

SELECTED FEDERAL COMPUTER-BASED INFORMATION SYSTEMS

Edited by SAUL HERNER and MATTHEW J. VELLUCCI



INFORMATION RESOURCES PRESS
WASHINGTON, D.C. 1972

Copyright © 1972 by Information Resources Press, a division of Herner & Company.

All Rights Reserved.

Printed in the United States of America. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher.

Available from

Information Resources Press

2100 M Street, N.W.

Washington, D.C. 20037

Library of Congress Catalog Card Number 72-85016

ISBN 0-87815-007-2

PREFACE

Selected Federal Computer-Based Information Systems is an enlarged and updated version of *Selected Mechanized Scientific and Technical Information Systems* which was originally published in 1968. The original version consisted of 13 system descriptions; the present version consists of 35 descriptions. Reflecting recent developments in computer-based information systems, the present volume includes some exemplary systems that are not strictly concerned with science and technology. Thus, the change in title.

As in the case of the original version, the present volume was prepared in cooperation with the Panel on Operational Techniques and Systems of the Committee on Scientific and Technical Information (COSATI). In addition to helping to select and solicit the cooperation of Federal agencies maintaining major operational computer-based systems, the Panel aided in developing a checklist of subjects to be included in the system descriptions.

The descriptions comprising this volume are in two parts, verbal and graphic. Each selected agency was asked to furnish a word description of its system, plus flow charts to illustrate and clarify it. The verbal and graphic descriptions were edited and rewritten, as necessary, for clarity, completeness, and uniformity and returned to the submitting agencies for approval, emendation, last-minute updating, etc. The cycle from original solicitation and final approval of descriptions consumed approximately one year.

We are deeply indebted to COSATI and to Andrew A. Aines, who, as Acting Chairman, agreed to the cooperative role of the Panel in the preparation of this volume. We are also indebted to the members of the Panel and, most particularly, to Charles De Vore, who, as Panel Secretary, coordinated our efforts with those of the Panel. We also acknowledge with profound thanks the work of Gene P. Allen, Paula D. Kal, and Thomas A. Mulholland, who played major roles in rewriting and editing the verbal and graphic descriptions and preparing the manuscript for publication. Finally, we owe a deep debt of gratitude to the agencies that submitted descriptions and reviewed and corrected our sometimes garbled interpretations of them.

While many people were involved in the preparation of this volume, we assume full responsibility for its contents. We apologize to the contributors and the readers for any errors of commission, omission, or obsolescence due to the length of the editorial processes.

Washington, D.C.
July 1972

CONTENTS

U.S. DEPARTMENT OF AGRICULTURE

- Cataloging and Indexing System (CAIN) / 1
- Current Research Information System (CRIS) / 5

U.S. DEPARTMENT OF COMMERCE

- International Cooperation in Information Retrieval Among Patent Offices (ICIREPAT) / 20
- Maritime Research Information Service (MRIS) / 24
- National Climatic Center (NCC) / 27
- National Environmental Satellite Service (NESS) / 33
- National Meteorological Center (NMC) / 40
- National Oceanographic Data Center (NODC) / 43
- National Technical Information Service (NTIS) / 48
- Office of Computer Information (OCI) / 54

U.S. DEPARTMENT OF DEFENSE

- Defense Documentation Center (DDC) / 59

U.S. Department of the Air Force

- Management and Scientific Information System (MASIS) / 73

U.S. Department of the Army

- Army Research and Development Information System (ARDIS) / 79
- Documentation Automated Retrieval Equipment (DARE) / 84
- WRAIR Chemical/Biological Information System (WRAIR) / 86

U.S. Department of the Navy

- Marine Corps Automated Readiness Evaluation System (MARES) / 90
- Marine Corps Unified Materiel Management System (MUMMS) / 92

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

- Medical Literature Analysis and Retrieval System (MEDLARS) / 94
- Statistical Table Assembly and Retrieval (STAR) / 103

U.S. DEPARTMENT OF THE INTERIOR

- Bonneville Power Administration/Selective Dissemination of Information (BPA-SDI) / 108
- Engineering Reference Branch (ERB) / 113
- National Water Data System (NWDS) / 119
- Oil and Gas Field Study (OGFS) / 124
- Water Resources Scientific Information Center (WRSIC) / 129

U.S. DEPARTMENT OF TRANSPORTATION

Transportation Research Information System (TRIS) / 133
Highway Research Information Service (HRIS) / 142
Transportation Noise Research Information Service (TNRIS) / 159
Transportation Research Activities Information Service (TRAIS) / 162

U.S. ATOMIC ENERGY COMMISSION

Atomic Energy Commission/Division of Technical Information Extension
(DTIE) / 167
Nuclear Desalination Information Center (NDIC) / 179

THE LIBRARY OF CONGRESS

Machine-Readable Cataloging (MARC) / 185

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

NASA Scientific and Technical Information System (NASA) / 191

SMITHSONIAN INSTITUTION

Catalog of American Portraits (CAP) / 195
Natural History Information Retrieval System (NHIR) / 203
Science Information Exchange (SIE) / 209

U.S. DEPARTMENT OF AGRICULTURE

CAIN—System Overview / 4
CRIS—System Overview / 8
CRIS—MT/ST and Digi-Data Processing / 9
CRIS—Pre-Edit Process / 9
CRIS—Table Load Process / 10
CRIS—Edit, Validate, Maintenance Process / 11
CRIS—Keyword Maintenance Process / 12
CRIS—Split Abstract History Process / 13
CRIS—Print Field Error Process / 14
CRIS—History Subsystem / 15
CRIS—Catalog and Index Process / 16
CRIS—Résumé Subsystem / 17
CRIS—Special Inquiries Subsystem / 18
CRIS—Recurring Report Selector / 19

U.S. DEPARTMENT OF COMMERCE

ICIREPAT—System Overview / 22
ICIREPAT—Input Process / 23
NCC—System Overview / 32
NESS—Input Process / 35
NESS—Earth Location Process / 36
NESS—Picture Mapping Process / 37
NESS—Raw Signal Input Process / 38
NESS—ATS Picture Acquisition / 39
NMC—Global Telecommunication System / 41
NMC—Forecast System / 42
NODC—File Creation Process / 47
NTIS—System Overview / 52
NTIS—Input Process / 53
OCI—System Overview / 57
OCI—Input Process / 58

U.S. DEPARTMENT OF DEFENSE

DDC—System Overview / 67
DDC—Input and Announcement Process / 68
DDC—Work Unit (DD Form 1498)—Input Process / 69
DDC—Bibliography and Work Unit Report Request Process / 70
DDC—Document Request Process / 71 and 72

U.S. Department of the Air Force

MASIS—System Overview / 76
MASIS—Update Process / 77
MASIS—Report Process / 78

U.S. Department of the Army

ARDIS—System Overview / 80
ARDIS—Input Process / 81
ARDIS—Publication Process / 82
ARDIS—Reference and Query Response Process / 83

DARE—System Overview / 85

WRAIR—Input Process / 88
WRAIR—Retrieval Process / 89

U.S. Department of the Navy

MARES—Input Process / 91

MUMMS—Requisition Process / 93

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

MEDLARS—System Overview / 100
MEDLARS—Input Process / 101
MEDLARS—Publication Process / 102
MEDLARS—Retrieval Process / 102

STAR—Input Process / 106
STAR—Reference and Query Response Process / 107

U.S. DEPARTMENT OF THE INTERIOR

BPA-SDI—System Overview / 112

ERB—Input Process / 116
ERB—Publication and Announcement Process / 117 and 118

NWDS—Ground-Water Data Processing / 121
NWDS—Water-Quality Data Processing / 122
NWDS—Surface-Water Data Processing / 123

OGFS—System Overview / 126
OGFS—Input Process / 127
OGFS—Search and Retrieval Process / 128

WRSIC—Input Process / 131
WRSIC—Publication and Announcement Service / 132

U.S. DEPARTMENT OF TRANSPORTATION

TRIS—Conceptual Model / 134
TRIS—Acquisition and Selection Process / 135
TRIS—Abstracting, Indexing, Classification and Coding Process / 136
TRIS—Keyboarding, Editing and Storage Process / 137 and 138
TRIS—File Processing / 139
TRIS—Retrieval Process / 140
TRIS—Output Process / 141

International Network for Highway Research Information / 146
International Flow of Information on Current Highway Research / 147
International Flow of Abstracts for Highway Research Literature / 148

HRIS—System Overview / 149
HRIS—Document Flow / 150
HRIS—Development of Thesaurus and Index Terms / 151
HRIS—Coverage and Selection of Highway Research Information / 152
HRIS—Entry of User Profiles and Special Requests / 153
HRIS—Retrieval of Accession Data / 154
HRIS—Retrieval and Assembly of Information / 155
HRIS—Dissemination, Reply, Evaluation and Modification / 156
HRIS—Entry of Document Information / 157
HRIS—Storage and Listing of Document Records and Retrieval Data / 158

TRAIS—Acquisition and Storage Process / 165
TRAIS—Retrieval Process / 166

U.S. ATOMIC ENERGY COMMISSION

AEC/DTIE—System Overview / 172
AEC/DTIE—Descriptive Cataloging Keyboarding Process / 173
AEC/DTIE—Descriptive Cataloging Input and Publishing Process / 174
AEC/DTIE—Abstracting, Indexing and Keywording Process / 175
AEC/DTIE—Abstract Input and Publishing Process / 176
AEC/DTIE—Publication and Announcement Process / 177
AEC/DTIE—Reference and Query Response Process / 178

NDIC—System Overview / 181
NDIC—Input Process / 182
NDIC—Publication Process / 183
NDIC—Request Process / 184

THE LIBRARY OF CONGRESS

MARC—Editorial Office Workflow / 189
MARC—System Overview / 190

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

NASA—System Overview / 194

SMITHSONIAN INSTITUTION

CAP—Pattern of Operation / 198
CAP—System Overview / 198
CAP—Core Sequence / 199
CAP—Outline of Catalog Division Operations / 200
CAP—Outline of Minifile Division Operations / 201
CAP—Outline of Graphics Division Operations / 202

NHIR—Input Process / 207
NHIR—Query Response Process / 208

SIE—Role in the National Information Network / 213
SIE—Input Process / 214
SIE—Output Process / 215

U.S. Department of Agriculture
National Agricultural Library
Beltsville, Maryland 20705

(301) 345-6200

System Manager: VERN J. VAN DYKE, Chief, Computer Applications, National Agricultural Library

For further information, contact:

DEPUTY DIRECTOR FOR RESOURCE DEVELOPMENT, Ext. 204 (all NAL source data)
COMPUTER APPLICATIONS, Ext. 322 (sale of CAIN data base on magnetic tape)

BACKGROUND

The Cataloging and Indexing System (CAIN), an expansion of the system previously used for the Pesticides Information Center, began operation in January 1970.

GENERAL SYSTEM DESCRIPTION

Major Components and Services

CAIN validates transactions; updates files; generates catalog cards, bibliographies, book catalogs, and a New Book Shelf pamphlet; and searches its data base in response to inquiries.

Scope

The CAIN data base contains literature pertaining to agriculture and associated subjects, gathered on a worldwide basis. Major subject fields include agriculture (general), agricultural economics and rural sociology, agricultural products (economics and technology), animal science, chemistry, engineering, entomology, food, human nutrition, home economics, forestry, life sciences, natural resources, pesticides, physical sciences, plant science, social sciences, soils and fertilizers, water resources, and other miscellaneous fields.

Users/Audience

The present users, by respective data base, are as follows: Data Base I (source data) includes NAL Cataloging and Indexing, NAL New Book Section, NAL Bibliographic Section (previously *Bibliography of Agricul-*

ture), and Agricultural Economics Center; Data Base II (herbicides) consists of the Herbicides Information Center (Agricultural Research Service); Data Base III (tree diseases) consists of INTREDIS: International Tree Disease Register (Forest Service). Control of input is the responsibility of each user. The Computer Applications Staff of the Library coordinates user needs and controls all runs.

Staff

Twenty-eight professional and 19 nonprofessional employees perform the internal production functions. Of these 47 employees, Acquisitions has 10, Cataloging has 25, and Indexing and Vocabulary, 12. The system also requires 2 computer systems analysts and 4 programmers for maintenance and support.

Equipment

The traditional keypunching of input data may be replaced by CRT-to-magnetic tape using 16 CRT's. The data will be entered by the operator and temporarily stored. A reviewer will recall the data from storage, revise it if necessary, and transmit the correct transaction to magnetic tape. This is expected to improve throughput time by several weeks. Although the system is run on an IBM 360/65 (os) with 2314's, it can also run on a model 40 having 265k and six tape drives.

Relationship to Other Systems

Magnetic tape records in the MARC internal format are created for entry into the National Serials Project on all serials newly cataloged or recataloged by NAL since January 1970.

PRESENT OPERATIONS

Acquisitions

Resources Although the National Agricultural Library has an enormous permanent collection in many physical forms, CAIN contains only items cataloged since January 1970 and selected journal articles. From January to August 1970, these items totaled 66,467. All data from 1967–1969 published in the *Pesticides Documentation Bulletin* were converted to the new format and used as a base.

Growth Approximately 8,000 citations are added each month. The number is expected to reach 10,000 to 12,000 a month. Items added to CAIN are not deleted from the file.

Description and Analysis

Cataloging Data elements in the master file include the following: identification number (unique by year); file code; subject category (two allowed); English indicator (text); translation available indicator; language of text (standard 3-character codes); restrictor code; date of document (numeric date for searching plus printed form of date); date record was last revised; identification of main entry if not author; title tracing code for catalog cards; document type (monograph, serial, or journal article); user codes (up to 5 per data base); source of document within USDA (originating agency); title (vernacular title, alternate title, and/or translated title); personal authors (up to 10); corporate authors (up to 2); biographical data of major author; title of journal in which article appears; imprint (monographs); pagination or collation; call number; citation number by user (if published in a bibliography or book catalog); subject terms (single or nested; up to 50 allowed); notes; report, patent, grant, contract, or analysis number (up to 2); series statement; abstract or extract; additional tracings beyond those generated by the foregoing data (up to 3; normally limited to related works or noncommercial publisher); nonvocabulary cross-references not verified against the controlled vocabulary (up to 2 sets). Individual records from the master file may be retrieved on any one or a combination of the foregoing data elements. All elements are printed on the linear proof output listing.

In addition to the master file, a controlled vocabulary contains the following data elements: generic term; specific term; type of relationship (scope note, use, use for, narrower term, broader term, related term, journal title/call number, call number/journal title); postability code; category in which term can be used (up to 10);

user code (which users consider the term part of their vocabulary); nonreversible code (if this code is not set, the system automatically generates a reverse action for every transaction, e.g., "Cotton, use for gossypium" is also reversed to "Gossypium, use cotton"); date term was approved for use; and unique number for generic term. This controlled vocabulary file is used on searches and updates as a control, but no individual record as such is retrievable.

Classifying and Coding The National Agricultural Library uses the Library of Congress classification system. Serials and monographs cataloged before 1966 still retain the old NAL classification.

Indexing The controlled vocabulary is based on the *Agricultural/Biological Vocabulary* developed by NAL. Approximately 20,000 terms are currently used, with 300 to 400 terms added each year.

CAIN can handle basic terms, scope notes, narrower terms, broader terms, related terms, use, and use for terms. However, at present the system includes only basic terms which can be used as single terms or as subheadings.

The average number of subject index terms per item is currently 2.5, including subheadings. The low number is based on the finding that the title of a document normally contains most of the descriptors necessary. On journal articles, therefore, title enrichment is used rather than specific subject terms. Cataloged items carry the full complement of subject headings.

Publications include the following sections: bibliography (by subject category and main entry); personal author index; corporate author index; biographical index; title index; subject index; and taxonomic index. The provision for a taxonomic index in the system is not currently used.

Abstracting NAL staff does not perform abstracting services.

Storage and Retrieval

Searching Simple searches or searches using Boolean parameters are possible on any one or a combination of data elements in the master record. The search subsystem can handle up to 200 descriptors at one time. A descriptor of more than 60 characters is counted as two descriptors. Weighting of elements is allowed.

User Products and Services

Requests for Information Search requests are currently processed only to answer internal Library requirements.

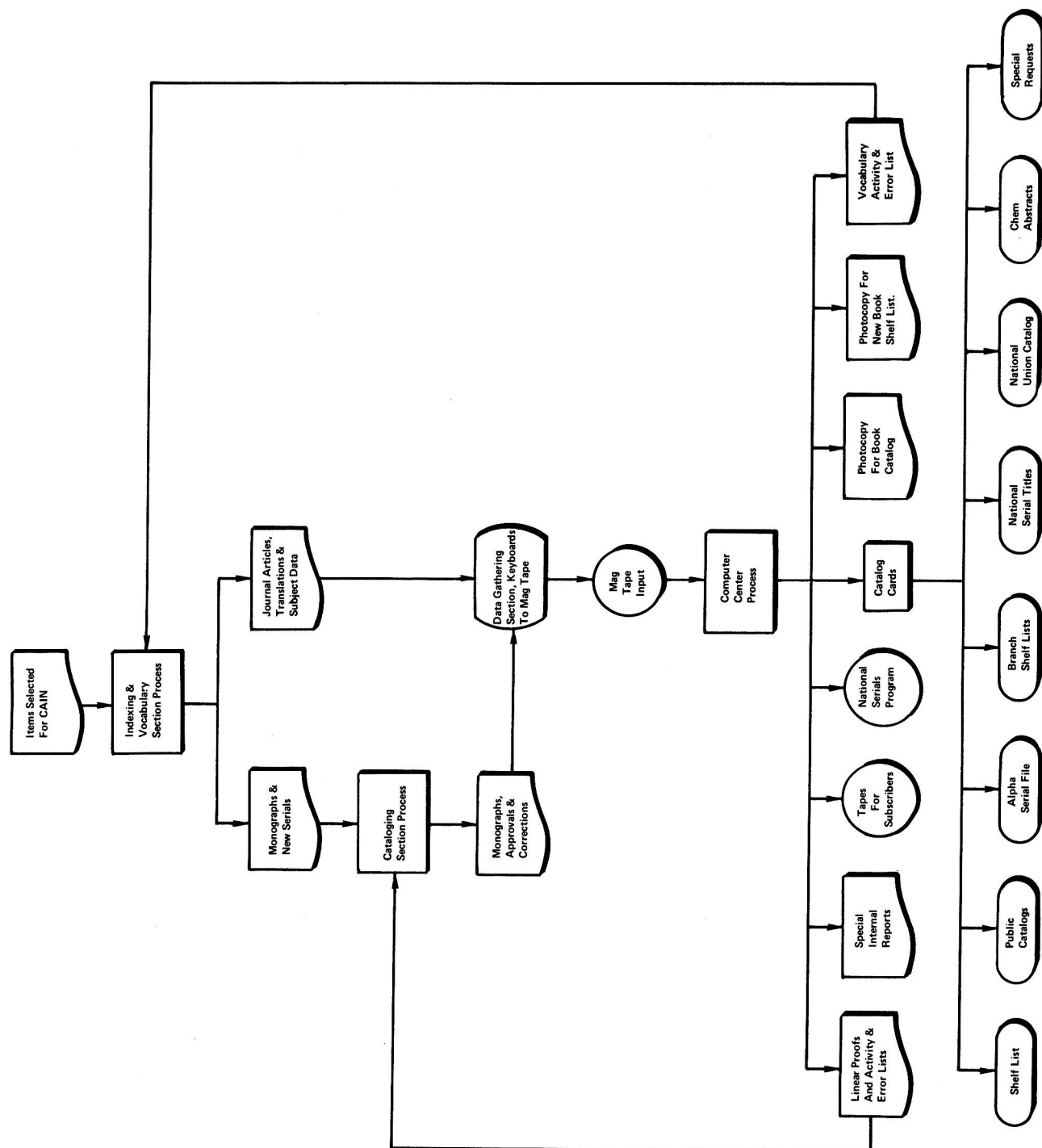
Publications All publications from the CAIN data base are printed in upper and lower case. Publications can be prepared either on photocopy paper by a computer printer or on magnetic tape in Linotron format.

The *National Agricultural Library Catalog* includes all newly cataloged items for the Library, listed by main entry, and with indexes for personal and corporate authors, titles, and subjects. It is prepared on photocopy paper. Printing and distribution are handled by Rowman and Littlefield, Inc., New York, N.Y. The annual subscription rate is \$72.

The *Bibliography of Agriculture* is produced by CCM

Information Corporation from magnetic tapes containing bibliographical data prepared by the Library. The subscription rate is \$85 per year.

Other Services The CAIN data base on magnetic tape is sold at \$60.00 for the first reel and \$47.50 for additional reels created at the same time. The cumulative data base and monthly additions of approximately 8,000 items are available. This tape is 9-track, 800 bpi, with variable length records, blocked 2, in EBCDIC with standard IBM/360 labels.



CAIN—SYSTEM OVERVIEW

U.S. Department of Agriculture
Science and Education Staff
South Building, Room 6818
Washington, D.C. 20250

(202) 388-2480

System Manager: JOHN R. MYERS, Director 388-7273

BACKGROUND

Increasing pressure from government, industry, and scientists for better and more complete agricultural research information led the U.S. Department of Agriculture (USDA) to establish a task force in July 1964, to investigate the feasibility of developing an automated system for the storage and retrieval of this research information. As a direct result of task force recommendations, the Current Research Information System (CRIS) was set up in April 1966, and became operational July 1, 1969.

GENERAL SYSTEM DESCRIPTION

Objectives

CRIS is designed to provide ready access to information regarding the research activities of the USDA, the state agricultural experiment stations, and other cooperating institutions by (1) improving communications among agricultural research scientists, especially with regard to current research work, and (2) providing research managers with up-to-date and coordinated information on the total research programs of these agencies.

Scope

The information base for CRIS derives from the descriptions of research efforts sponsored by 6 USDA agencies (Agricultural Research Service, Forest Service, Economic Research Service, Cooperative State Research Service, Farmer Cooperative Service, Statistical Reporting Service), 53 state agricultural experiment stations, and 25 other cooperating state institutions.

The basic unit for documentation and reporting is the Research Work Unit, i.e., a research activity at a single location which focuses on a clearly definable problem, a sizable phase of a very big problem, or a few closely related elements of interest which together form a logical and manageable package.

Users/Audience

The eligible users of the system are all researchers or research managers in the total agricultural community previously mentioned.

Staff

The CRIS staff totals 17, including 1 biologist, 1 systems analyst, 2 programmers, and 2 technical information specialists.

Equipment

The basic equipment of the system is the IBM Model 40 computer. CRIS is currently being run at the USDA's Washington Data Processing Center. The system itself does not own or lease any computer equipment.

Relationship to Other Systems

CRIS is not part of any larger information system. Current interaction with other programs is limited, for the most part, to providing the Science Information Exchange (SIE) with tapes of updated information for its information file of agricultural research projects. In addition, the Highway Research Information Service (HRIS) is supplied annually with pertinent printouts of CRIS project descriptions.

PRESENT OPERATIONS

Acquisitions

Resources The Research Work Unit is the basic data unit collected in CRIS. Approximately 28,000 project descriptions are currently in CRIS. The total description for each current unit consists of a research résumé, a classification report, a funds and manpower report, and a progress report. The two latter reports are submitted annually, while new résumés and classifications are submitted for revised projects.

Growth A Research Work Unit project description is entered into CRIS at the time the project is initiated, and it remains in the active file, unless revised or discontinued, for two years beyond its termination date. Each year, approximately 4,000 new Research Work Units are initiated, and the same number retired from the file. The average active life of an individual unit is five years.

Description and Analysis

Cataloging The data elements entered into CRIS are taken from the four types of forms included in the Research Work Unit project description. Virtually all data elements (almost 100) on the forms are accessible for retrieval purposes.

Classifying and Coding All input documents are classified and coded in accordance with the Department-wide classification scheme. Categories include: type of research (basic, applied, developmental); activity or general purposes; general and specialized commodities; field of science or discipline; research problem area; PPBS code; special-interest categories such as pollution, weather, medical, or health-related; and percentage of effort in each category.

Indexing A CRIS authority file has been developed from an initial accumulation of keywords provided by the investigator. These are generally in conformance with the style and form of terms in NAL's *Agricultural/Biological Vocabulary*. Other authorities used in connection with the CRIS program include the *Thesaurus of Engineering and Scientific Terms* (DOD and EJC), *Water Resources Thesaurus*, the National Library of Medicine's *Medical Subject Headings* (MESH), *Thesaurus of ERIC Descriptors*, *Medical and Health Related Sciences Thesaurus*, and the *NASA Thesaurus*. Additional subject-matter guidance and consultation are available from standard and specialized dictionaries, glossaries, and word lists in the areas of the life and

physical sciences, as well as from within the Department itself.

At present, the authority file is an alphabetical keyword list with no cross-reference or generic-specific hierarchies. There are approximately 11,000 keywords in the system, with new terms being added at the rate of about 1,000 per year. Concepts, either inherent or implied in the classification, are used for generating or modifying index terms. An average of about eight subject index terms are assigned per item. The authority file undergoes periodic revision to achieve higher levels of retrieval efficiency and user satisfaction and greater compatibility with the NAL *Vocabulary*.

Storage and Retrieval

Storage Forms and Files CRIS is a batch processing system. All recorded data elements are stored on magnetic tape. The documents themselves, i.e., the Research Work Unit forms, are maintained by an in-house accession number, randomly assigned. The keyword index, a primary tool for accession to the data base, is maintained in inverted format on a disk pack for more efficient processing.

Searching The average number of indexing variables per query is 20. The Boolean selection logic in the search system allows for the use of the connectors AND and OR either alone or in conjunction with NOT. Each data element can be compared with another or with a constant to establish an equal, less than, or greater than relationship. CRIS utilizes a combination of coordinate indexing, subject classification, and coding for the technical retrieval subsystem.

The total number of search requests that can be processed during one pass of the query system is determined primarily by the complexity of the search logic. Approximately 20 queries of average complexity can be processed at one time. A flexible schedule for query processing usually makes it possible to process the maximum number in one pass.

User Products and Services

Requests for Information The information retrieval services performed are based on programs designed to retrieve, accumulate, arithmetically manipulate, sort, and print the recorded data elements in variable formats. The major output is the Standard Technical Retrieval (Form AD 357). This format includes basic fields of project identification (e.g., accession number, responsible agency, investigators, contract or grant number), as well as descriptive textual fields such as title, objective, keywords, and publications. The other significant

output of CRIS varies according to the needs of the requester. Custom reports can be developed which, by means of extensive formatting capabilities in the retrieval subsystem, provide for any number of combinations of data elements in a variety of formats.

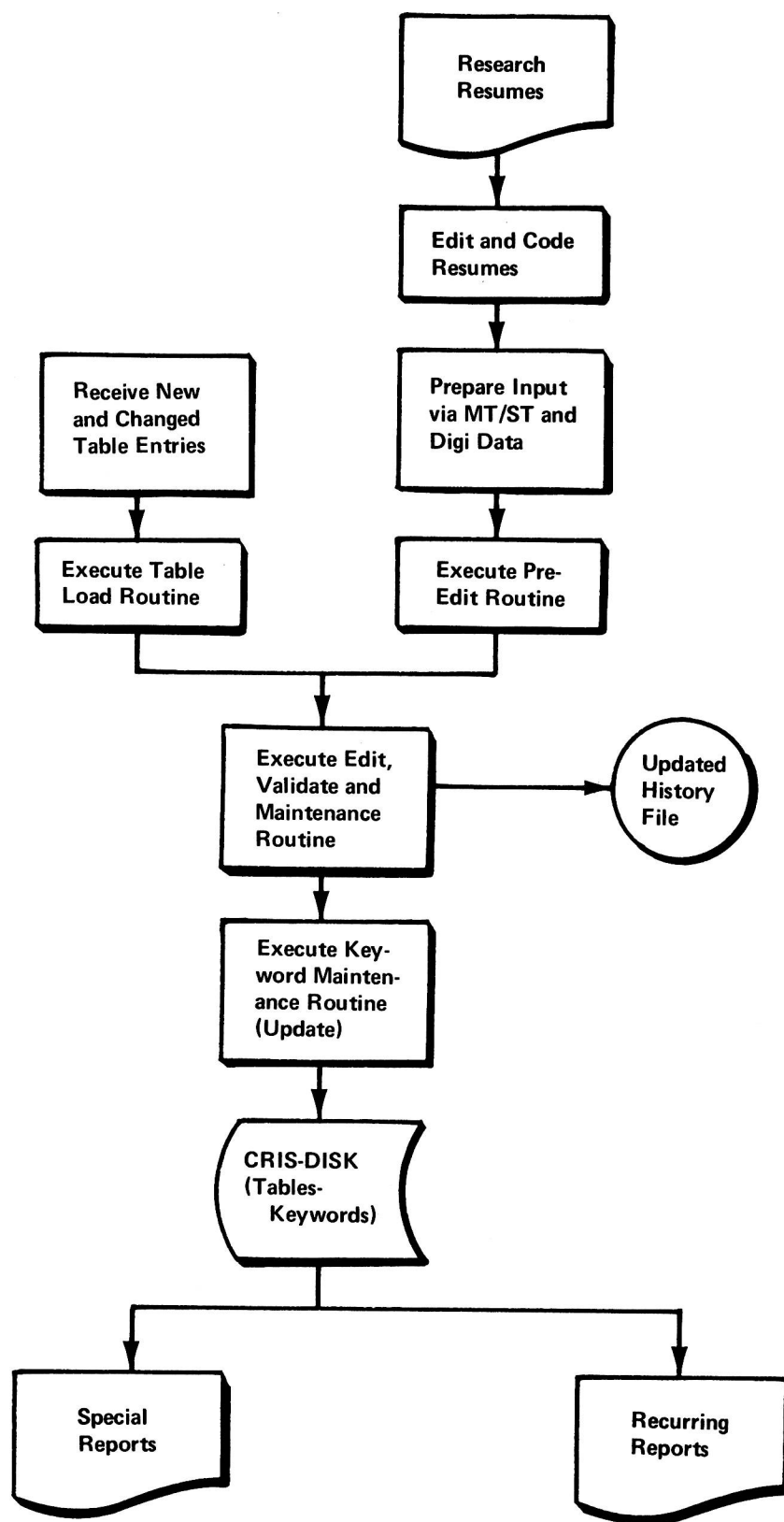
The average time between receipt of a query and delivery is two weeks, with many requests being processed in less time. CRIS is now processing about 100 information requests per month. At the present time, there is no cost to users since CRIS is currently being financed by assessments of all USDA research agencies.

Evaluation and Testing

All CRIS printouts are transmitted to users with an appropriate form letter requesting appraisal and comments on the information supplied. A User Appraisal form and a self-addressed return envelope are included with the letter.

FUTURE PLANS

Indefinite at this time.



CRIS—SYSTEM OVERVIEW