ²A nearly up-to-date qualitative description of current capabilities available through MASTER CONTROL is presented on pages 14-26, and 81-148 of UCRL-73192, *Functions and Activities of the Physioal Data Group at Lawrence Radiation Laboratory, Livermore*, Robert J. Howerton, May 12, 1971.
- ³Livermore Time Sharing System, Chapter - III - 207, Edition-2.
- ⁴Scott J. Buginas and Neil B. Crow, *A Computerized File Management System - A Tool for the Reference Librarian*. (Preprint UCRL-73355) Presented at the Boston meeting of the Special Libraries Association, June 1970. Extract published in the Atomic Energy Commission Technical Information Bulletin No. 28, page 6, September 1972.
- ⁵John T. Engle, *ELF*, UR-202, June 1, 1970 and Octopus Communique No. 376.
- ⁶Dennis R. Elchesen, *A Correlation of Bibliographic Data Elements for Use in a Generalized File Management System*, UCRL-73827, March 24, 1972.
- ⁷C.F. Bender and B.R. Kowalski, *PATTERN RECOGNITION - I: A Powerful Approach to Interpreting Chemical Data*, Journal of the American Chemical Society, Vol. 94, p. 5632 (1972).

REFERENCES (continued)

⁸ B.R. Kowalski and C.F. Bender, *PATTERN RECOGNITION - II: Linear and Nonlinear Methods for Displaying Chemical Data*, UCRL-74027, Submitted for publication in the Journal of the American Chemical Society.

⁹ Chemical and Engineering News, August 28, 1972, pp. 14-15.

NOTICE

"This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Atomic Energy Commission, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately-owned rights."