

LOCKHEED MARTIN ENERGY RESEARCH LIBRARIES



3 4456 0515441 8

cy.2

# OAK RIDGE COMPUTERIZED HIERARCHICAL INFORMATION SYSTEM (ORCHIS) STATUS REPORT — JULY 1973

A. A. Brooks

OAK RIDGE NATIONAL LABORATORY  
CENTRAL RESEARCH LIBRARY  
DOCUMENT COLLECTION

## **LIBRARY LOAN COPY**

DO NOT TRANSFER TO ANOTHER PERSON

If you wish someone else to see this  
document, send in name with document  
and the library will arrange a loan.

UCN-7969  
(3 3-67)



# OAK RIDGE NATIONAL LABORATORY

OPERATED BY UNION CARBIDE CORPORATION • FOR THE U.S. ATOMIC ENERGY COMMISSION

Printed in the United States of America. Available from  
National Technical Information Service  
U.S. Department of Commerce  
5285 Port Royal Road, Springfield, Virginia 22151  
Price: Printed Copy \$5.45; Microfiche \$0.95

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.

ORNL-4929  
UC-32 - Mathematics and Computers

Contract No. W-7405-eng-26

COMPUTER SCIENCES DIVISION

OAK RIDGE COMPUTERIZED HIERARCHICAL  
INFORMATION SYSTEM (ORCHIS)  
STATUS REPORT - JULY 1973

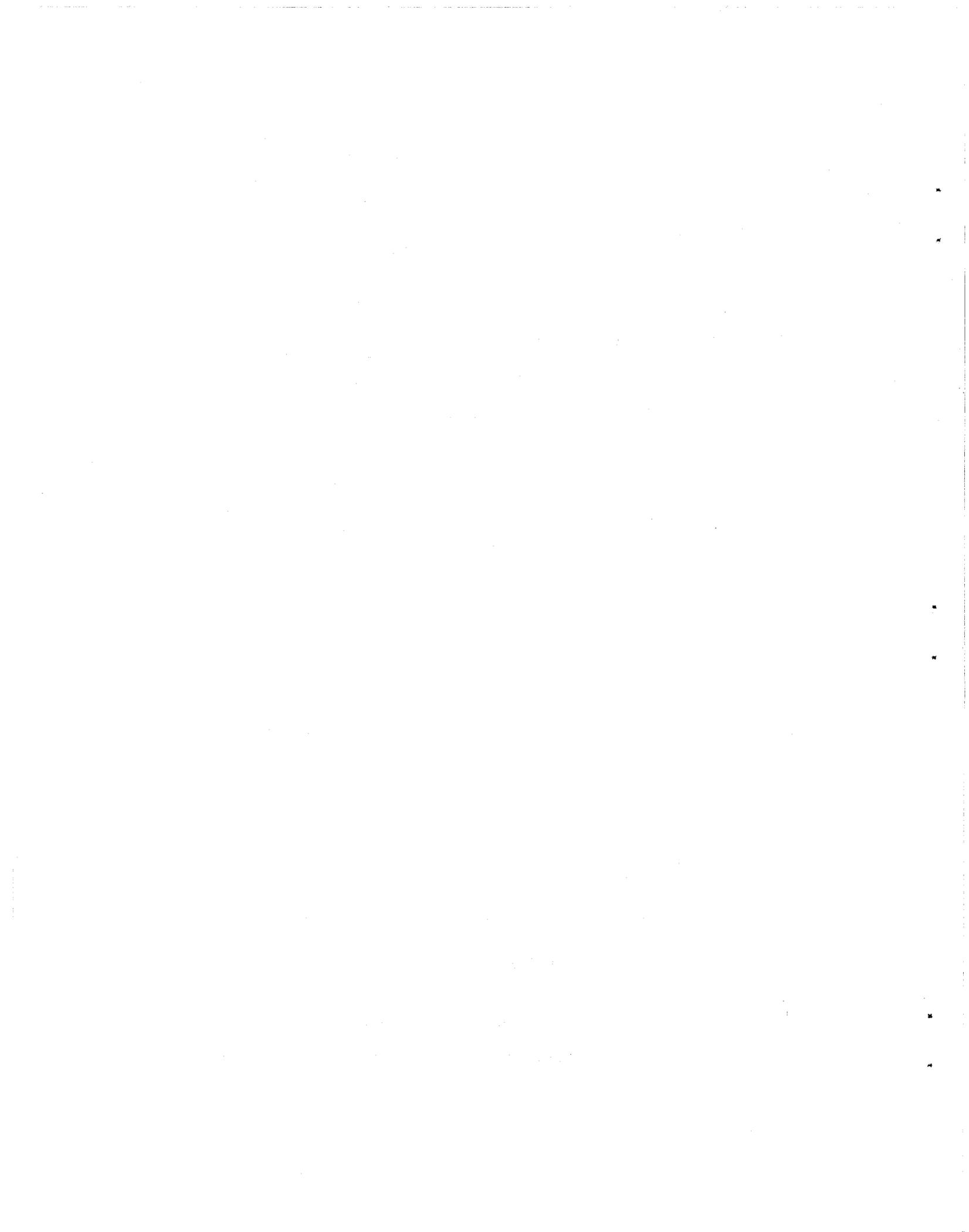
A. A. Brooks

JANUARY 1974

OAK RIDGE NATIONAL LABORATORY  
Oak Ridge, Tennessee 37830  
operated by  
UNION CARBIDE CORPORATION  
for the  
U.S. ATOMIC ENERGY COMMISSION

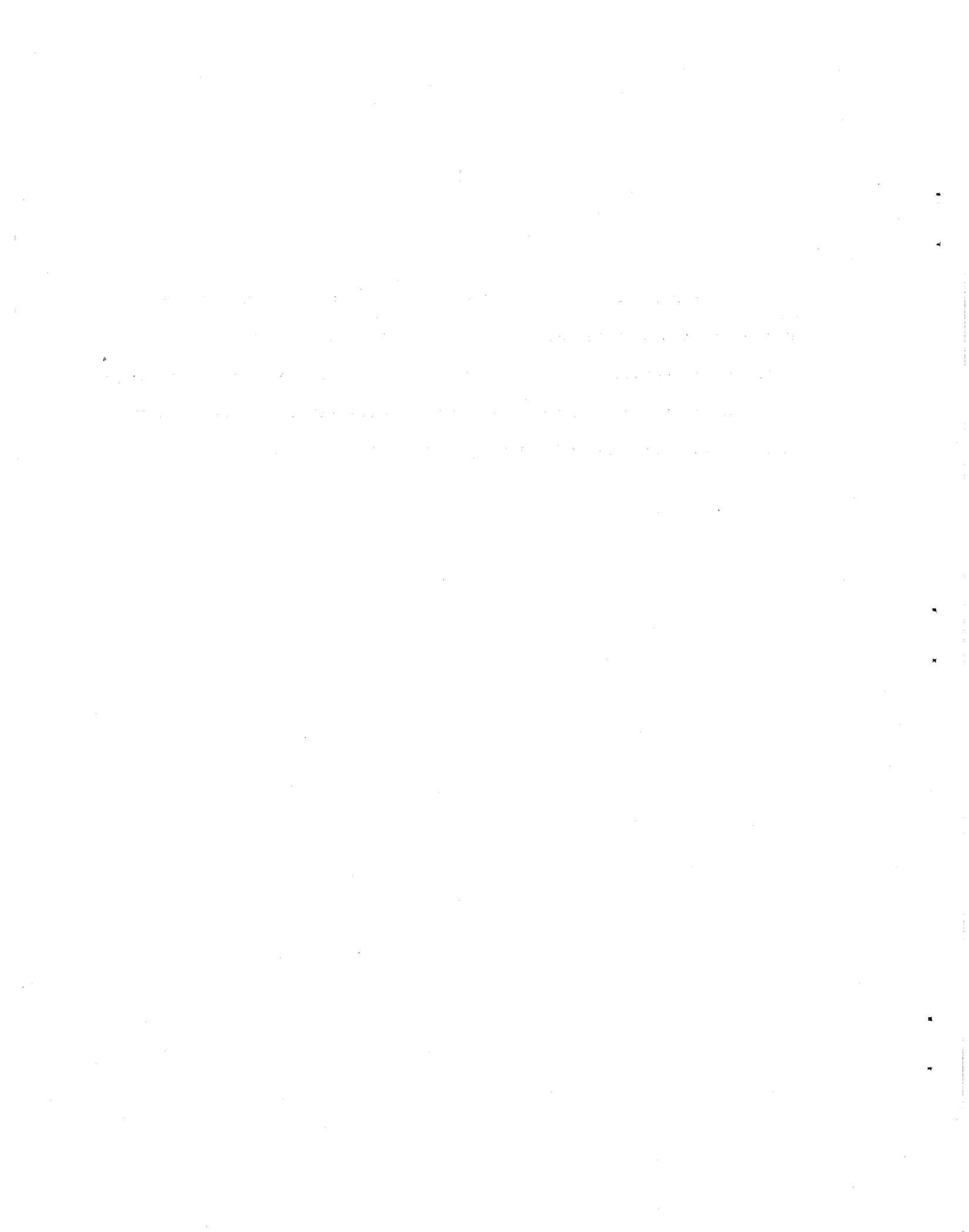


3 4456 0515441 8



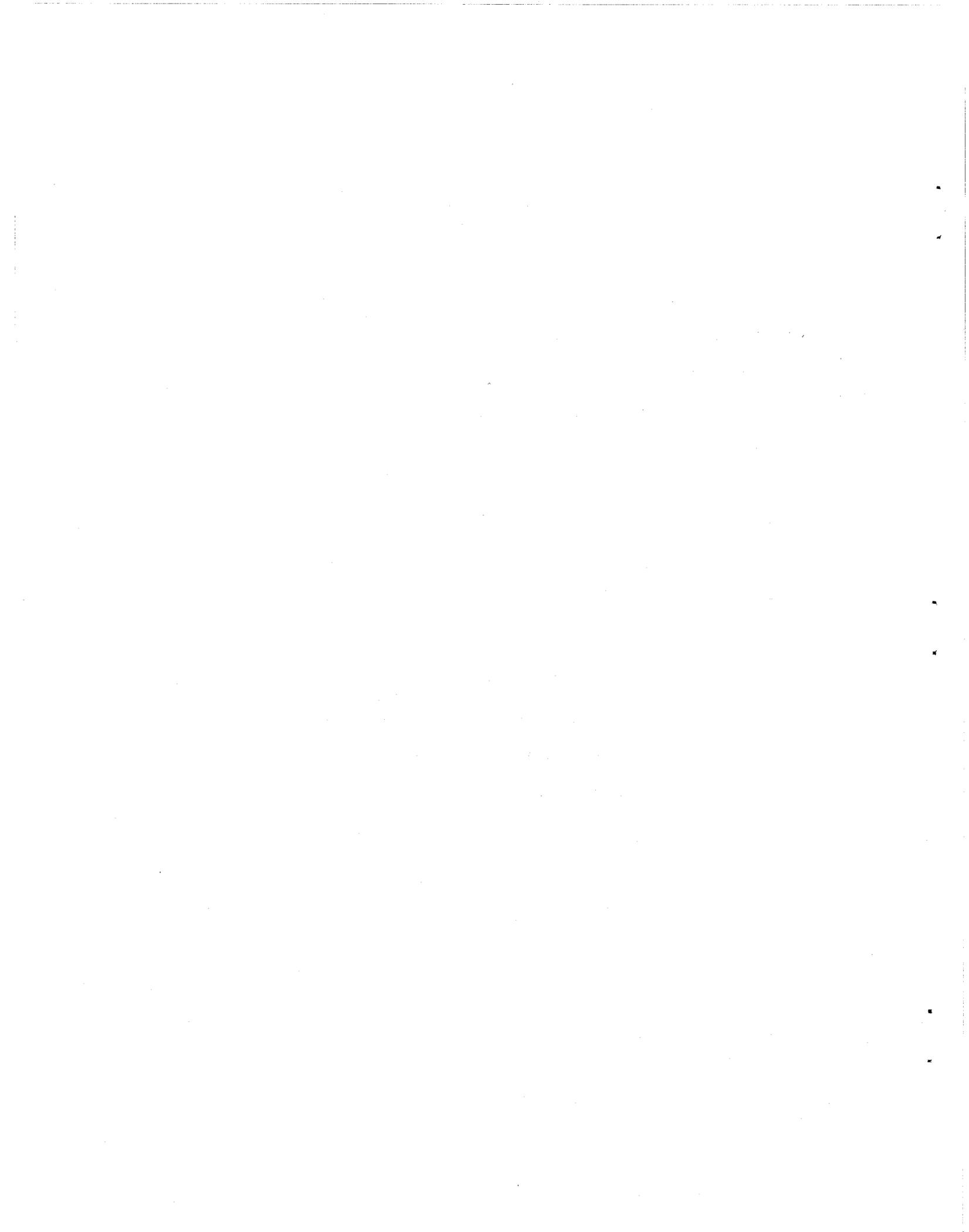
## ACKNOWLEDGMENTS

The author wishes to acknowledge the contributions made by the many persons of the Computer Sciences Division staff to this project. Their names are associated with the numerous reports and programs. Acknowledgments are also due to the several users, especially Dr. G. U. Ulrikson and the Environmental Information Systems Office staff.



## CONTENTS

	PAGE
ABSTRACT . . . . .	1
INTRODUCTION . . . . .	1
OBJECTIVE . . . . .	2
DATA STRUCTURE CONCEPTS . . . . .	2
SOFTWARE CONCEPTS . . . . .	3
LOGICAL CONCEPTS . . . . .	8
SYSTEM STATISTICS . . . . .	13
SYSTEM CONTENTS . . . . .	14
FUTURE OBJECTIVES . . . . .	15
APPENDICES . . . . .	17
I. Program Descriptions . . . . .	19
II. Summary of ORCHIS Data Bases . . . . .	37
III. Annotated Bibliography . . . . .	47
IV. Glossary of Acronyms, Names, and Computing Terms . . . . .	55



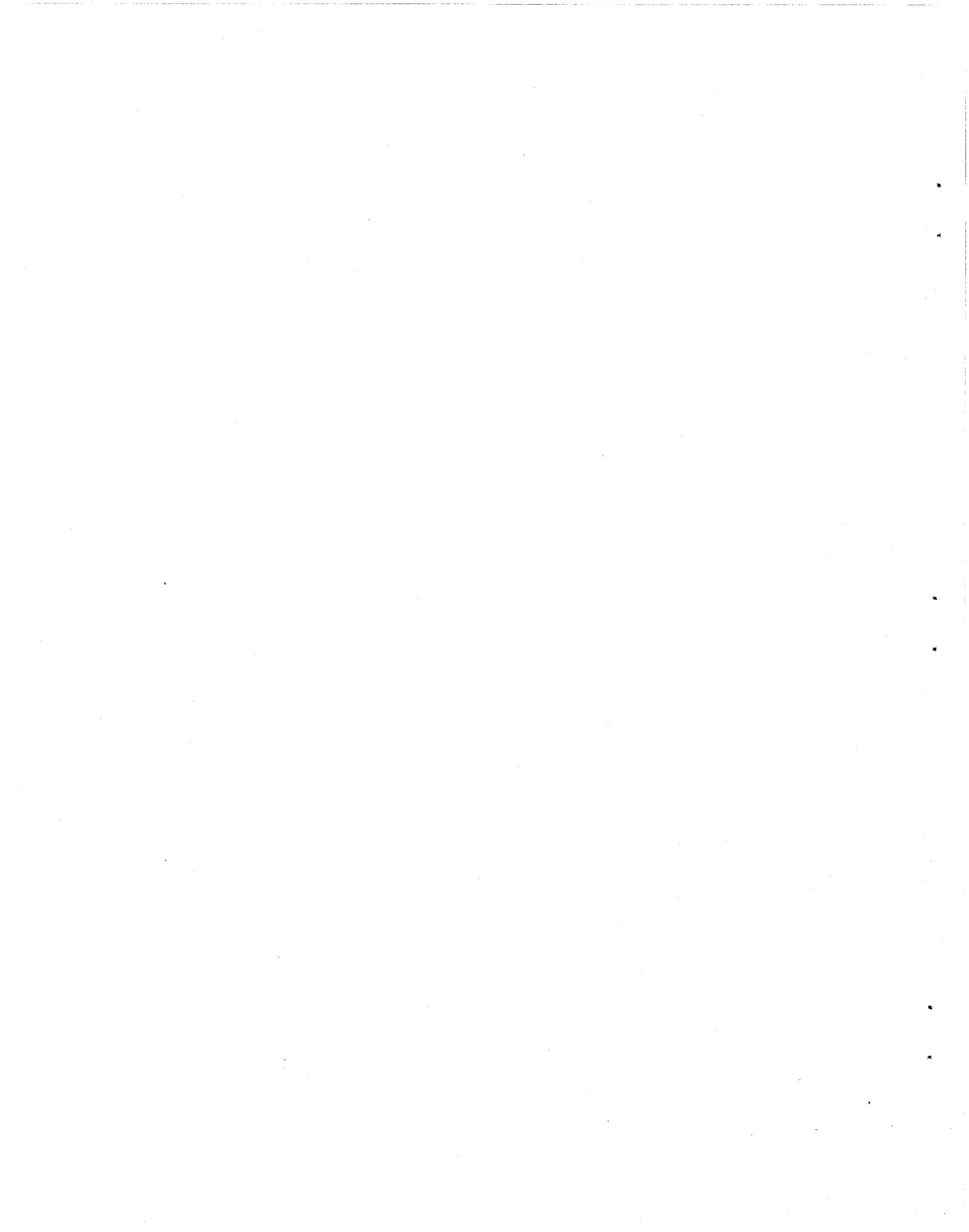
## LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1a. Model for "Information-Systems." . . . . .	4
1b. Model for Computerized Surrogate and Information Collection Files . . . . .	4
2a. Hierarchical Data Structure . . . . .	5
2b. Data Elements Defined . . . . .	5
3. ADSEP Identifier Attribute Conventions for Substructure Mode and Type . . . . .	6
4a. Program Module Model . . . . .	7
4b. A Typical Functional Linkage of Program Modules to Produce Indexed Hard Copy and On-Line Searches . . . . .	7
5. Pictorial of ORCHIS System . . . . .	11



LIST OF TABLES

<u>Table</u>	<u>Page</u>
I. Software Inventory . . . . .	9
II. Summary of ORCHIS Data Bases . . . . .	39



OAK RIDGE COMPUTERIZED HIERARCHICAL  
INFORMATION SYSTEM (ORCHIS)  
STATUS REPORT - JULY 1973

A. A. Brooks

ABSTRACT

This report summarizes the concepts, software, and contents of the Oak Ridge Computerized Hierarchical Information System. This data analysis and text processing system has been developed as an integrated, comprehensive information processing capability to meet the needs of an on-going multidisciplinary research and development organization.

---

INTRODUCTION

The ORCHIS system is a comprehensive, integrated information system serving many AEC/ORNL research and development projects. A brief summary of the ORCHIS concepts, software, data bases, and an annotated bibliography (see Appendices I, II, and III) are presented in this document. A glossary of terms and acronyms is given in Appendix IV. The Computer Sciences Division has integrated the information processing software requirements of several customers to produce the total system. The major funded activities are the ORCON and ORSDI subsystems and the EISO information processing activities. In this report, emphasis is placed on the system aspects of ORCHIS rather than on the contents of the more than 100 varied data bases which are processed by the common software.

## OBJECTIVE

The ORCHIS system is intended to serve the research and development community and its projects with an integrated information processing, retrieval, and analysis system which can deal with both alphanumeric and digital information. Retrospective search, selective dissemination of current information, comprehensive Boolean selection of structured data elements, and statistical analysis are included as well as the capacity to produce hard copy documents having comprehensive working indices and tabular data display. The system obtains its broad capabilities for handling a wide variety of information and providing a variety of services by adopting data structure concepts and software concepts which are sufficiently broad and open-ended that the ORCHIS system can be expanded as new demands are placed upon it. It is essential to the integrity of such a system that such concepts be well defined and faithfully adhered to during the implementation of the system so that compatibility of the software subsystems can be obtained.

## DATA STRUCTURE CONCEPTS

The ORCHIS system reaches its objective of storing and accessing a large number of varied information collections by adopting a hierarchical model sufficiently general to meet the needs of the many users. It reaches its objective of storing and processing a wide variety of types of information by adopting a flexible hierarchical structure into which a diversity of data elements can be stored. The information base

model (see Fig. 1a) is a hierarchical set of global indices, surrogate files, information files and sets of experimental data. The file model (see Fig. 1b) is that of an extended population of unique individuals with whom the information is permanently associated. The data structure at the record level (see Fig. 2a) is a hierarchical structure storing data elements (see Fig. 2b) of arbitrary definition at the leaf nodes. This hierarchical structure contains its own system of pointers which permits a dynamic variation in the size of the sets stored and in the completeness of the information. The generic structure and data description (see Fig. 3) of a particular data set are recorded in the first record of each data set. The data structure used is reducible to a simple case for high-volume use of simple structures. Bulk information may be described within the system and stored for occasional retrieval in less structured and less automated forms.

#### SOFTWARE CONCEPTS

The ORCHIS system reaches its objective of providing a wide variety of data processing services by providing a collection of software subsystems which can be linked together in varied sequences and combinations to provide the necessary processing functions to meet the needs of a wide variety of users. The program model is shown in Fig. 4a which depicts the input and output data sets. Since the input and output master sets are identical in format, any number of programs (i.e., functions) can be linked together (Fig. 4b) to reach a needed objective.

Global Index				
Global Surrogate File				
System 1		2	...	n
System 1 Index				
System 1 Surrogate File				
Information Base 1	2			
Index File				
Surrogate File				
Information Collection				

Typical Information Collections:

- 1) Libraries, report files
- 2) Microfiche collection
- 3) Machine readable research data-nonstandard formats
- 4) Machine readable research data-standard formats
- 5) Resource information; projects, vitae, etc.

Fig. 1a. Model for "Information-System."

External Concept

Population

Individual

Structured Information

Computer Concept

Data Set

Record

Hierarchical Data Structure

Fig. 1b. Model for Computerized Surrogate and Information Collection Files.



C = word or byte count of subset.

$P_i$  = pointer relative to its own location.

$C_i$  = data field count or nested subset count.

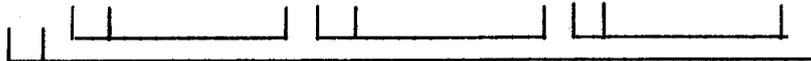
The fully nested, ordered, contiguous data structure with pointers relative to their own location is self-describing with respect to form and extent. Subsets and data elements are associated with alphanumeric identifiers by means of a table containing hierarchical position and data element type information as well as input/output controls. Thus, defined data elements are manipulated by subroutines transparent to the user.

Fig. 2a. Hierarchical Data Structure.

1. Item - (machine word)

2. List -

3. Correlated multilists



4. Delimited text fields - three levels of substructure

5. FORTRAN arrays for regular numerical arrays

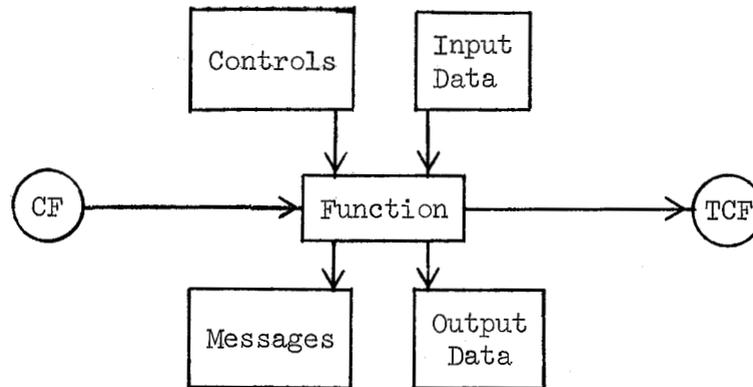
6. Byte-oriented structures for high-volume text data bases

Fig. 2b. Data Elements Defined.

Substructure (Data Element)				Word	Data List	Multi-list	Array	Sub-set
Type				01	02	03	04	09
Mode								(4)
External	Internal							
Formatted	(1)		00	X	X	X	X	X
Free Form (2)	Integer	I*4	10	X	X	X	X	X
	Real	R*4	20	X	X	X	X	X
	Character	R*4	30	X	X	X(3)	X(3)	X
	Text	R*4	40	-	X	-	-	X
	Logical	L*4	50	X	X	X	X	X
	Bit	I*4	60	X	X	X	X	X
	Dictionary	I*4	90	X	X	X	X	X

- (1) External fields and internal mode are determined by the FORTRAN format supplied and may consist of mixed types and modes.
- (2) The external free form fields are described elsewhere.
- (3) Restricted to four or fewer characters per word.
- (4) Subset may contain information of any mode in any substructure.

Fig. 3. ADSEP Identifier Attribute Conventions for Substructure Mode and Type.



CF = computerized file in system format.

TCF = transformed computerized file in system format. Transformation may mean update, conditional selection, reordering, restructuring, or other transformation.

N.B. One or more input or output streams may be nonexistent.

Fig. 4a. Program Module Model.

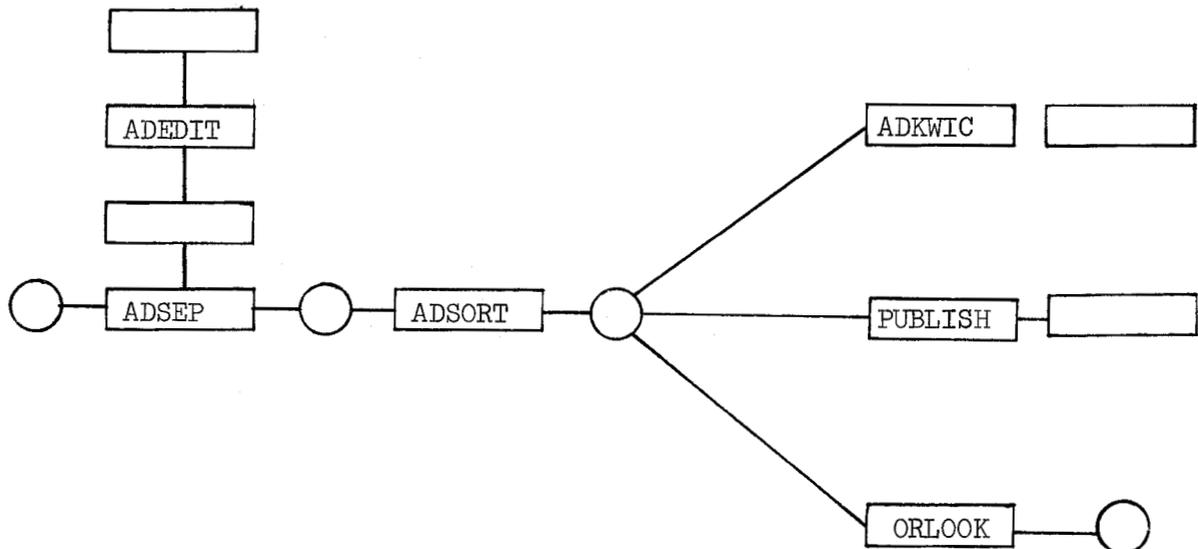


Fig. 4b. A Typical Functional Linkage of Program Modules to Produce Indexed Hard Copy and On-Line Searches.

The record structures are defined by tables or by declaration. The programs are table-driven and deal with arbitrary data elements designated by alphanumeric identifiers. The data elements can be sets of defined structures appropriate to the specific data processing problems. These data structures are manipulated by subroutines transparent to users. The system consists of both batch and on-line programs (see Table I and Appendix I) appropriate to their use and the ORNL environment.

Special access techniques applicable to research problems and substance oriented research are included, such as access to subsets of specified subgraphs of data sets described by directed graphs and access through CAS Registry number, chemical structure, and substructures.

The hierarchical concepts are extended to the format of the reports and tables which can be generated. Arbitrary simple field, permuted words by field, and Cartesian product indices are provided to give superior working indices of the reports produced.

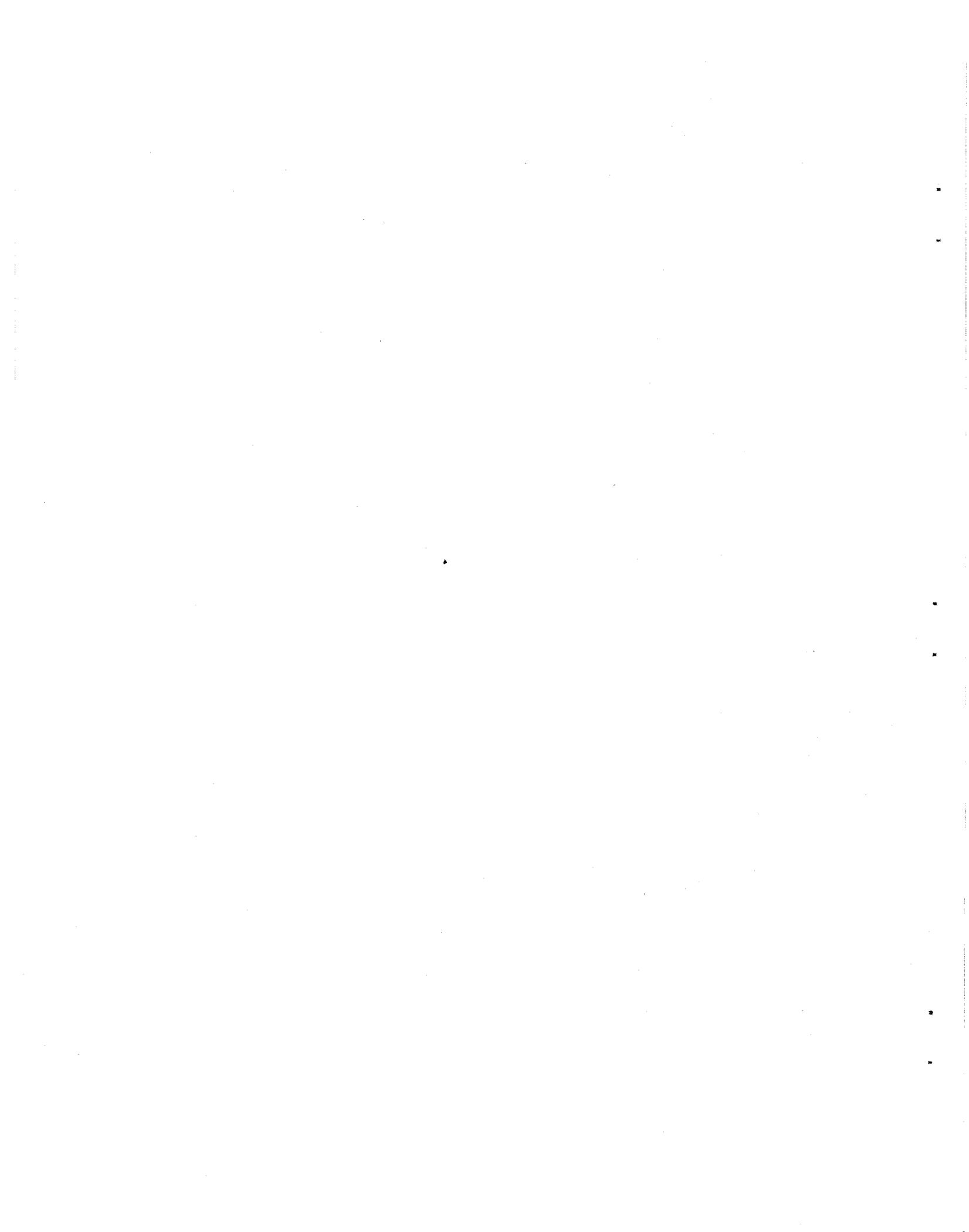
Figure 5 is a pictorial display of the ORCHIS software. The central oval depicts the several data bases. The surrounding annular ring shows the several programs grouped by function. The double-ended arrows represent the ability of the program to produce a modified or augmented data base to enable program linking. The two outside rings describe the functions of each program or group of programs.

#### LOGICAL CONCEPTS

A future objective of the ORCHIS system is to base all of its logical operations on a single, well-defined system of symbolic logic at the

Program or subsystem	Function	Input		Output			Reference
		Master	Data	Master	Data	Copy	Appendix III
AEDIT	Pre-edits text input to ADSEP		X		X		2.9
ADSEP	Data set input and management	X	X	X		proof	2.0
ADSORT	Data set sort keys and reordering	X		X			10.0
ADQIND	Produces Cartesian product indices	X				reproduction	
DUT	Taped deck processor for bulk data		X		X		3.2
CONTAD	Conversion to ADSEP input		X		X	proof	
ADKEYS	Prepared compound compressed keys	X			X		
PUBLISH	Paginates structured report	X				reproduction	10.0
ADKWIC	Simple and permited indices	X				reproduction	4.0
ADTABLE	Tabular displays of text fields	X				reproduction	
ADLABEL	Mailing services, labels, and cards	X				mailing	
MASINDX	Produces global indices	G		G	X	reproduction	
ADLIST	Conditional service listings and copies	X		X		proof	
ADINDX	Service indices and edits	X			X	proof	
CONFAD	Conversion from ADSEP to other systems	X			X	proof	
ORLOOK	String search selection (on-line)	X		X		display, printed	6.0
ORSBI	Current selective dissemination	I	B	X		printed	9.0
ORCON	Retrospective indexed search (on-line)	I	B	X		display, printed	5.0
SADS	Boolean subsetting and analysis	X		X	X	summary	3.0
TRACE	Directed subgraph selection	X		X		summary	7.0
CHEMACES	Chemical name, substructure access		X		X		11.0
ADLIN	ADSEP to LYNOTRON input	X			X	proof	
ADCLEAN	Cleanup and editing	X			X	proof	

Table I. Software Inventory.



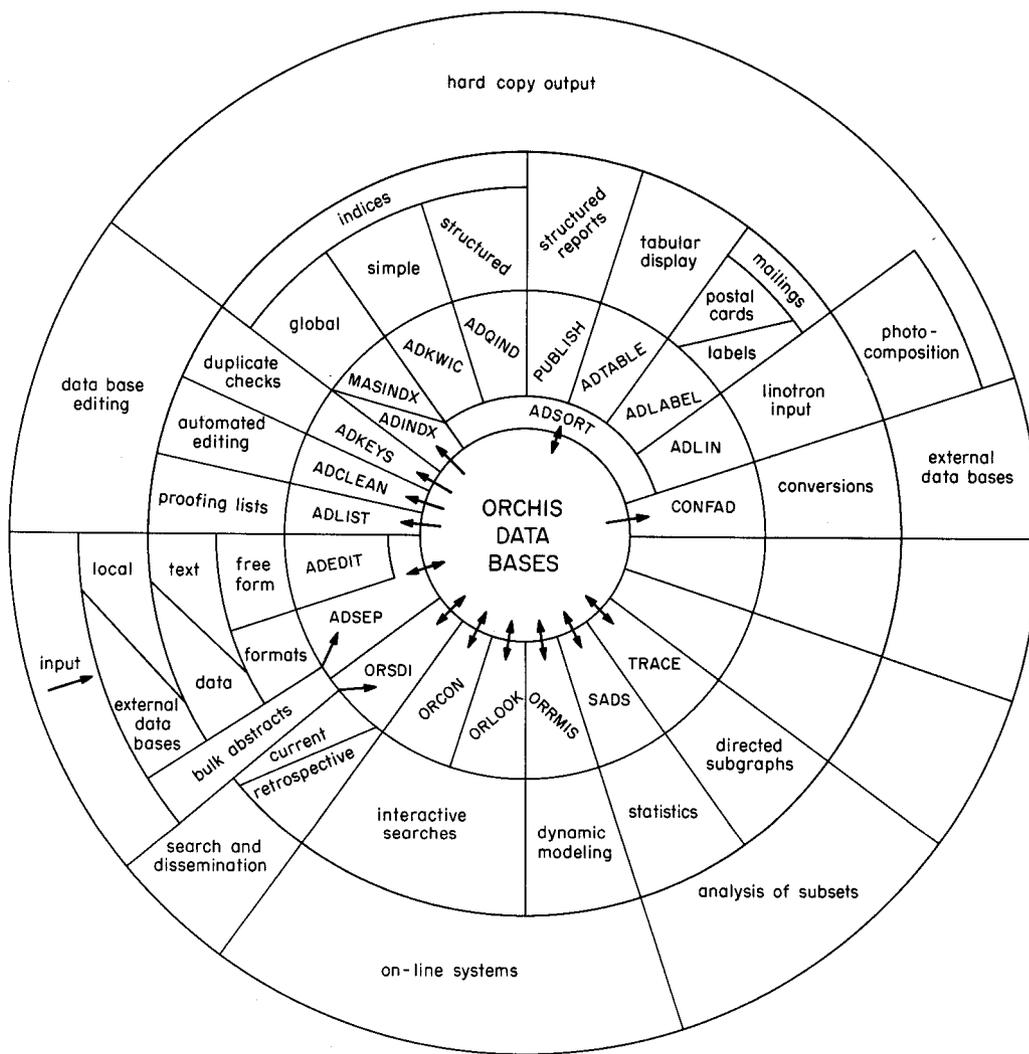
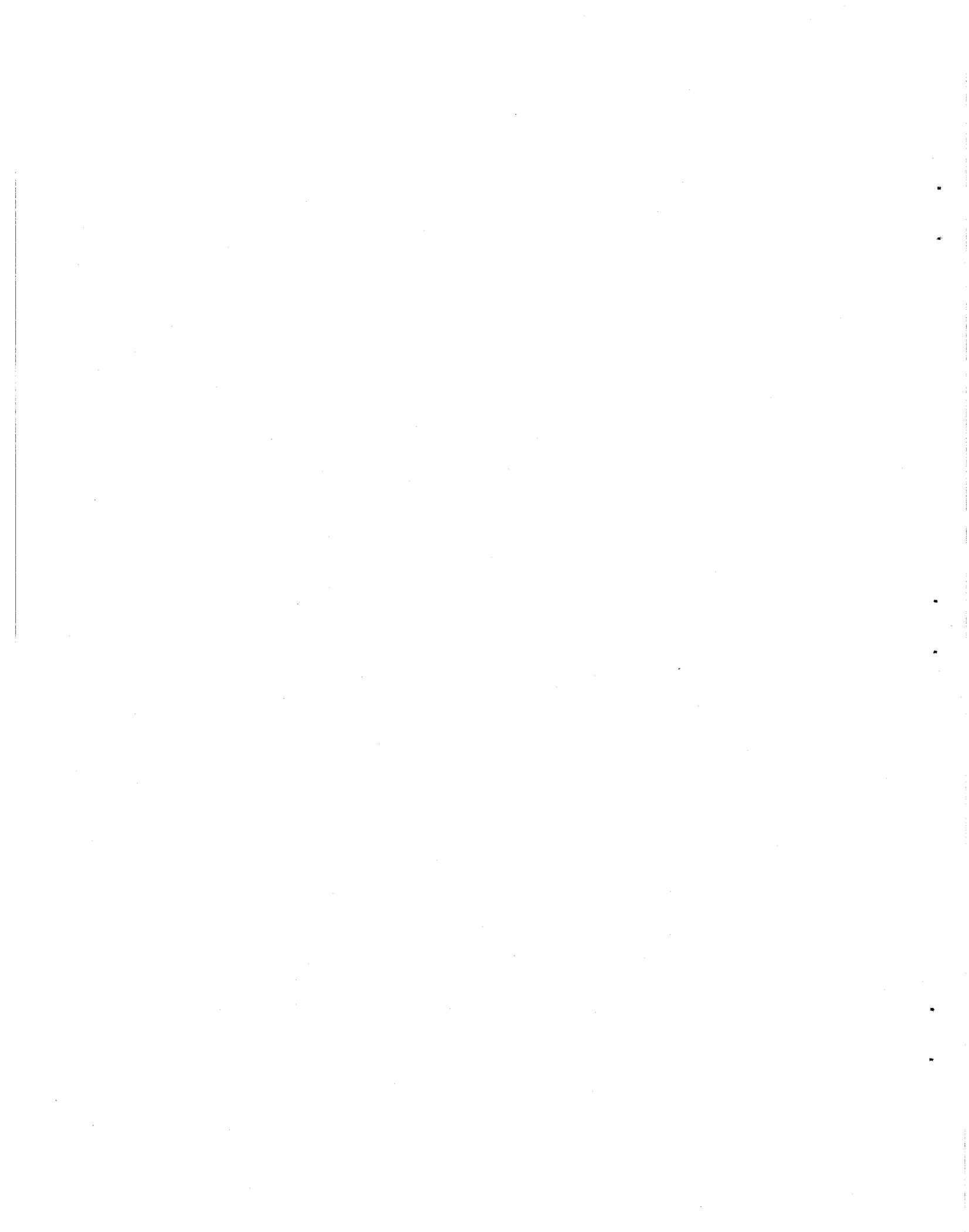


Fig. 5. Pictorial of ORCHIS System.



predicate calculus level and appropriate to hierarchical structures. The user will see less formal and topic-oriented command languages precisely defined in the symbolic logic used. Implementation can then take advantage of the methods of symbolic logic. In this manner common algorithms can be used to process an arbitrary variety of arithmetic expressions and logical expressions containing relational operators, including those appropriate for processing sets of information.

#### SYSTEM STATISTICS

The following table provides a brief summary of some system statistics indicating the variety of data contained in the system and the variety and extent of the services provided by the system to the research and development effort in Oak Ridge and elsewhere:

Number of subsystems		20
Number of data bases		123
ORCON	5	
ORSDI	6	
ADSEP	100	
Other	10	
Number of hard copy reports		50-60
Number of Organization Users		
ORNL		25
Non-ORNL		26

## Number of SDI users

Individuals	138
Projects	41
Profiles	289

## ORCON Statistics

Terminals	15
Organizations	13
Scheduled Time	10 hrs/day, 5 days/week
Number of References	300,000

## Users

Regular	140
Secondary	1000

## Number of ORLOOK users

Individuals	15-20
Searches	100/month

## SYSTEM CONTENTS

This report is not intended to document the many information bases which are currently in or have been processed through the ORCHIS system. More than 100 data bases have been involved, ranging from simple bibliographic through tabular text and research data bases containing mixed alphanumeric and digital information to purely digital data bases. The criteria for being contained within the ORCHIS system are that the data

base is (1) an ADSEP data base, (2) an ORCON or ORSDI data base, or (3) a non-ADSEP data base which has been analyzed by SADS. Since some of the data bases are descriptions of other collections of information, the total number of data collection locatable through the system is considerably larger than 100. Appendix II contains a partial list of data base titles, an indication of size, and type.

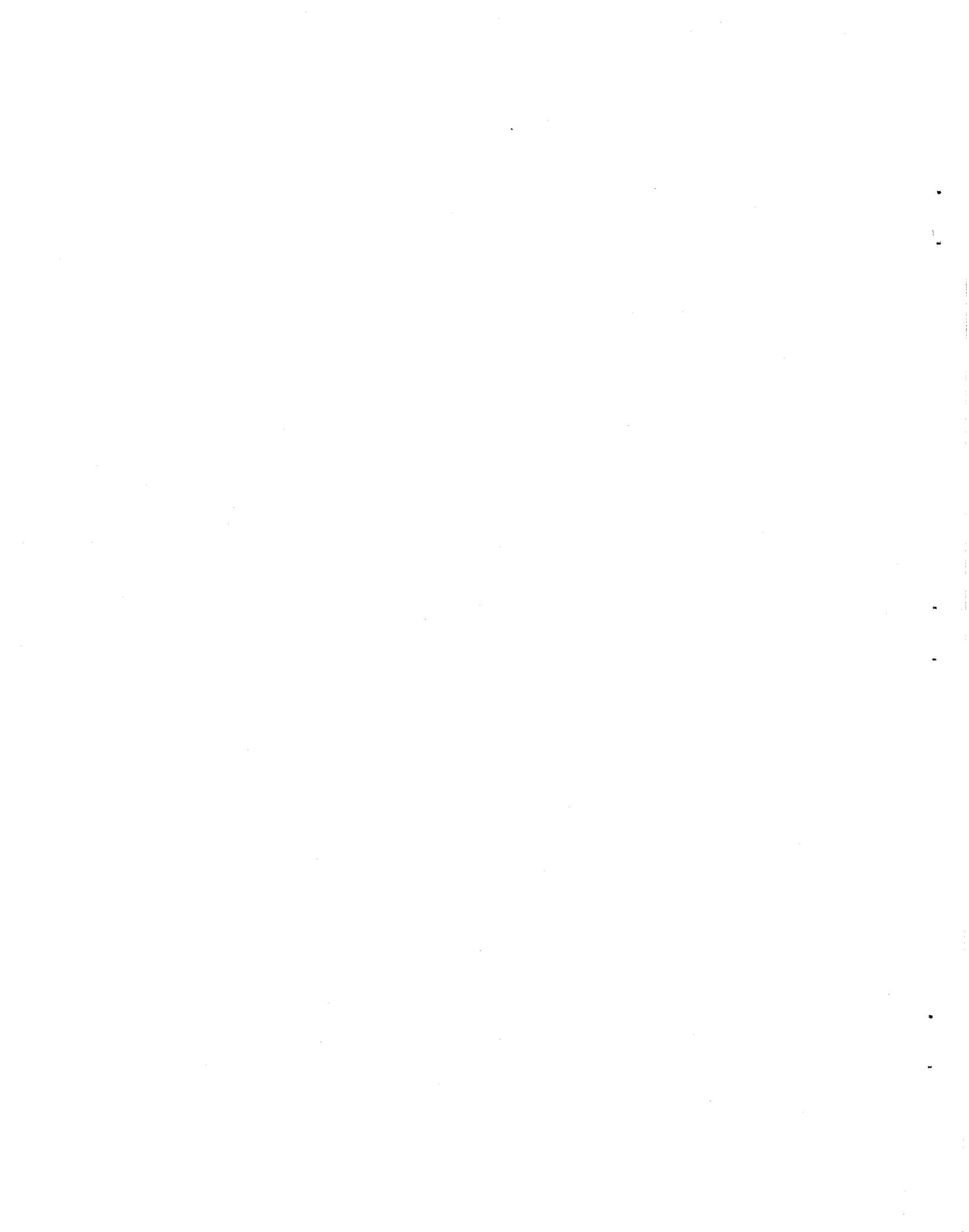
#### FUTURE OBJECTIVES

The following is a list of some tasks which could be undertaken to increase the services provided and extend the data processing capabilities of the system.

1. Completion of existing defined functions.
2. Increased system content and use.
3. Increased on-line services.
4. More "palatable" systems for quantitative analysis of structured data.
5. Increased automated processing of structured data by means of an expanded formal model.
6. Non-population models such as relational data bases.
7. Query processing based on predicate calculus.
8. Extensible "natural" language query systems.



APPENDICES



## APPENDIX I

## PROGRAM DESCRIPTIONS

This section of the outline describes each subsystem of program in Table I in terms of the program model in Fig. 4. A brief description of the program function, controls, input data, messages, output data, and its use of master data sets is given. The user should refer to other documentation (see Table I and Appendix III) for operating details. Many of the programs are still being developed and the documentation is not yet complete.

Name: AEDIT

Language: FORTRAN IV

Requirements: Core: 90K

Data Sets: Sequential

Function: This program provides error checking for ADSEP free-form text input. Input processed through AEDIT will be free of errors detectable in ADSEP except for incompatibilities with the contents of the master file.

Input: Control cards and the ADSEP text input.

Output: Error messages containing a brief diagnostic and indicating the line causing the error. Hard copy of the input may also be obtained.

Master Data Sets: One

Programmers: C. H. Gard, P. G. Fowler

Name: ADSEP

Language: FORTRAN IV

Requirements: Core: 220K

Data Sets: Sequential (1-4)

Function: This program provides the data management for the ORCHIS/ADSEP data bases, including updating (add, delete, correct, etc.), pooling with merging by field, editing including restructuring, and listing for proofing purposes. Input may be free-form text or formatted fields.

Input: The input contains request cards and updating information not necessarily on the same data set. The data set identifier table may be included as input.

Output: The output contains error messages, a transaction log, and an optional hard copy of updated records.

Master Data Sets: The program creates or manipulates master data sets as indicated above using from one to three data sets as required for the function requested.

Programmers: C. Gard, A. A. Brooks

Name: ADSORT

Language: FORTRAN IV

Requirements: Core: 225K

Data Sets: Sequential (2)  
Scratch (5-8)

Function: This program sorts an ADSEP data base on any number of arbitrary fields. The program also has the capability of forming a new ADSEP record for each  $I^{\text{th}}$  element of the Cartesian cross-product of any number of subsets from the existing ADSEP record. This record would contain the  $I^{\text{th}}$  element from the cross-product as a subset along with any other specified fields.

Input: The input contains the field names along with several control parameters.

Output: The output contains error messages as well as information about the fields being sorted (i.e., field missing, etc.).

Master Data Set: The program generates a sorted master data set identical to the original except in the cross-product case.

Programmer: F. D. Hammerling

Name: ADQIND

Language: PL/I

Requirements: Core: 130K

Data Sets: Sequential (1)

Function: This program, used in conjunction with the cross-product function of the ADSORT program, produces an index from an ADSEP data base with any arbitrary field as a reference. Each index term consists of a multiply qualified access phrase derived from the Cartesian cross-product previously formed.

Input: The input consists of several page formatting options, the field name of the cross-product subset, and the field name of the reference field.

Output: The output consists of error messages and a one- or two-column listing of the index suitable for publication purposes.

Master Data Sets: One

Programmer: F. D. Hammerling

Name: Deck Utility (DUT)

Language: FORTRAN IV

Requirements: Core: 130K

Data Sets: Sequential (2)

Function: The DUT program provides for the manipulation of card decks stored in a sequential file, usually a magnetic tape. Each deck is enclosed within a name card and an end card. The program provides for inputting and updating such data sets and verifies the sequence of cards when these cards represent hex program decks. The program can be used to produce copies of selected decks, listings of selected decks, and other utility functions associated with the manipulation of taped deck images.

Input: The input consists of the master tape containing a number of tape decks, the input stream containing control cards, and new decks to be added.

Output: The output consists of printed listings and updated master tapes.

Master File: The master file for this program consists of a sequential data set containing deck images.

Programmer: A. Hume

Name: CONTAD

Language: PL/I

Requirements: Core: Variable K

Data Sets: Sequential (3)

Function: This program, Conversion to ADSEP, provides for the conversion of some data bases to ADSEP input. Authority files and editing are provided. The framework can be used for other data bases by providing alternate paths. A field-by-field logic is used.

Input: Authority files and control cards for each field are required as well as the file to be converted.

Output: The ADSEP input file resulting from the conversion and the debugging output are provided.

Programmers: A. A. Brooks

Name: ADKEYS

Language: PL/I

Requirements: Core: 200K  
Sort Utility

Data Sets: Sequential

Function: This program offers the ORCHIS system with a method for the checking and removal of duplicate entries within data files. The contents of data fields within an entry are represented in a more concise form known as a key or a 32-bit machine word which is generated by performing a summation (logical add and carry) of all the machine words in the field. This technique permits every character in the string to be represented in the key. Weights are assigned to each key field to indicate relative importance. After records are sorted on highly reliable key fields and compared, acceptance or rejection logic is applied by quantitatively combining weights and testing against one or more threshold values.

Programmers: C. A. Giles, A. A. Brooks

Name: PUBLISH

Language: PL/I

Requirements: Core: 250K

Data Sets: Sequential (1-2)

Function: This program produces output from an ADSEP data base suitable for publication purposes. The report format may vary from a linear unordered structure to a three-level hierarchical structure of the subject material. Any text fields can be selected from the data base for publishing with arbitrary titles. A provision for editing the fields is also included.

Input: The input consists of several page-formatting options and the field names to be published with their titles.

Output: The output consists of the formatted material as described above, error messages, and number of records processed. The formatted output may be routed to a data set for printing.

Master Data Sets: One

Programmers: F. D. Hammerling, A. A. Brooks

Name: ADKWIC

Language: FORTRAN IV and Assembler

Requirements: Core: Simple (125K)  
Permuted (925K)

Data Sets: Sequential (1)  
Scratch (7)

Function: This program produces a simple or permuted index on any ADSEP field. The index may be produced in several formatting options for publication purposes.

Input: The input contains page-formatting options along with the field names to be indexed.

Output: The output contains error messages and the index in the format specified.

Master Data Sets: One

Programmers: D. L. Wilson, G. K. Haeuslein

Name: ADTABLE

Language: FORTRAN IV, Assembler

Requirements: Core: 170K  
IBM 360

Data Sets: Sequential

Printer: Universal character set

Function: This program provides flexible tabular reports from the ORCHIS/ADSEP data bases for publication. Conceptually, publications may be structures to produce volume, chapter, and page. Records must be pre-ordered, and the text must be edited for automatic sequential processing. A variable number of record fields are automatically text-aligned to be contained in columns of variable width. Each row of the table is a rectangular block that varies in height to contain the largest column of information from the record. Page size is optional with automatic page numbering.

Input: Control input is free form on cards. The command language consists of "English" keyword syntax.

Output: The report may be output to a high-speed printer with a universal character set. Page size and ribbon quality are optional. Mylar ribbons give sufficient quality for reproduction. Up to two pages are used if the table is wider than one page; pages may be joined for reproduction purposes.

Programmers: V. A. Singletary, A. F. Joseph, C. H. Gard, A. A. Brooks

Name: ADLABEL

Language: PL/I

Requirements: Core: 140K

Data Sets: Sequential (1-2)

Function: This program provides for labels or post cards produced from ADSEP data bases. Lists are conditional upon simple Boolean expressions. A variety of formats for variable and fixed message fields is provided.

Input: Standard format controls, request controls by arbitrary field, print messages, and selection criteria are placed in the input stream.

Output: Printed labels and post cards with variable fields and fixed messages.

Master Data Sets: Sequential (1)

Programmers: R. Slusher, A. A. Brooks

Name: MASINDEX

Language: PL/I

Requirements: Core: 75K

Data Sets: Sequential (3+)

Function: This program produces a master or global index for up to 100 data sets using the simple indices for each data set from ADINDEX.

The global index contains the access term, the data base numerical codes, and the number of occurrences of the access term in each data base.

Input: Control cards, separate indices, and old global indices.

Output: Updated global index and listing of global index.

Programmers: R. Slusher, A. A. Brooks

Name: ADLIST

Language: PL/I

Requirements: Core: 100K

Data Sets: Sequential (1-2)

Function: This program provides for selective lists based on simple Boolean expressions. Several standard general formats using computed, variable PL/I format statements are provided and others can be simply added. The standard lists are intended for utility purposes but can meet many listing needs.

Input: The input consists of standard format controls, request controls by fields, and selection criteria.

Output: Sequential output file.

Master Data Sets: Sequential (1)

Programmers: R. Slusher, A. A. Brooks

Name: ADINDX

Language: PL/I

Requirements: Core: 160K

Data Sets: Sequential (3)

Function: This program permits the formation of simple indices from one or more arbitrary fields in an ADSEP data base. Drop lists and minimal string lengths are provided. Redundant access terms and appended "S" plurals are removed. In forming an access term from arbitrary fields, punctuation or other characters can be removed as well as redundant blanks. The access terms are prepared as ADSEP input for KEYWORD fields.

Input: The drop lists and field controls are placed in the input stream.

Output: A sequential file of text to be used as input to the ADSEP data base.

Master Data Set: Sequential (1)

Programmers: R. Slusher, A. A. Brooks

Name: CONFAD

Language: PL/I

Requirements: Core: Variable

Data Sets: Sequential (1-2)

Function: This program provides a skeleton for converting ADSEP data bases to other formats. The program requires that the user supply a procedure that performs the necessary initialization, processes or edits the field, and outputs the resulting data in the desired format.

Input: The input consists of the field names the user wishes to process along with other control parameters.

Output: The output contains error messages and the number of records processed. Hard copy of the data in the new format may also be obtained.

Master Data Sets: One

Programmer: F. D. Hammerling

Name: ORLOOK

Languages: FORTRAN IV, Assembler

Requirements: IBM 360

On-Line: Environment: TSO Core: 200k  
 Disks: 2314 (1600 Trks Scratch)/User, Random Access  
 Data Sets: On-Line, Sequential (1-99), <4000 Records

Off-Line: Environment: OS (Batch) Core: 330k  
 Disks: 2314 (Entire Pack) Random Access  
 Data Sets: On-Line/Off-Line Sequential (1-30)  
 <25000 Records

Function: This program is designed for performing numerical or text searches on and obtaining output from subsets of ORCHIS/ADSEP data bases. It is particularly useful for user-interaction and decision-making during execution via teleprocessing facilities. An off-line noninteractive version is available for searching large data bases. Classed or unclassified searches may be made on simple or compound (hierarchically structured) key fields specified in a search profile. Logical functions may be requested on combinations of key fields in the search. Resulting subsets may be subsequently searched, combined logically, and output.

Input: Control input for the on-line version is entered through remote typewriter terminals and transmitted over standard telephone lines. Command language input is syntactical but input in free form. Off-line input is identical and is entered as free-form card input.

Output: Output includes error messages, optionally selected fields of records of the entire data set, or any search result subset. The on-line version may have records output to the remote terminal or to the high-speed line printer. A summary table can be displayed for the searches or combinations of subsets.

Programmers: V. A. Singletary, D. W. Altom, C. H. Gard, J. D. McDowell, A. A. Brooks, A. F. Joseph, J. S. Crowell

Name: ORSDI

Language: 360 Assembler, PL/I

Requirements:

Core for Search: 170K  
 Core for Largest Conversion: 330K  
 On-line: Direct Access Storage Device  
 (2311, 2314, or 2321)  
 Terminals: IBM 2260 or 2741

Data Sets: Sequential, Direct  
 Access

Function: This collection of programs provides ORCHIS with a valuable tool for searching very large bibliographic or text data bases, such as those produced by commercial suppliers. These can best be searched in a batch environment. Both retrospective and current awareness services are available. Profiles of researchers' interests are entered into the system via computer terminals and are stored and maintained on-line. Useful techniques available for profile construction include term and threshold weighting, Boolean logic, left- and right-term truncation, and different sorting techniques. A single profile can be written to search any number of data bases. A sequential search is done by matching terms in the profile on a character-by-character basis with terms in the data base.

The profile management, search, and output programs are written in 360 BAL. All data bases are converted into a standardized directory-oriented file structure which uses variable length data fields. The conversion programs and several utility programs are written in PL/I. Programs are basically table-driven; and optional card, paper, or tape output is available.

Current Data Files Include: Nuclear Science Abstracts, Biological Abstracts, Bio-research Index, Chemical Abstracts Condensates, Chemical

Biological Activities, NTIS Bibliographic Data File, and Cataloging and Indexing (U.S. Department of Agriculture). Future plans are to include Searchable Physics Information Notices and Metals Abstracts.

Programmers: C. A. Giles, A. H. Tindell

Name: ORCON

Language: IBM 360 Assembly

Requirements: Core: 250K

Data Sets: Direct Access ISAM,  
Sequential Indexed

Function: The ORCON program provides an on-line interactive retrospective search capability for inverted index files. The principal file consists of five years of Nuclear Science Abstracts with a number of smaller, locally produced files. The program services a dedicated multidrop line to 15 terminals located from the East Coast to the West Coast. The terminals are a CRT-type terminal with an optional teletype-compatible print device. The user may select a file and obtain a subset of that file based upon a sequence of simple Boolean operations. The user has the option of displaying the selected subset on the screen, typing it at his terminal, or requesting printout at the central facility. An interface between the ADSEP data bases and the ORCON format exists.

Input: There is no user-submitted input stream for this type program, rather the user submits a sequence of commands at his local terminal.

Output: The program optionally produces a local display and typing of selected subsets as well as an optional printing of the contents of the selected subset at the central facility.

Master Data Sets: The system uses one or more large linear files containing the bibliographic information as well as two levels of ISAM files

which form the index to the linear file and on which the logical operations are performed.

Programmers: G. K. Haeuslein, A. Culkowski, W. B. Malthouse, R. E. Plemens, D. G. Taylor

Name: SADS

Language: FORTRAN IV, Machine Language

Requirements: Core: 470K  
Overlays (2)

Data Sets: Sequential (1)

Function: SADS, Statistical Analysis of Data Subsets, provides Boolean expressions for processing hierarchical structures of mixed alphanumeric and digital fields. The selected data can be subjected to statistical analysis including linear regression. The SADS program is a compiler for a block structure, declarative language which has a FORTRAN IV interface to accommodate extended analytical capabilities.

The Boolean operators are defined for processing hierarchical structures and are modified by the declaration of the left-hand (variable) operand. The language thus permits processing hierarchical structures of unspecified length as all looping is controlled by the data base pointer system. Qualified structured constants may be specified as the right-hand operand.

Input: Program text and dictionary, if used, are required.

Output: Program diagnostics and statistical (or other) analysis output.

Master Data Set: ADSEP or other appropriate.

Programmers: C. H. Gard, A. Hume, A. A. Brooks

Name: TRACE

Language: PL/I

Requirements: Core: Dynamic

Data Sets: Sequential (1-6)

Function: This is an information retrieval program designed to trace the family tree, i.e., ancestors and progeny, of a given node or nodes in a directed graph and to retrieve information associated with each element of the tree, if desired.

The TRACE program provides the ability to select subgraphs from the directed graph representing a data base where the predecessor or successor relationship is entered as a field in each record. The five varieties of subgraphs which can be selected are (1) all the predecessors of an individual, (2) the successors of an individual, (3) all the predecessors and successors of an individual, (4) all of the successors of the first generation predecessors of an individual, and (5) all of the predecessors of the first generation successors of an individual. The program produces a display of the structure of the directed graph and itemizes the number of nodes and closed cycles associated with the subgraph.

Input: The user supplies a set of instruction cards as input to the program for his particular problem. These instructions include, for instance, such features as setting up new problems, retrieving old data sets, editing data sets, and presenting the output in various ways.

Output: The main output of the program is a pictorial display of the tree of any node in the directed graph. These traces can also be saved in a data set for analysis in other programs the user might have.

Master Data Sets: The program can also generate and edit master data sets as required.

Programmer: P. R. Coleman

Name: Name-Match

Language: Assembly PL/I

Requirements: Direct Access Device

Data Sets: Direct Access

Core: 336K

Sequential

Function: A modification of the Chemical Abstracts Service (CAS)

Name-Match System is the basis for all chemical information retrieved through the registry number by the Oak Ridge Computer Sciences Division. The data base for retrieval is the CAS Master Nomenclature-Registry File, a machine-readable file of chemical nomenclature containing information for 2.3 million compounds. Input to the system may be either a specific chemical name or a CAS registry number, a unique identification number assigned to a compound and associated with all names for the compound. Output includes the CAS registry number, CAS preferred index name, synonyms, and molecular formula. Registry numbers and synonyms obtained through Name-Match are very useful in searching bibliographic data bases for chemical information.

A subset, the Common Data Base, consisting of 20,000 compounds from the Master File, exists in ADSEP internal format. This subset has the Wiswesser Line Notation, not available on the Master File, for each entry as well as the registry number, molecular formula, and all synonyms.

Programmer: K. L. Kannan

Name: ADLIN

Language: PL/I

Requirements: New Program Under  
Development

Data Sets:

Function: This program will provide an arbitrary field-by-field conversion of ADSEP data bases to arbitrary LINOTRON input. Procedures will be established for classes of total field conversion, editing of specific types of fields to LINOTRON input, and lastly the conversion of control flags in ADSEP fields to be translated into LINOTRON control flags.

Programmer: F. D. Hammerling

Name: ADCLEAN

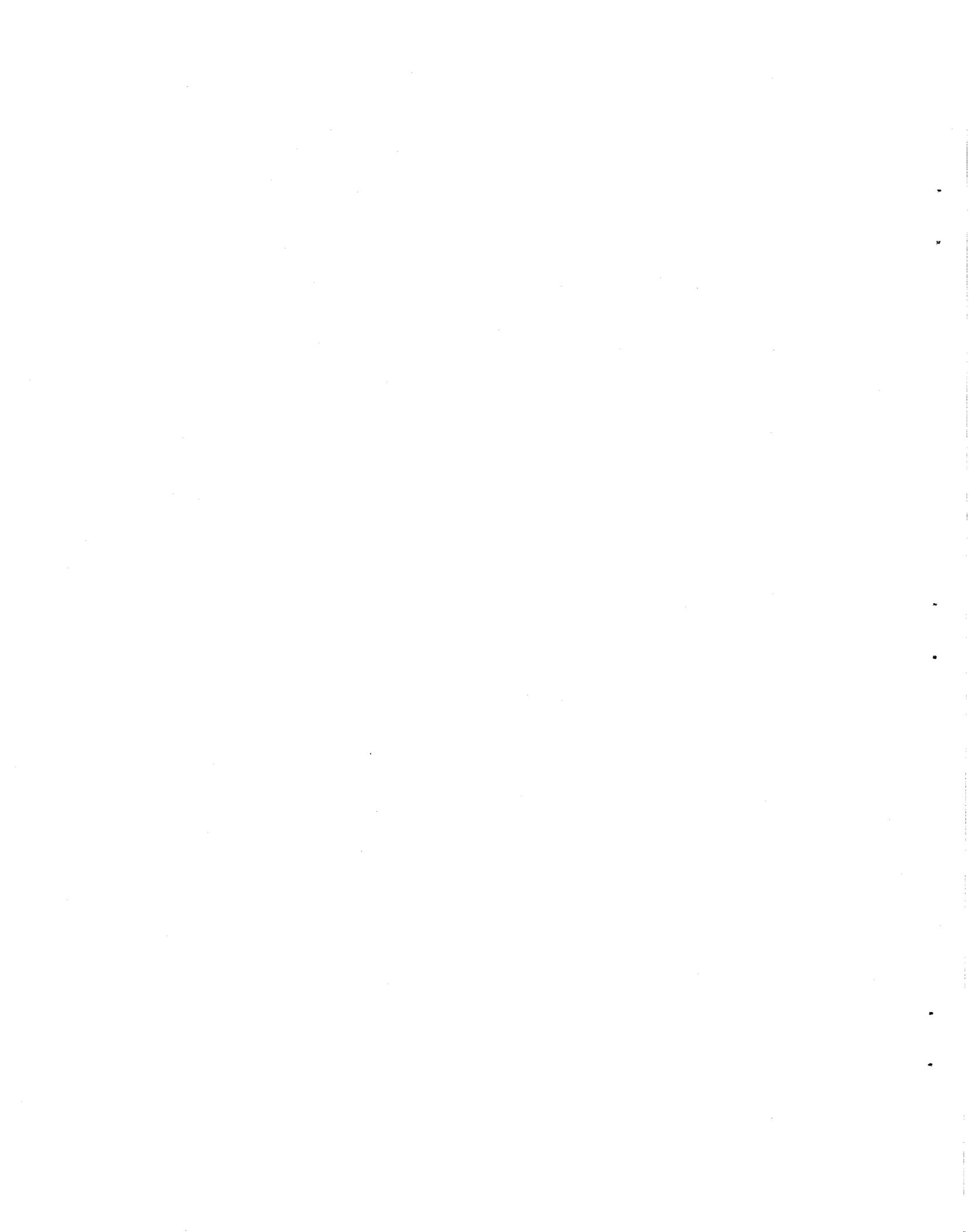
Language: PL/I

Requirements: New Program Under  
Development

Data Sets:

Function: This program will provide for the cleanup of ADSEP data bases with such features as inversion of author names, conversion of all upper case to lower case, generation of access terms from arbitrary ADSEP fields with drop lists, and so forth.

Programmer: F. D. Hammerling



## APPENDIX II

## SUMMARY OF ORCHIS DATA BASES

Table II presents a summary of ORCHIS data bases by title, number of entries, and type to illustrate the variety of information being processed. Further documentation is being prepared in the form of a searchable ADSEP file and hard copy report which details the content and availability of these data bases.

The type of each data set is coded as follows:

## 1. ADSEP Data Sets

- B - bibliographic
- A - alphanumeric other than bibliographic
- D - numeric data
- M - mixed (A&D)
- T - tabular display

## 2. Other Data Sets Used in ORCHIS

- R - RECON
- S - selective dissemination
- C - card decks analyzed by SADS
- N - data sets analyzed by TRACE

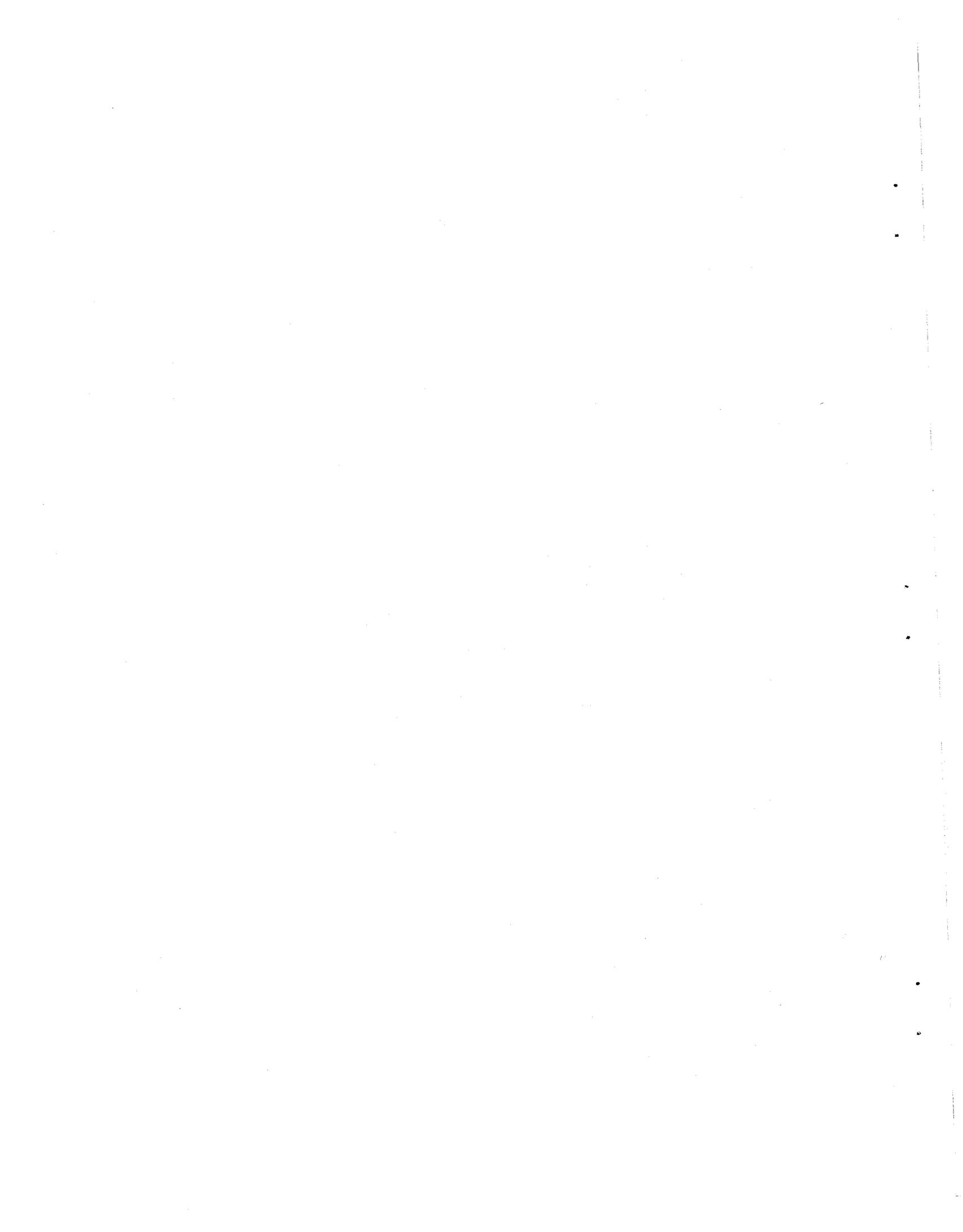


TABLE II

## SUMMARY OF ORCHIS DATA BASES

Data Set No.	Descriptive Data Set Name	No. of Records	Type
1	ORCHIS Data Set Inventory	130	A
2	CAIN, USDA Bibliography	--	B
3	Environmental Personnel File	4,700	A
4	Biomedical Radionuclide Bibliography	10,500	B
5	Nuclear Reactor Site Description	116	M
6	Urban Growth-Overcrowding	411	B
7	EISO Document Acquisition	15,525	A
8	Chemical Mutagenesis Bibliography	10,008	B
9	Chemical Mutagenesis Tabular Information	1,200	T
10	Chemical Mutagenesis Bibliographic Abstract	300	M
11	Energy Technology Bibliography	2,280	B,R
12	Ecological Effects of Strip Mining	963	B
13	Energy R&D Project Inventory	4,631	M,R
14	Biomedical Project Inventory	400	B
15	Biomedical Project Inventory (proprietary)	1,017	B
16	Environmental Directory	11,534	A

Data Set No.	Descriptive Data Set Name	No. of Records	Type
17	Environmental Impact of Nuclear Power Plants	1,734	B
18	Fish Protection Technology	100	B
19	Forest Service Elsner Recreation Bibliography	411	B
20	Forest Service Thesaurus	200	B
21	Forest Service Bibliography	6,000	B
22	Species Diversity	89	B
23	Environmental Terminology Index	--	A
24	Inventory of World Energy Resources	64	A
25	Environmental Law Abstracts	1,556	B
26	Regional Modeling Bibliography (I-V)	3,581	B
27	Biological Effects of Indium	208	B
28	Biological Effects of Selenium	284	B
29	Biological Effects of Technetium - 99m	2,145	B
30	Cadmium Toxicity	731	B
31	Clinical Toxicology of Commercial Products (CTCP)	1,052	B
32	Biological Implications of Dioxin	275	B
33	The Toxicology of Hexachlorocyclohexane	170	B

Data Set No.	Descriptive Data Set Name	No. of Records	Type
34	Toxicology and Biological Effects of Hydrazine and Derivatives	289	B
35	Insecticide Effects of Enzymes	457	B
36	Toxicology of Lead	495	B
37	Polychloroicphenyls in the Environment	880	B
38	Formulations of Commercial Products	18,000	B
39	TIRC Query-Response File	1,910	B
40	Toxicity of Selected Pesticides	10,000	B
41	Biological Effects of Saccharin	117	B
42	Air Pollution by Heavy Metals	2,225	B
43	Environmental Modeling	9,819	B
44	Environmental Quality - Biological Indicators	600	B
45	Environmental Transport	495	B
46	Mercury in the Environment	562	B,R
47	Toxic Materials in the Environment	4,335	B,R
48	Urban Technology - Liquid Wastes	414	B
49	Urban Technology - Solid Wastes	2,000	B

Data Set No.	Descriptive Data Set Name	No. of Records	Type
50	Ecological Niche Width Theory	60	B
51	Walker Branch Watershed Nutrient Data File (10)	100-200 each	N
52	TVA Forest Resources	431	M
53	Biogeochemical Ecology	1,998	B
54	Ecological Sciences	1,308	B
55	Ecosystems Analysis	231	B
56	Environmental Impact on Aquatic Organisms	100	B
57	Environmental Plutonium	2,200	B
58	Radioecology	1,352	B
59	Radionuclide Cycling in Plants and Soils	1,255	B
60	Thermal Effects on Aquatic Organisms	1,059	B
61	International Biological Program Abstract Journal	1,082	B
62	Bird Census Data	1,000	B
63	IBP Environmental Data Set Descriptions	78	A
64	Forest Resources Inventory, USDA	69,000	B
65	Modeling of Ecological Systems	312	B
66	Nitrogen Measurement in Ecosystems	255	B

Data Set No.	Descriptive Data Set Name	No. of Records	Type
67	Oak Ridge Weather Data	7,115	D
68	Ecological Succession	303	B
69	Expedient Shelter for Civil Defense	72	B
70	Tunneling Technology	1,400	B
71	Biological Abstracts	--	S
72	Biological Previews	--	S
73	Chemical - Biological Activities	--	S
74	Chemical Condensates	--	S
75	Nuclear Science Abstracts	295,000	R
76	U.S. Government Research Abstracts (GRA)	--	B
77	Desalination Materials Abstracts	2,000	B
78	Nuclear Desalination Processes	870	B
79	Desalination Reference File	1,762	B
80	Saline Water Conversion	1,500	B
81	Heated Effluents, Part I	1,856	B
82	Heated Effluents, Part II	2,314	B
83	Molten Salt Reactor	373	B
84	Magnetic Materials	304	B

Data Set No.	Descriptive Data Set Name	No. of Records	Type
85	Refractive Indices of Optical Materials	126	B
86	Radiation Shielding Computer Codes	212	B
87	RSIC Transactions Requests	100	A
88	Mammalian Metabolism and Dosimetric Aspects	50	B
89	Health Physics Info. Sys.	104	B
90	Nuclear Data Keyword and Reference File	15,331	B
91	Mechanical Properties of Metals	1,000	M
92	Environmental Impact Statement Questions	2,000	A
93	Materials Resources and Recycling	324	B
94	Radioecology Symposium Index	--	B
95	Butylated Hydroxytoluene Bibliography	172	B
96	AEC Chairman Reference File	200	B
97	Insecticide Use for Forest Products	--	T
98	Forest Trees Diseases	--	A
99	Instrumentation for Environmental Monitoring	10	B

Data Set No.	Descriptive Data Set Name	No. of Records	Type
100	Nuclear Desalination Reference File	1,762	B
101	TVA Forest Plot Data	4,000	M
102	Low Level Radiation Experimental Data	22,000	M
103	Census LEED File	7,135	M
104	Census Migration File	180,000	M
105	Pilot Plant Data (classified)	--	D
106	Production Plant Data (classified)	--	D
107	Production Line Data (classified)	--	M
108	Mouse Pathology Data (Air Contaminants)	--	M
109	INDIANA File	2,000	B
110	Water Resource Abstracts	--	R
111	Atomic and Molecular Radiation Physics	In Progress	M
112	Atomic Physics	In Progress	M
113	Comparative Animal Research Lab	In Progress	M
114	Bio-Application of Chemicals	In Progress	M
115	Environs	In Progress	M

---

Data Set No.	Descriptive Data Set Name	No. of Records	Type
116	Internal Exposure	In Progress	M
117	Medical Physics of External Dosimetry	In Progress	M
118	Low-Level Effects	In Progress	M
119	Pathology	In Progress	M
120	Health Aspects of Plutonium	In Progress	M
121	Reference Man	In Progress	M
122	Epidemiological Studies	In Progress	M
123	Theoretical Radiation Physics	In Progress	M

---

## APPENDIX III

## ANNOTATED BIBLIOGRAPHY

The following references (collected by topic) include the majority of documents defining and documenting the ORCHIS system. Some of these documents are in progress but are included here so that the reader will be aware of their future existence as they will provide information necessary to the use of the system. The CF memos not listed as "Internal Use Only" are or will be available as reports.

- 1.1 D. Rose, et al., "The Case for National Environmental Laboratories," ORNL-TM-2887 (December 1969).

The need for a National Environmental Laboratory is discussed. This discussion includes the need for comprehensive information processing capabilities to meet the needs of the interdisciplinary projects essential in environmental work.

- 1.2 J. L. Liverman, et al., A Proposal to the National Science Foundation - The Environment and Technological Assessment, March 1970.

This proposal to the National Science Foundation was the basis for the discussion during the summer of 1970 which led eventually to the formation to the ORCHIS project.

- 1.3 J. L. Liverman, et al., The Environment and Technological Assessment Progress Report June - December 1970; ORNL/NSF-EP-3, February 1971, Chapter II-0.

This report summarizes the discussions of the summer of 1970 wherein the needs and broad goals of information processing support for environmental and technical programs are summarized.

- 1.4 J. L. Liverman, et al., The Environmental and Technological Assessment Research Proposal - Submitted to NSF, February 1971.

This proposal submitted to the National Science Foundation covers the first year of this project at the Oak Ridge National Laboratory. It outlines in more detail the objectives of the information processing project as it relates to the environmental proposal.

- 1.5 N. B. Gove and A. F. Joseph, "Preliminary Survey of Information Retrieval Systems for IRRPOSI," ORNL-CF-71-1-27 (January 1971).

This memo is a survey of information and retrieval systems for IRRPOS. The study was limited to bibliographic-type retrieval systems where text information was the primary content of the data base. Information gained from the specifications of these systems contributed to the ultimate plans for the ORCHIS system.

- 1.6 A. A. Brooks, "Project for Information Processing," ORNL-CF-71-5-37 (May 20, 1971).

This memo is the first presentation of plans for the ORCHIS project. It is the first description of plans which include bibliographic and text searching SDI services and the preparation and analysis of research-type data bases. It represents the formal inception of the ORCHIS project.

- 1.7 A. A. Brooks, "Capital Equipment for Information Retrieval," ORNL-CF-71-11-22 (November 16, 1971).

This memo, which is for "Internal Use Only," examines the need for on-line storage space and other capital equipment for the ORNL Computer System which would permit the development of the ORCHIS system and its implementation on the ORNL computers.

- 2.1 A. A. Brooks, "A Data Set Edit Program and the All Mighty Mouse Subprogram," K-1748 (June 24, 1968).

This report describes the DESEP Update Program which was the original data set management program for the storage structure which is common to all ADSEP data bases. While this program has been superseded, this memo is the only documentation for the thesaurus term processing program and subroutines.

- 2.2 A. A. Brooks, "ADSEP - An Automated Data Set Editing Program," CTC-34 (May 1970).

The ADSEP program, which is the extension of the DESEP program, is discussed and documented. This report with a small collection of addenda represents the principal documentation of the update system. The concepts and implementation of the variably dimensioned structures are presented in this document.

- 2.3 A. A. Brooks, "Adoption of Generalized Record Format, ORNL-CF-71-1-9 (January 1971).

This memo documents the adoption of the generalized record format as a common internal format for the ORCHIS system. The nature of the structure used and its applications are discussed in detail.

- 2.4 A. A. Brooks, "ADSEP Text Input," ORNL-CF-71-9-3 (September 6, 1971).

This memo describes an addition to ADSEP input routines which permits the entry of free-form text into the system. This extension is the basis for all of the data bases containing text. This memo outlines the constraints which are placed upon the formation of text fields in the ADSEP program.

- 2.5 A. A. Brooks and D. L. Wilson, "ADBASE - A Subprogram to Access ADSEP Data Sets," ORNL-CF-71-9-21 (September 20, 1971).

This memo documents a subroutine which is designed to permit the casual programmers to access ADSEP data sets. The subroutine exists in two forms: (a) a FORTRAN subroutine which provides for the complete facilities of manipulating ADSEP data bases, and (b) a PL/I procedure which will permit access an ADSEP data base for the purpose of retrieving data only.

- 2.6 A. A. Brooks, "Implementation Specifications in the Extended ADSEP Input Subroutines, Their Description and Use," ORNL-CF-72-1-4 (January 1972).

The specifications presented in this memo describe extended input subroutines. This description documents the development of the structure concept with substructure fields being implemented in subroutines with open-ended definitions in order that the routines can be extended in the future to meet additional requirements.

- 2.7 A. H. Culkowski, D. W. Cardwell, R. B. Hofstra, A. M. Legg, and N. B. Gove, "A Study of Four Methods of Entering Text Into Computers," ORNL-TM-3600 (November 1, 1971).

This memo compares several methods of entering text into computers. A brief description of each of the four systems is presented - strong points, advantages, and disadvantages of each system being given. Conclusions are drawn as to which method would be the most advantageous.

- 2.8 C. J. Oen, N. F. Ferguson, and D. K. Truby, "Guide to the Generalized Bibliographic Format for the Environmental Information System," ORNL-EIS-71-3 (February 1972).

This memo documents a generalized abstract format which is an example of the free-form text input to the ADSEP program. It can be considered a prototype for other text entry formats for other purposes.

- 2.9 P. G. Fowler and A. A. Brooks, "PREEDIT - A PREEDIT Program for ADSEP Text Input," ORNL-TM-3897 (August 1972).

A FORTRAN program written for the purposes of editing ADSEP text input to locate errors in input sequence and input format, and for the absence of essential fields.

- 2.10 S. W. Nestor and C. W. Nestor, Jr., "Guide to the Use of the ORNL CRBE System for Text Entry, Text Editing and Information Retrieval," ORNL-CF-72-6-26 (July 1972).

This memorandum was prepared to assist the novice user of the ORNL CRBE system in text entry, text editing, and rudimentary information retrieval. It is assumed that the user will consult the ORNL CRBE manual for detailed instructions. This summary of operations is intended as an outline by which the new user may recall procedures demonstrated to him in an introductory session.

- 3.1 A. Hume and A. A. Brooks, "SADS, Statistical Analysis of Data Subsets, A Status Report," K-DP-3310 (July 26, 1968).

This status report covers the original tentative definition of the SADS Language. While a more up-to-date description of the SADS language is available, this report contains some alternative definitions which may be of value to the SADS users.

- 3.2 A. Hume and A. A. Brooks, "The SADS Program - Statistical Analysis of Data Subsets," CTC-19 (October 1969).

This report describes the program documentation for the SADS program. The first 70 pages contain the final definition of the language and examples for its use. The report also contains a description of the program maintenance system and listings for the program.

- 4.1 G. K. Haeuslein and Ann S. Klein, "The Oak Ridge National Laboratory KWIC Index," ORNL-4536 (May 1970).

This report documents the ORNL version of the KWIC index program which produces permuted word indices as well as simple indices.

- 4.2 D. Wilson, "The KWIC Program and ADSEP Storage Structure," ORNL-EIS-15.

This report documents the interfaces of the KWIC program to the ADSEP storage structure so that indices of arbitrary fields from ADSEP data bases can be readily produced.

- 5.1 "AEC/RECON Operation Manual," Lockheed Missile and Space Company (April 5, 1971).

This document contains the operating procedures for the Lockheed RECON program both for its use from the query terminals and its management from the master terminal.

- 5.2 A. A. Brooks, A. H. Culkowski, N. B. Gove, C. E. Hammons, G. K. Haeuslein, W. B. Malthouse, R. E. Plemens, H. S. Pomerance, and D. G. Taylor, "The AEC/RECON Project at ORNL," ORNL-4913 (in progress).

This report provides the most comprehensive description of the new AEC/RECON program and the activities of the RECON project.

6. V. A. Singletary and A. F. Joseph, "ORLOOK: Phase I Project Support Information Retrieval System," ORNL-EIS-15.

This report documents the ORLOOK program which is a text-matching technique applied to arbitrary fields of an ADSEP data base. It permits retrieval of selected subsets based upon Boolean logic expressions of information stored in text formats.

- 7.1 P. R. Coleman, "TRACE - An Information Retrieval Program For Tree-Structured Data," ORNL-4709 (October 1971).

This program permits a user to define the interrecord relationships in a data base as a cyclic-directed graph and permits retrieval of subsets of the data base as follows: all the predecessors of an individual record, all the successors of an individual record, a combination of all the predecessors and the successors of an individual record, all the successors of the first generation predecessors, and all of the predecessors of the first generation successors.

- 7.2 P. R. Coleman, "Use of the TRACE Program with the ADSEP Storage Structure," ORNL-CF-72-1-24 (in progress).

This document describes the interfacing of the TRACE program with the ADSEP storage structure. The predecessor-successor relationships of the cyclic-directed graph presenting the data base can be stored in any arbitrary list field of the storage structure. The first pass of this program through the data base constructs the directed graph information, and the second pass permits the extraction of the appropriate subsets of the data base.

8. R. C. Durfee, "ORRMIS - Oak Ridge Regional Modeling Information System," ORNL-CF-72-1-25 (in progress).

The ORRMIS program is a tool to be used for modeling purposes on large data bases. The data base is the direct-access type and only those cells which change as a function of time are updated. The ORRMIS program will be compatible with the ADSEP storage structure. Thus, the standard data base can be easily used as source information for modeling activities.

- 9.1 C. A. Giles, "The University of Georgia Tape Search and Profile Management Programs," ORNL-CF-72-1-26 (in progress).

This memo documents the University of Georgia Tape Search program and Profile Management program which constitutes the SDI component of the ORCHIS system. The profile management programs are on-line programs, and the tape search program can process many of the large abstract tapes currently available in a match mode.

- 9.2 C. A. Giles, "Data Base Description for CBAC in Text Search Format," ORNL-CF-72-11-14 (December 1972).

This report describes conversion of the CBAC data base to the University of Georgia standard file format so that it may be used as a part of the SDI system.

- 9.3 C. A. Giles and H. Pomerance, "Data Base Description for GRA - Government Reports Announcements," ORNL-CF-73-1-42 (January 1973).

This report describes the conversion of the Government Reports Announcement Data Base to the standard file format so that it can be used in the SDI system.

10. F. D. Hammerling and A. A. Brooks, "PUBLISH - A Program for the Composition of Pages from ADSEP Data Bases," ORNL-CF-72-1-27 (in progress).

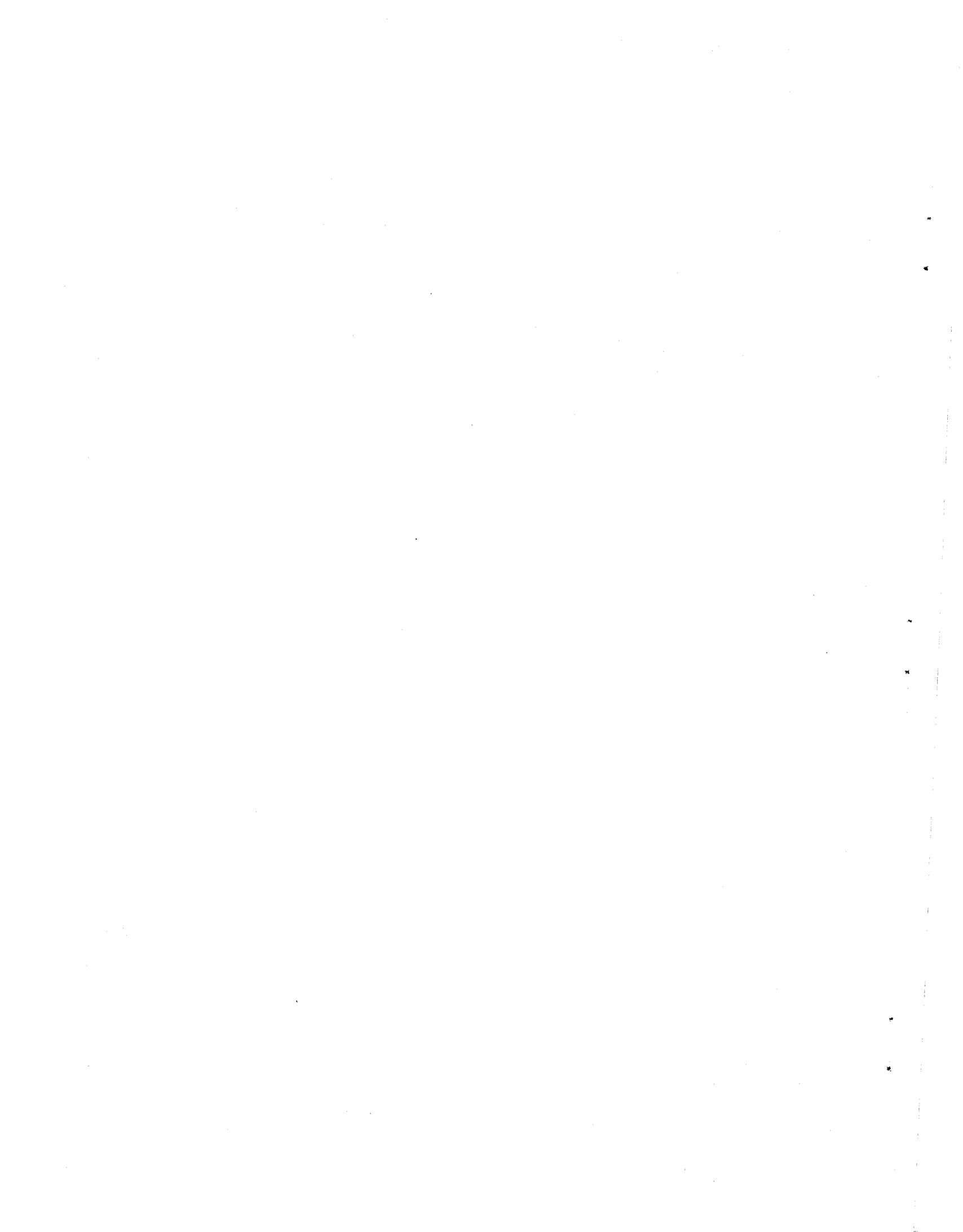
This program will produce a three-level hierarchical structured report from ADSEP data base. The nonstructured report can also be produced. The program will accomplish the pagination of the material into a two-column format with column balancing. A program to sort an ADSEP data base into order on arbitrary compatinated subfields is available. The hierarchical structure of the report must be indicated in the records, and the records must be sorted into the order of the hierarchy. The user has control over the choice of subfields published, their order, and their title.

- 11.1 H. J. Hamilton, "Streamlining the Chemical Abstracts Service's Name-Match System," ORNL-CF-72-3-38 (May 1972).

This report describes the changes made in the Chemical Abstracts Service Name-Match program to the ORNL environment. These changes were necessary to convert the program from an environment intended to produce the abstract literature to one which used the programs for access of that literature.

- 11.2 H. J. Hamilton, C. A. Giles, and D. Altom, "Implementation and Modification of the Chemical Abstracts Service Name-Match System - Phase I," ORNL-TM-3814 (July 1973).

This report is a more detailed description of the changes made in the Chemical Abstracts Name-Match system during its conversion to the ORNL environment.



## APPENDIX IV

## GLOSSARY OF ACRONYMS, NAMES, AND COMPUTING TERMS

Acronyms and Program Names

ADCLEAN - ADSEP Clean-Up Program (see p. 35)

AEDIT - ADSEP Edit Program (see p. 19)

ADINDX - ADSEP Indexing Program (see p. 27)

ADKEYS - ADSEP Key Program (see p. 23)

ADKWIC - ADSEP Key Word in Context Indexing Program (see p. 24)

ADLABEL - ADSEP Label Program (see p. 26)

ADLIN - ADSEP Linotron Input (see p. 35)

ADLIST - ADSEP Listing Program (see p. 27)

ADQIND - ADSEP Qualified Indexing Program (see p. 21)

ADSEP - Automated Data Set Edit Program (see p. 20)

ADSORT - ADSEP Sort Program (see p. 20)

ADTABLE - ADSEP Table Generation Program (see p. 25)

APTIC - Air Pollution Technology Information Center

CAIN - Catalogue and Indexing Data File

CAS - Chemical Abstracts Service

CONFAD - Conversion From ADSEP Format (see p. 28)

CONTAD - Conversion To ADSEP Input Program (see p. 22)

DUT - Deck Utility Program (see p. 22)

EISO - Environmental Information Systems Office/ORNL

GRA - Government Report Abstracts

IBP - International Biological Program

LEED - Longitudinal Employee Employer Data

MASINDX - Master Index Program (see p. 26)

NAME-MATCH - A program to match chemical names against the Chemical and registry numbers (see p. 34)

NSF - National Science Foundation

ORCHIS - Oak Ridge Computerized Hierarchical Information System

ORCON - Oak Ridge version of the RECON Program (see p. 31)

ORLOOK - An on-line conversational text search program (see p. 28)

ORNL - Oak Ridge National Laboratory

ORSDI - Oak Ridge Selective Dissemination of Information Program (see p. 30)

PUBLISH - A Program to Produce Line Printer Reproduction Copy (see p. 24)

SADS - Statistical Analysis of Data Subsets (see p. 32)

SDI - Selective dissemination of information

TIRC - Toxicology Information Response Center

TRACE - A Program to Retrieve Directed Subgraphs (see p. 33)

TVA - Tennessee Valley Authority

USAEC - United States Atomic Energy Commission

USDA - United States Department of Agriculture

#### Computing Terms

Alphanumeric - Information stored in the computer as alphabetic, numeric, and special characters primarily for display purposes.

Batch (batchwise) - A computer program generally processed simultaneously with several other programs without regard to real-time.

Boolean - See symbolic logic.

Cartesian product (Cartesian cross-product) - A Cartesian cross-product is every possible combination formed from the elements of two or more sets.

- Data base - Synonym for information base, sometimes connotating research data.
- Data element - Data element consists of a single number, a single string of alphanumeric information, a vector, an array, a set, or some defined collection of the former.
- Digital information - Information stored in the computer in the form of numbers to be used for computational purposes.
- Dynamic - Changing in character during the execution of a computer program.
- File - A set of computer records which are collected together and identified as an entity for computer purposes, sometimes called a data set.
- Hierarchical structure - A structure wherein narrow concepts are logically contained within broader concepts.
- Information base - One or more collections of information which are brought together into some logical system and which can be accessed through appropriate indices or other access device.
- Logical expression - An expression containing alphanumeric variables and constants, digital variables and constants combined with relational operators further combined with the logical operators "and", "or", and "not". These expressions have the value of true or false.
- On-line program (on-line mode) - A computer program which is run under the control of the user by means of a terminal. The program may or may not be conversational or interactive which implies a real-time interplay between the computer program and the user.

Population model - A model for a collection of information involving a large number of conceptual individuals with whom the information is permanently associated.

Predicate calculus - A level of a symbolic logic in which relationships between premises and conclusions can be stated subject to the range of variables.

Record - A computer term for a segment of information contained in a computer file which is obtained from the file by one read operation.

Record structure - A record structure is the collection of pointers and data elements which constitute a single record. A record structure is the computer embodiment of a data structure.

Relational operator - Operators such as equal, not equal, less than, greater than, is contained in, etc., used to state the relationship between variables or variables and constants.

Retrospective search - A search on a complete time ordered data base - see Selective Dissemination of Current Information.

Selective dissemination of current information - Information selected periodically from current sources and disseminated to users.

Software - The collection of computer programs designed to carry out data processing functions.

Software system - A collection of programs which are related and which can be used in concert with one another.

Structured data elements - Data elements comprised of structured information such as vectors, matrices, sets of information, and other multidimensional structures, especially alphanumeric information.

Symbolic logic - One of the Boolean algebras consisting of operands and logical operators, the results of the operations being true or false.

## INTERNAL DISTRIBUTION

- |   |                        |
|---|------------------------|
| 1-3. Central Research Library                                   | 178. J. A. Elkins      |
| 4. Biology Library  | 179. J. L. English     |
| 5. ORNL - Y-12 Technical Library,<br>Document Reference Section | 180. D. E. Ferguson    |
| 6. ORGDP Library  | 181. N. Ferguson       |
| 7. Patent Office  | 182. P. G. Fowler      |
| 8-107. Computer Sciences Division Library                       | 183. W. Fulkerson      |
| 108-129. Laboratory Records Department                          | 184. C. H. Gard        |
| 130. Laboratory Records, ORNL R.C.                              | 185. D. A. Gardiner    |
| 131. H. I. Adler  | 186. C. A. Giles       |
| 132. N. G. Anderson   | 187. J. A. Gillcrist   |
| 133. L. L. Anthony  | 188. J. S. Gillen      |
| 134. S. I. Auerbach   | 189. H. E. Goeller     |
| 135. J. A. Auxier   | 190. R. O. Greene      |
| 136. P. S. Baker  | 191. H. E. Gregg       |
| 137. J. H. Barnes   | 192. G. K. Haeuslein   |
| 138. C. F. Barnett  | 193. F. D. Hammerling  |
| 139. R. O. Beauchamp  | 194. G. S. Henderson   |
| 140. S. E. Beall  | 195. R. W. Henderson   |
| 141. C. Becker  | 196. R. F. Hibbs       |
| 142. M. Bender  | 197. D. J. Horen       |
| 143. S. R. Bernard  | 198. J. W. Huckabee    |
| 144. G. Blaylock  | 199. F. C. Hutton      |
| 145. C. J. Borkowski  | 200. J. G. Jones       |
| 146-155. A. A. Brooks   | 201. K. L. Kannan      |
| 156. F. R. Bruce  | 202. S. V. Kaye        |
| 157. J. R. Buchanan   | 203. O. A. Kelly       |
| 158. R. L. Burgess  | 204. E. M. Kidd        |
| 159. D. W. Cardwell   | 205. J. W. Kidd        |
| 160. R. S. Carlsmith  | 206. A. S. Klein       |
| 161. J. A. Carpenter  | 207. T. A. Lincoln     |
| 162. S. F. Carson   | 208. R. Livingston     |
| 163. H. P. Carter   | 209. F. C. Maienschein |
| 164. G. M. Caton  | 210. H. V. Malling     |
| 165. Z. Combs   | 211. W. B. Malthouse   |
| 166. T. F. Connolly   | 212. B. F. Maskewitz   |
| 167. E. D. Copenhaver   | 213. F. K. McGowan     |
| 168. W. B. Cottrell   | 214. E. J. Nall        |
| 169. C. C. Coutant  | 215. D. J. Nelson      |
| 170. K. E. Cowser   | 216. C. J. Oen         |
| 171. C. W. Craven   | 217. J. S. Olson       |
| 172. J. S. Crowell  | 218. R. V. O'Neill     |
| 173. F. L. Culler   | 219. J. H. Owings      |
| 174. A. H. Culkowski  | 220. J. H. Paylor      |
| 175. B. E. Dinger   | 221. R. E. Plemens     |
| 176. J. C. Dougherty  | 222. H. S. Pomerance   |
| 177. R. C. Durfee   | 223. H. Postma         |
|   | 224. C. E. Price       |

- |                       |                         |
|-----------------------|-------------------------|
| 225. M. E. Ramsey     | 243. I. Thomas          |
| 226. A. M. Read       | 244. A. H. Tindell      |
| 227. S. A. Reed       | 245. D. B. Trauger      |
| 228. D. E. Reichle    | 246. D. K. Trubey       |
| 229. J. P. Robertson  | 247-271. G. U. Ulrikson |
| 230. T. H. Row        | 272. R. I. Van Hook     |
| 231. H. H. Shugart    | 273. N. H. Van Wie      |
| 232. W. D. Shults     | 274. J. S. Wassom       |
| 233. V. A. Singletary | 275. A. M. Weinberg     |
| 234. R. Slusher       | 276. G. W. Westley      |
| 235. R. H. Strand     | 277. J. C. White        |
| 236. E. G. Struxness  | 278. G. E. Whitesides   |
| 237. J. G. Sullivan   | 279. C. S. Williams     |
| 238. R. C. Taeuber    | 280. D. L. Wilson       |
| 239. R. Talmi         | 281. M. B. Wright       |
| 240. R. A. Tannert    | 282. S. L. Yount        |
| 241. D. G. Taylor     | 283. A. Zucker          |
| 242. E. H. Taylor     |                         |

## EXTERNAL DISTRIBUTION

284. J. D. Allen, Research and Monitoring, Environmental Protection Agency, Washington, D.C. 20545
285. D. D. Bacon, Forest Hydrologist, U.S. Department of Agriculture, Forest Service, 50 Seventh Street, NE, Atlanta, Georgia 30309
286. Yates Barber, Bureau of Sport Fisheries and Wildlife, U.S. Department of the Interior, Washington, D.C. 20240
287. N. F. Barr, Assistant Director for Radiological Physics, Division of Biomedical and Environmental Research, U.S. Atomic Energy Commission, Washington, D.C. 20545
288. F. P. Baxter, Division of Forestry, Fisheries, and Wildlife, Forestry Building, Tennessee Valley Authority, Norris, TN 37828
289. R. Bell, National Aeronautics and Space Administration, Washington, D.C. 20545
290. Dr. W. R. Bibb, U.S. Atomic Energy Commission, Oak Ridge Operations, Oak Ridge, TN 37830
291. J. C. Bresee, U.S. Atomic Energy Commission, Division of Applied Technology, Washington, D.C. 20545
292. R. A. Card, Corporate Research Department, Union Carbide Corporation P. O. Box 363, Tuxedo, NY 10987
293. R. Chartrand, Library of Congress, 10 First Street, SE, Washington, D.C. 20540
294. E. F. Codd, IBM, San Jose, CA
295. F. E. Croxton, Library of Congress, Director of the Administrative Department, 10 First Street, SE, Washington, D.C. 20540
296. Dr. R. Cumley, M. D. Anderson Hospital and Tumor Institute, Texas Medical Center, Houston, TX 77025
297. J. W. Curlin, Environmental Policy Division, Congressional Research Service, Library of Congress, Washington, D.C. 20540
298. J. J. Davis, Office of Directorate of Regulatory Standards, U.S. Atomic Energy Commission, Washington, D.C. 20545

299. Ruth Davis, National Bureau of Standards, Washington, D.C. 20234
300. M. Day, National Library of Medicine, Bethesda, MD 20014
301. P. B. Dunaway, Office of Effect Evaluation, U.S. Atomic Energy Commission, Nevada Operations Office, P.O. Box 1400, Las Vegas, NV 89114
302. R. J. Engleman, Division of Biomedical and Environmental Research, U.S. Atomic Energy Commission, Washington, D.C. 20545
303. E. H. Fleming, Assistant Director for Peaceful Nuclear Explosives, Division of Applied Technology, U.S. Atomic Energy Commission, Washington, D.C. 20545
304. J. F. Franklin, Deputy Director, Coniferous Forest Biome, Forestry Sciences Laboratory, 3200 Jefferson Way, Corvallis, OR 97331
305. R. R. Freeman, Environmental Science Information Center, Environmental Data Service, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, Rockville, MD 20852
306. R. Goor, National Science Foundation, Division of Environmental Systems and Resources, 1800 G Street, NW, Washington, D.C. 20550
307. N. B. Gove, IAEA, Karntner Ring 11, P.O. Box 645, A-1010, Vienna, Austria
308. J. A. Harris, U.S. Atomic Energy Commission, Division of Management Information and Telecommunications Systems, Washington, D.C.
309. J. M. Harvin, M. D. Anderson Hospital and Tumor Institute, Texas Medical Center, Houston, TX 77025
310. V. Hemple, Mail Drop L-380, University of California, Lawrence Livermore Laboratory, P.O. Box 808, Livermore, CA 94550
311. H. Hollister, U.S. Atomic Energy Commission, Division of Biology and Medicine, Washington, D.C. 20545
312. Dr. R. L. Holt, National Cancer Institute, National Institutes of Health, Bethesda, MD 20014
313. M. A. Huffer, Department of Interior, Director of Office of Library Services, 18th and C Streets, NW, Washington, D.C. 20240
314. T. E. Hughes, Chief, Systems Development Branch, Office of Information Services, U.S. Atomic Energy Commission, Washington, D.C. 20545
315. D. Hummel, National Library of Medicine, Bethesda, MD 20014
316. G. W. Johnson, Division of Applied Technology, U.S. Atomic Energy Commission, Washington, D.C. 20545
317. Dr. V. A. Kamath, Scientific Advisor, Attention: P. K. Patwardhan, Bhabha Atomic Research Centre, Trombay, Bombay, India
318. N. Kawakami, U.S. Atomic Energy Commission, Technical Information Center, P.O. Box E, Oak Ridge, TN 37830
319. H. Kissman, SIS/TIP, National Library of Medicine, Bethesda, MD 20014
320. I. Lebow, Technical Information Center, U.S. Atomic Energy Commission, P.O. Box E, Oak Ridge, TN 37830
321. H. Lieth, Department of Botany, University of North Carolina, Chapel Hill, NC 27514
322. Dr. J. L. Liverman, U.S. Atomic Energy Commission, Division of Biomedical and Environmental Research, Washington, D.C. 20545
323. T. G. McCreless, U.S. Atomic Energy Commission, Office of the Director of Regulations, Bethesda, MD 20014
324. R. D. McCulloch, IAEA, Karntner Ring 11, P.O. Box 645, A-1010, Vienna, Austria

325. C. T. Meadows, U.S. Atomic Energy Commission, Division of Management Information and Telecommunications Systems, Washington, D.C. 20545
326. Dr. E. A. Mirand, Roswell Park Memorial Institute, 666 Elm Street, Buffalo, NY 14203
327. B. Murray, National Cancer Institute, National Institute of Health, Bethesda, MD 20014
328. M. O'Hara, Forest Service, Washington, D.C. 20545
329. C. L. Osterberg, Division of Biomedical and Environmental Research, U.S. Atomic Energy Commission, Washington, D.C. 20545
330. A. Pepmueller, Organization 3150, Sandia Laboratory, Kirtland Air Force Base, E, Albuquerque, NM 87115
331. J. Powers, Defense Documentation Center, Defense Supply Agency, Cameron Station, Alexandria, VA 22314
- 332-333. M. Prostein, Argonne National Laboratory, P.O. Box 2528, Idaho Falls, ID 83401
334. R. Rabin, National Science Foundation, Division of Environmental Systems and Resources, 1800 G Street, NW, Washington, D.C. 20550
335. J. N. Rogers, Division 8321, Sandia Laboratories, P.O. Box 969, Livermore, CA 94550
336. Dr. M. E. Rose, Mathematical and Computer Sciences Program, Molecular Sciences and Energy Research, Division of Physical Research, U.S. Atomic Energy Commission, Washington, D.C. 20545
337. D. Ruchotzke, U.S. Atomic Energy Commission, Oak Ridge Operations, Oak Ridge, TN 37830
338. Dr. M. Savitz, Office of Energy Conservation, Department of Interior, Room 4458, 18th and C Street, NW, Washington, D.C. 20240
339. Dr. J. H. Schneider, National Cancer Institute, National Institutes of Health, Department of Health, Education, and Welfare, Bethesda, MD 20014
340. M. H. Schwartz, U.S. Atomic Energy Commission, Division of Management Information and Telecommunications Systems, Washington, D.C. 20545
341. R. L. Shannon, Technical Information Center, U.S. Atomic Energy Commission, Oak Ridge Operations, Oak Ridge, TN 37830
342. Dr. V. Slamecka, Department of Communications and Information Systems, Georgia Institute of Technology, Atlanta, GA 30332
343. N. M. Snell, U.S. Atomic Energy Commission, Division of Management Information and Telecommunications Systems, Washington, D.C. 20545
344. E. E. Stokley, U.S. Atomic Energy Commission, Century Office Building, Germantown, MD 20767
345. C. M. Tarzwell, Director, National Marine Water Quality Laboratory, U.S. Environmental Protection Agency, P.O. Box 277, Kingston, RI 02892
346. H. Taylor, National Wildlife Federation, Frazier Memorial Library, 1412 Sixteenth Street, NW, Washington, D.C. 20036
347. L. Tilley, Environmental Protection Agency, Region V, 1 North Wacker Drive, Chicago, IL 60606
348. J. Tow, University of Florida, Gainesville, FL 32601
349. J. Trew, Library of Congress, Washington, D.C. 20540

350. W. M. Vaden, Technical Information Center, U.S. Atomic Energy Commission, Oak Ridge Operations, P.O. Box E, Oak Ridge, TN 37830
351. G. M. Van Dyne, College of Forestry and Natural Resources, Colorado State University, Fort Collins, CO 80521
352. B. Vasta, Specialized Information Services, National Library of Medicine, Bethesda, MD 20014
353. H. Voress, U.S. Atomic Energy Commission, Washington, D.C. 20545
354. G. Ward, National Science Foundation, 1800 G Street, NW, Washington, D.C. 20550
355. V. A. Wentz, National Aeronautics and Space Administration, Washington, D.C. 20545
356. M. Wolfe, Department of the Army, U.S. Army Engineer District, Philadelphia Corps of Engineers, Custom House, 2D and Chestnut Streets, Philadelphia, PA 19106
357. Dr. S. Zimmerman, M. D. Anderson Hospital and Tumor Institute, Texas Medical Center, Houston, TX 77025
358. Research and Technical Support Division, AEC, ORO
359. Patent Office, AEC, ORO
- 360-528. Given distribution as shown in TID-4500 under Mathematics and Computers category (25 copies - NTIS)

