

Lawrence Berkeley National Laboratory

Recent Work

Title

ERDA INTERLABORATORY WORKING GROUP FOR DATA EXCHANGE (IWGDE)

Permalink

<https://escholarship.org/uc/item/7jz9r9qb>

Author

Merrill, Deane

Publication Date

1976-09-01

ERDA INTERLABORATORY WORKING GROUP FOR DATA EXCHANGE (IWGDE)

c. 1

*Edited by
Deane Merrill and Donald Austin*

September 1976

RECEIVED
LAWRENCE
BERKELEY LABORATORY

JAN 11 1977

LIBRARY AND
DOCUMENTS SECTION

For Reference

Not to be taken from this room



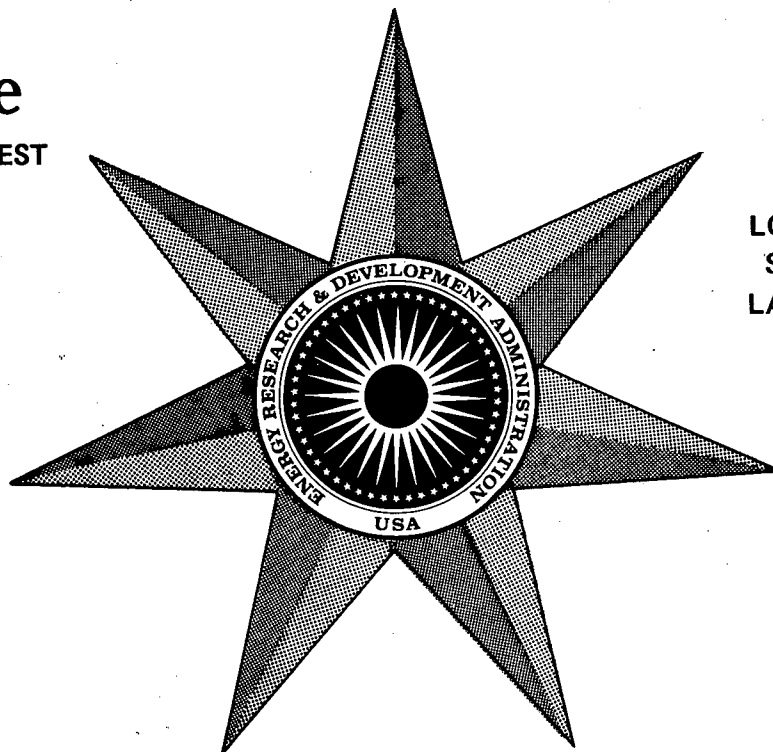
LAWRENCE
BERKELEY
LABORATORY



LOS ALAMOS
SCIENTIFIC
LABORATORY



Battelle
PACIFIC NORTHWEST
LABORATORY



BROOKHAVEN
NATIONAL
LABORATORY



OAK RIDGE
NATIONAL
LABORATORY



ARGONNE
NATIONAL LABORATORY



LAWRENCE
LIVERMORE
LABORATORY

c. 1

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal 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. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

LBL-5329

ERDA INTERLABORATORY WORKING
GROUP FOR DATA EXCHANGE (IWGDE)
ANNUAL REPORT FOR FISCAL YEAR 1976

Argonne National Laboratory
Brookhaven National Laboratory
Lawrence Berkeley Laboratory
Lawrence Livermore Laboratory
Los Alamos Scientific Laboratory
Oak Ridge National Laboratory
Pacific Northwest Laboratory

Edited by

Deane Merrill and Don Austin
Computer Science and Applied Mathematics
Department
Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720

September 9, 1976
(Revised)

ABSTRACT

In this report the activities of the ERDA IWGDE (Interlaboratory Working Group for Data Exchange) during fiscal year 1976 are discussed. The IWGDE was created in response to an evident need for increased sharing of resources among ERDA installations. During fiscal year 1976 it was supported through funds allocated to the Energy Analysis and Assessment programs of seven multipurpose ERDA national laboratories. Principal accomplishments included the interlaboratory exchange of socio-economic, environmental, demographic, and energy-related data bases, liaison development through interlaboratory meetings, creation of a national index of energy-related models and data bases, and definition and partial implementation of a computer-independent standard for exchange of data via magnetic tape.

This work was done with support from the U.S.
Energy Research and Development Administration.

FOREWORD

This report is stored in the Program Storage System (PSS) of the Lawrence Berkeley Laboratory computing facility. To get a copy of the current version (including any or all of its appendices and some of its computer-stored references) do the following.

Call Don Austin at FTS-451-5313 if you do not presently have a valid LBL computer account number. Suppose your account number, password, and name are 999999, MOTCLEF, and JOHNDOE respectively. Set your terminal to 30 char/sec, full duplex, even parity. Dial (FTS)-451-5752. Type the following commands, ending each line with 'return':

```
>LOG, IWGDE.999999, JOHNDOE
MOTCLEF (omit for non-password accounts)
^LOAD, GETFY76, IWGDE
^RUN
```

and await further instructions. Ref. LBL.27, which contains detailed citations not included in this report, is one of the computer-stored references that may be so obtained.

Alternatively, persons at LBL may obtain a microfiche listing and a hardcopy 8x11 listing of the report including appendices A,B,C,D by typing at a terminal:

```
>LOG, IWGDE.999999, JOHNDOE
MOTCLEF (omit for non-password accounts)
>Y (optional, to see what's happening)
^LOAD, PRTFY76, IWGDE
^RUN
>WIPE (to sign off)
```

or by submitting the following deck:

```
IWGDE.999999, JOHNDOE
$ID=MOTCLEF (omit for non-password
accounts)
*B, PSS
LIBCOPY, IWGDE, CONTROL/XR, PRTFY76.
(end-of-job card)
```

In either case, the 8x11 hardcopy listing and microfiche will be found at the LBL Computer Center (first floor, building 50B). The listing will be filed under your account number 999999, and the microfiche under your name JOHNDOE.

To get the current IWGDE status report, log in to LBL's interactive command processor SESAME, by typing the first two lines in either of the first two examples above. Then type

```
^CLEAR
^LOAD,NEWS,IWGDE
^RUN
```

and await further instructions.

To send a message to Don Austin, log in to SESAME and type

```
^CLEAR
^LOAD,ANSWER,IWGDE
^RUN
```

and await further instructions.

To send a message to Deane Merrill, log in to SESAME and type

```
^CLEAR
^LOAD,EXEC,DWMMAIL
^RUN
```

and await further instructions.

For lengthy messages, transmission via this latter method is preferable as a machine-readable version of the message is preserved. If so requested, Deane Merrill will forward copies of messages received by him to Don Austin and/or other IWGDE members.

TABLE OF CONTENTS 09 SEP 76

| | |
|--|----|
| 1. IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 | 1 |
| 1.1 INTRODUCTION | 1 |
| ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION | 2 |
| ERDA ANALYSIS AND ASSESSMENT PROGRAM | 2 |
| INTERLABORATORY WORKING GROUP FOR DATA EXCHANGE | 3 |
| 1.2 ACCOMPLISHMENTS DURING FY 1976 | 5 |
| INTERLABORATORY DATA EXCHANGES | 5 |
| INTERACTIVE DATA RETRIEVAL | 6 |
| COMPUTER COMMUNICATIONS | 7 |
| NATIONAL INDEX OF DATA BASES AND MODELS | 8 |
| IWGDE EXCHANGE STANDARD | 9 |
| HIGHER LEVEL EXTENSIONS TO THE STANDARD | 10 |
| IMPLEMENTATION OF IWGDE EXCHANGE STANDARD | 11 |
| GEOGRAPHIC INTERCHANGE STANDARDS COMMITTEE | 11 |
| DATA BASE MANAGEMENT SYSTEMS | 13 |
| COMPUTER NETWORKS | 15 |
| 1.3 FUTURE PLANS | 16 |
| IMPLEMENTATION OF IWGDE EXCHANGE STANDARD | 16 |
| INTERLABORATORY DATA | 16 |

TABLE OF CONTENTS 09 SEP 76

| | |
|---|----|
| EXCHANGES | |
| DATA BASE MANAGEMENT SYSTEMS | 18 |
| COMPUTER NETWORKS | 19 |
| ON-LINE COMPUTER DOCUMENTATION | 19 |
| 1.4 SUMMARY AND CONCLUSIONS | 21 |
| 1.5 REFERENCES | 23 |
| ERDA HEADQUARTERS | 23 |
| ARGONNE NATIONAL LABORATORY (ANL) | 24 |
| BROOKHAVEN NATIONAL LABORATORY (BNL) | 25 |
| LOS ALAMOS SCIENTIFIC LABORATORY (LASL) | 26 |
| LAWRENCE BERKELEY LABORATORY (LBL) | 27 |
| LAWRENCE LIVERMORE LABORATORY (LLI) | 29 |
| OAK RIDGE NATIONAL LABORATORY (ORNL) | 30 |
| PACIFIC NORTHWEST LABORATORY (PNL) | 31 |
| NON-ERDA INSTITUTIONS | 32 |

APPENDICES

- A. ERDA COMPUTING FACILITIES
- B. IWGDE ADDRESSES AND PHONE NUMBERS
- C. INTERLABORATORY DATA EXCHANGES
- D. IWGDE DATA EXCHANGE STANDARD

1.1 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 1
INTRODUCTION

1.1 INTRODUCTION

In this report the activities of the ERDA IWGDE (Interlaboratory Working Group for Data Exchange) during fiscal year 1976 are discussed. The IWGDE was created in response to an evident need for increased sharing of resources among ERDA installations. During fiscal year 1976 it was supported through funds allocated to the Energy Analysis and Assessment programs of seven multipurpose ERDA national laboratories. Principal accomplishments included the interlaboratory exchange of socio-economic, environmental, demographic, and energy-related data bases, liaison development through interlaboratory meetings, creation of a national index of energy-related models and data bases, and definition and partial implementation of a computer-independent standard for exchange of data via magnetic tape.

This report consists of five sections, which are separately stored and periodically updated, in the Program Storage System (PSS) of the Lawrence Berkeley Laboratory Computer Center. The five sections of the report are:

ERDA Interlaboratory Working Group for Data Exchange (IWGDE) Annual Report for Fiscal Year 1976, Report LBL-5329, about 30 pages;

(Appendix A) Hardware and Software Capabilities of the Laboratories Involved in the Interlaboratory Working Group on Data Exchange, about 25 pages;

(Appendix B) IWGDE Mailing List, about 10 pages;

(Appendix C) Data Exchanges Between ERDA National Laboratories, Report UCID-3828, about 15 pages;

(Appendix D) Draft Proposal, An Extendable Standard for Information Interchange on Magnetic Tape, about 40 pages.

Instructions for obtaining a copy of the full report, including all appendices, are provided

1.1 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 2
INTRODUCTION

in the foreword of this report.

ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION

In 1975, the former U.S. Atomic Energy Commission became part of the U.S. Energy Research and Development Administration (ERDA), whose broader goal is to coordinate U.S. energy R+D, both nuclear and non-nuclear.

The ERDA-supported national laboratories, long noted for their expertise in nuclear science, have expanded their horizons to consider energy resource shortages, socioeconomic impacts of energy alternatives, advanced energy supply technologies, energy conservation measures, power plant siting criteria, and other related questions.

ERDA is not studying these questions alone, but in close cooperation with legislators, other government agencies, private industry, and the public. ERDA has the responsibility of using its own resources, and those outside ERDA, to the best possible advantage, avoiding unnecessary duplication of expertise or research facilities. With this end in mind, a comprehensive long-range plan, considering all phases of U.S. energy RD+D (research, development, and demonstration), has been developed. (Ref. ERDA.1).

ERDA ANALYSIS AND ASSESSMENT PROGRAM

Certain energy-related questions, such as decisions relating to power plant siting or large-scale mining operations, have profound socio-economic impacts on particular regions of the U.S. Such questions are being studied in ERDA's Analysis and Assessment Program, which includes programs formerly known as Regional Studies Programs. This program is administered by DBER, the Division of Biomedical and Environmental Research, within ERDA's

1.1 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 3
INTRODUCTION

Administration for Environment and Safety.

Beginning in FY 1976, parallel Analysis and Assessment (Regional Studies) programs were initiated at each of seven multipurpose ERDA National Laboratories, namely; Argonne National Laboratory (ANL), Brookhaven National Laboratory (BNL), Los Alamos Scientific Laboratory (LASL), Lawrence Berkeley Laboratory (LBL), Lawrence Livermore Laboratory (LLL), Oak Ridge National Laboratory (ORNL), and Pacific Northwest Laboratory (PNL). These laboratories, taken together, possess an awesome array of scientific expertise and computing capability unequalled in any other government agency. The computer resources of each laboratory are summarized in Appendix A.

The ERDA Analysis and Assessment Program encompasses a variety of studies which are too numerous to list here. For further details, refer to recent annual reports and Form 189 proposals of the individual laboratories. (See, for example, Refs. ERDA.2, ANL.1, ANL.2, BNL.2, LASL.1, LBL.1, LBL.3, LBL.4, ORNL.6, ORNL.8, and PNL.2). Although each laboratory has responsibility for a particular region of the U.S., research efforts are coordinated so as to achieve comparability of results at the national level.

INTERLABORATORY WORKING GROUP FOR DATA EXCHANGE

During FY 1976, a small portion (less than one full-time equivalent) of each laboratory's Analysis and Assessment funding was allocated for support of an interlaboratory group, which came to be known as the IWGDE (Interlaboratory Working Group for Data Exchange). The present membership of IWGDE, and the names of the Analysis and Assessment Program coordinators at each of the ERDA National Laboratories, are given in Appendix B. Each laboratory is represented by at least two members, representing expertise both in computer science techniques and in the specific data needs of the Analysis and Assessment Programs.

1.1 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 4
INTRODUCTION

The IWGDE evolved from an earlier group that had met informally several times during 1974 and early 1975. During this period, under separate funding, the LLL Information Research Group began the compilation of the National Index of Energy Models and Data Bases. (Refs. LLL.2, LLL.4). Preliminary draft 'Guidelines' for exchange of data bases were drawn up, also by LLL. (Ref. LLL.1). A 'Strawman' proposal for exchange of geographic data bases was proposed by PNL (Ref. PNL.3). These documents were circulated to all participating laboratories and to ERDA headquarters for modification and approval.

A substantially revised version of the 'Guidelines' was written by Gill Ringland of LBL. (Ref. LBL.18). Some guiding principles were outlined by John Wilson of ERDA Headquarters. (Ref. ERDA.3). A reply, consisting of a brief historical summary and an informal statement of purpose, was drafted by Don Austin of LBL. (Ref. LBL.19). Subsequent IWGDE activities during FY 1976 are the logical outgrowth of these early deliberations.

1.2 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 5
ACCOMPLISHMENTS DURING FY 1976

1.2 ACCOMPLISHMENTS DURING FY 1976

The primary purpose of the IWGDE is the face-to-face exchange of information, to ensure effective coordination of the Analysis and Assessment Programs at the ERDA national laboratories. Representatives from each of the laboratories met four times during FY 1976: in Albuquerque, New Mexico in July 1975; in Berkeley, California, in November 1975; in Richland, Washington in March 1976; and again in Berkeley in May 1976. (Refs. BNL.5, BNL.6, BNL.7). At each of these meetings, about 15 IWGDE members from the ERDA laboratories, plus an advisor from ERDA Headquarters, were present.

During the meetings, IWGDE members acquired a clearer understanding of the nature of the Analysis and Assessment Programs and the available resources at the ERDA laboratories. This knowledge quickly led to the exchange of energy-related and socio-economic data bases required for ongoing programs. Even more important, the IWGDE functioned as an interlaboratory liaison group, facilitating the interlaboratory transfer of specific data bases requested by Analysis and Assessment Program coordinators and others.

INTERLABORATORY DATA EXCHANGES

During FY 1976 a considerable number of socio-economic, environmental, and demographic data bases were copied to tape and exchanged between two or more ERDA national laboratories. For the most part, these exchanges were bilateral, and no attempt was made to re-format the data for the convenience of the user. A discussion of these data exchanges is provided in Appendix C.

Appendix C illustrates the high degree to which data are already being shared by the ERDA national laboratories. At the same time, it points up some of the inefficiencies which have resulted due to the absence of a data exchange standard.

1.2 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 6
ACCOMPLISHMENTS DURING FY 1976

INTERACTIVE DATA RETRIEVAL

Members of the IWGDE provided each other with courtesy guest computer account numbers at their own installations. Procedures for logging on to time-sharing systems, and for using data management systems (e.g. REAP, BDMS, SYSTEM 2000, MASTER CONTROL, ORLOOK, RECON, etc.) were explained. (Refs. LBL.25, LBL.13, NE.3, LLL.3, ORNL.3, ORNL.1). Using computer terminals at remote locations, IWGDE members successfully retrieved sample data from the following data bases, among others:

from ANL: an energy-related bibliographic data base;

Energy Supply and Demand data base;

Nuclear Power Plant data base;

from BNL: Energy Model Data Base;

an index of various air quality data bases;

NASN (National Aerometric Surveillance Network) Air Quality Data Base;

NEDS (National Emissions Data Service) Emissions Inventory Data Base;

from LASL: RMDB, an energy-related bibliographic data base (a Rocky Mountain state subset of a bibliographic data base in RECON);

from LBL: City-County Data Book;

1970 U.S. Census Fifth Count summary;

OBERS Series C and E Projections;

1.2 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 17
ACCOMPLISHMENTS DURING FY 1976

Census of Agriculture;

LBLIRI, the LBL Interactive
Resource Index;

from LLL: National Index of Energy
Models and Data Bases;

from ORNL: (in ORLOOK) several
energy-related bibliographic
and numerical data bases;

(in RECON) several
energy-related bibliographic
data bases;

from PNL: on-line index of
computer-stored files.

Most of these data bases are hierarchically stored in data base management systems, so that their content and structure can be easily understood by the user not having supplementary documentation. In all of the examples cited, data were displayed at a terminal but not actually transferred from one computer center to another. Interlaboratory exchange of computer files over computer networks is awaiting implementation of links to the ARPANET at additional ERDA facilities. (Refs. LBL.9, LBL.14, LBL.15, LBL.16, NE.1, NE.2).

COMPUTER COMMUNICATIONS

Between meetings, IWGDE members communicated with each other using computer 'mailbox' facilities, as well as by more conventional means. (As an example, the use of IWGDE mail facilities at LBL is explained in the foreword of this report.) Status reports, informal notes, and draft versions of this report were stored in the LBL PSS (Program Storage System), where they could be inspected by IWGDE members at remote terminals. Appendix A of this report, written by Dick Wiley of LASL, was transmitted directly to the LBL PSS from a cassette terminal at LASL. Draft versions of Appendix D were stored

1.2 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 8
ACCOMPLISHMENTS DURING FY 1976

in the ORNL PDP-10 computer, and read from terminals at LBL.

Recently, the technique of teleconferencing (in which participants at several locations send messages to one another via terminals connected to a central site) is being used in communications among ERDA installations, including ERDA headquarters. (Ref. NE.5).

NATIONAL INDEX OF DATA BASES AND MODELS

An important prerequisite for interlaboratory exchange of resources is an adequate catalog of energy-related models and data bases, both within and outside the ERDA national laboratories. In Spring 1975 a comprehensive nationwide survey was undertaken by the Information Research Group at LLL under the direction of Viktor Hampel. (Refs. LLL.2, LLL.4). Information obtained from the following six sources was incorporated into a single computerized data base:

Naval Oceanographic and Atmospheric Administration (NOAA): information on 3151 data bases relevant to energy and environmental research;

ORNL: a subset of the ORCHIS Information System, containing information on 82 data bases and 4437 calculational models;

LLL: descriptions of 90 environmental calculational models;

National Technical Information Service (NTIS): information on 516 data bases;

Federal Energy Administration (FEA): information on 326 energy-related data bases;

Questionnaire Survey: information on 140 data bases and 68 calculational models, obtained from questionnaires completed by the ERDA national laboratories, the

1.2 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 9
ACCOMPLISHMENTS DURING FY 1976

Environmental Protection Agency (EPA), the United States Geological Survey (USGS), the United States Bureau of Mines, the Thermonuclear Properties Research Center, the Purdue Information Center, and the Economic Sciences Corporation.

The resulting data base, known as the National Index of Energy and Environmentally Related Data Bases and Computational Models, contains bibliographic information, abstracts, and detailed descriptions of 4215 data bases and 4595 calculational models. To facilitate retrieval of selected information, author and subject indices, containing more than 10,000 and 13,000 entries respectively, have been provided. The LLL Index is presently the most comprehensive single index of energy-related computer resources available to the ERDA National Laboratories.

The Index is a hierarchical data base in the format of MASTER CONTROL, a data base management system used at LLL and LASL. It is available for remote interactive access in the LLL CDC-6600 computer (G machine). In addition, it has been installed on the ORNL ORCHIS data management system, and will be made available on ORLOOK and RECON as soon as space and indexing problems are resolved. Alternatively, the Index is available from LLL in microfiche form (15,000 pages including author and subject indices).

IWGDE EXCHANGE STANDARD

The most important substantive accomplishment of the IWGDE was the definition of an exchange standard for the interchange of data via magnetic tape. (See Appendix D). The draft of the standard included in this report is substantially the final version, although editing of the standard is not yet complete. The proposed standard is fully consistent with, and an extension of, the ANSI (American National Standards Institute) standard for exchange of bibliographic data on magnetic tape. (Ref. NE.4).

1.2 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 10 ACCOMPLISHMENTS DURING FY 1976

Preliminary discussions with the ANSI/X3L5 committee indicate that the IWGDE standard embodies much of this committee's thoughts on an exchange standard for a wide variety of devices, and may serve as a point of departure for such a standard with few changes required. A member of the IWGDE has been asked to serve on the ANSI/X3L5 committee.

The proposed Extended Format standard was also reviewed by the U.S. Department of Health, Education, and Welfare (HEW) and by the Library of Congress. The initial reaction was cautiously positive from HEW, slightly negative from the Library of Congress. Al Brooks feels that some misunderstanding may have caused the negative response.

An important feature of the IWGDE exchange standard is its generality and flexibility. An introductory record on the standard-format tape defines the format and content of all successive records, so that a general reading routine can accept any standard-format tape, with no auxiliary information being necessary. Although the standard is defined for magnetic tapes, it can be used as well for file transfer via other means, for example the ARPANET.

Because the IWGDE's funding was derived from ERDA's Analysis and Assessment Programs, the primary emphasis so far has been on the transfer of socio-economic, environmental, demographic, and energy-related data bases. Nevertheless, the proposed (level 0) exchange standard is sufficiently general to allow the transmission of arbitrary bibliographic, numerical, or geographic data.

HIGHER LEVEL EXTENSIONS TO THE STANDARD

Higher level extensions to the IWGDE exchange standard are treated as special cases within the level 0 standard. Directions for future work were clearly identified.

A proposal for handling hierarchical data

1.2 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 II ACCOMPLISHMENTS DURING FY 1976

structures was written by Dave Richards of LBL (Ref. LBL.20). Concepts from this paper have been included in Section 3.5.1 of Appendix D. Another proposal was presented by J. Nardi and J. Heller of BNL (Ref. BNL.1). This scheme can be considered as a special case of the Cartesian label (cross product) format described in Appendix D. However, certain features of the BNL proposal were considered too restrictive for inclusion within the IWGDE Level 0 standard as presently conceived. Geographic data bases, such as those exchanged in POLYVRT format (see below), can be handled within the context of the level 0 exchange standard.

IMPLEMENTATION OF IWGDE EXCHANGE STANDARD

Implementation of the IWGDE exchange standard, namely creation at each ERDA installation of the software necessary to read and write standard-format tapes, is being accomplished independently at several ERDA laboratories. As an example, Ref. LBL.7 contains the documentation of PADRE, a software package written in BLIMP language, for writing standard-format tapes at LBL. Programmers at ORNL are using PL/1 to produce similar tape reading and writing routines for their own system. At BNL, tape writing routines, still incomplete, are being written mostly in FORTRAN.

Unfortunately, IWGDE funding during FY 1976 was insufficient to complete this task at all of the laboratories. FY 1977 funding, if received, will permit (among other benefits) a uniform and orderly implementation of the exchange standard at all ERDA laboratories.

GEOGRAPHIC INTERCHANGE STANDARDS COMMITTEE

In this section we discuss the activities of the Geographic Interchange Standards Committee (GISC), whose activities closely parallel those

1.2 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 12
ACCOMPLISHMENTS DURING FY 1976

of the IWGDE. Many individuals are members of both committees, and so for reasons of convenience the GISC and IWGDE scheduled their meetings to coincide. After a time the membership and activities of the two committees overlapped so greatly that the formal distinction between the two committees was largely forgotten.

Both the IWGDE and the GISC had their formal beginnings in a two-day workshop-meeting at ERDA headquarters in Germantown in April 1975, which was attended by representatives from the ERDA national laboratories. Dr. John Wilson of ERDA's DBER was chairman of the workshop, whose theme was Coordination of Exchange of Data Bases and Models. Geocoding Standards was one of the many subjects discussed. As a result of those discussions Dr. Wilson asked Paul Dionne of PNL to head a group which was named the Geographic Interchange Standards Committee.

The GISC was formed in order to develop guidelines for data transfer using geographic coordinates. The committee began its work after the distribution of LLL's survey of ERDA energy data bases and models (Refs. LLL.2, LLL.4). The IWGDE identified appropriate staff members from the laboratories to serve on the GISC. Presently active GISC participants are listed in Appendix B.

The first GISC meeting was held in conjunction with the ERDA-wide conference on Computer Support of Environmental Science and Analysis, in Albuquerque, in July 1975. In order to initiate discussion, Dick Hill of PNL provided a 'Strawman' paper on 'Standards of Geographic Data Transfer' (Ref. PNL.3).

The second meeting of the GISC was held in Berkeley in November 1975. Dick Hill prepared 'Remarks on Strawman' and 'Cartographic Data Base Characteristics' (Ref. PNL.4). Don Austin of LBL discussed POLYVRT and GEOGRAPH, computer programs available at a cost of \$700 from the Harvard University Laboratory for Computer Graphics and Spatial Analysis. After the November GISC meeting, LBL converted a small cartographic data base (census tract boundaries

1.2 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 13 ACCOMPLISHMENTS DURING FY 1976

from the Reno, Nevada SMSA) into POLYVRT format. A tape was mailed to Dick Durfee at ORNL, who successfully reproduced the polygonal tract boundaries, and went on to produce a contour map of population densities, through a spatial interpolation of enumeration district data on the MEDX tapes.

At the third GISC meeting, in Richland in March 1976, Don Austin pointed out two serious limitations of the POLYVRT software package: (a) auxiliary data (e.g. county names, populations, etc.) cannot be easily associated with particular polygons in the cartographic data base; and (b) polygons completely internal to other polygons (e.g. contour maps) cannot be conveniently represented. Don also pointed out that data of the type for which POLYVRT is designed can be handled equally well by the proposed IWGDE Exchange Standard (Appendix D). As a result, the IWGDE decided to abandon POLYVRT as a mechanism for the exchange of cartographic data.

DATA BASE MANAGEMENT SYSTEMS

Much discussion in early IWGDE meetings centered on the selection of a data base management system (DBMS). Several of the laboratories had independently determined a need for such a system at about the same time. Some of them began development on their own because of the lack of, or the expense of, commercially available systems compatible with their computer configurations. Those laboratories needing a system during the last year found several alternatives available and realized that developing their own system from scratch was unnecessary.

Formal studies of a number of DBMS's were conducted by ANL, BNL, LASL, LBL, LLL, and PNL. The studies covered facilities provided by each DBMS, particular needs of the laboratories, availability on different hardware configurations, vendor support and documentation, and transportability. Commercial

1.2 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 14
ACCOMPLISHMENTS DURING FY 1976

systems considered included SYSTEM 2000, TOTAL, ADABAS, IMS, and ENVIR. These studies, and subsequent discussions within the IWGDE, resulted in the purchase of SYSTEM 2000 by ANL, LASL, and BNL. LBL, LLL, and ORNL felt satisfied with DBMS's already functioning at their installations. PNL, having convenient remote-job access to BNL and LBL, deferred the acquisition of a DBMS. The following DBMS's are presently available at the ERDA national laboratories:

| | |
|------|----------------------------------|
| ANL | SYSTEM 2000 |
| BNL | SYSTEM 2000, BDMS (soon) |
| LASL | SYSTEM 2000, MASTER CONTROL |
| LBL | BDMS, QWICK QUERY, STOFI, others |
| LLL | MASTER CONTROL |
| ORNL | ORCHIS, IMS, CICS |
| PNL | (remote access to LBL and BNL) |

The IWGDE also discussed transportability of the DBMS's developed by the laboratories themselves, namely; ORCHIS, MASTER CONTROL, and BDMS (Berkeley Data-base Management System). ORCHIS, developed at ORNL, is portable to other IBM 360/370 computers, but not readily portable to non-IBM equipment. Source decks are available on request for all published programs. MASTER CONTROL, developed at LLL, was successfully transferred to LASL. This transfer was easily made because LLL and LASL have the same type of computer (CDC 7600) and the same operating system (LTSS). Transfers of MASTER CONTROL to other ERDA laboratories, however, would prove considerably more difficult.

BDMS, the Berkeley Data-base Management System, was specifically designed for easy transportability. At LBL it runs on the CDC 6600 or CDC 6400. It has been successfully moved to an IBM-360 at the Rutherford High Energy Laboratory (in England) and is soon to be available on BNL's CDC 7600. A version compatible with DEC PDP's and other mini-computers is planned. The major drawback of BDMS is its developmental state. It presently lacks a report generator and other features of commercial systems. Unlike commercial systems, BDMS is available free of charge to the ERDA community and other federally

1.2 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 15
ACCOMPLISHMENTS DURING FY 1976

supported installations.

In early discussions, IWGDE members felt that interlaboratory data exchange should involve either (a) a common DBMS, or (b) programs for one-to-one translation between all the DBMS's in use. A more practical solution, however, proved to be the development of the exchange standard described in Appendix D, whereby each laboratory is responsible only for the interface between the common standard and its own DBMS's.

COMPUTER NETWORKS

Interlaboratory exchange of data via computer networks was actively discussed. Full ARPANET implementation, including user and server TELNET, and user and server FTP (File Transfer Protocol) is expected to be operational at ANL, BNL, and LBL by the end of FY 1976.

As was mentioned earlier, the IWGDE exchange standard, designed for data exchange on magnetic tapes, may be used as well for exchanges via computer networks. Since the ARPANET operates at a maximum bandwidth of 50 kilobits/second (and an average of 10-20 Kbs), exchange of data bases on the order of a few million characters may be practical. For larger data bases, however, this method will not replace exchange of magnetic tapes in the IWGDE Standard exchange format. Future research by the IWGDE will determine the feasibility of sharing manageable subsets of large data bases over computer networks.

1.3 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 16 FUTURE PLANS

1.3 FUTURE PLANS

The following projects will be pursued by the IWGDE during fiscal year 1977, if adequate funding is obtained.

IMPLEMENTATION OF IWGDE EXCHANGE STANDARD

As soon as possible, an IWGDE member at each laboratory will complete the software interface between the exchange standard and the data base management system(s) (e.g. BDMS, SYSTEM 2000, ORCHIS, MASTER CONTROL) in use at that installation. To the greatest extent possible, future interlaboratory data transfers will be performed in the format specified in Appendix D.

If, on the other hand, IWGDE funding is not obtained for fiscal year 1977, future interlaboratory exchanges of data will be the responsibility of Analysis and Assessment program coordinators, who are (in most cases) not acquainted with the technical problems involved. Furthermore, the subsequent decoding of non-standard tapes will be performed independently at each laboratory, with duplication of effort, without controls on uniformity, and without the benefits of collaborative discussion. The IWGDE recommends that sufficient funding be allocated to complete the implementation of the exchange standard at all ERDA installations as soon as possible.

INTERLABORATORY DATA EXCHANGES

The IWGDE exchange standard is not merely a computer science exercise, but a research tool urgently needed in programs already funded at the ERDA national laboratories. Although the standard may be used to exchange arbitrary data bases, computer models, and programs, its initial applications will be for the transfer of data bases required by Analysis and Assessment

1.3 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 17 FUTURE PLANS

Programs. The first exchange under the new IWGDE standard will most likely be the City-County Data Book, which is presently being converted to standard format at LBL for transmission to ORNL and BNL. (The same data have already been sent to ANL and LASL, but not in the standard exchange format.)

In addition to data exchanges already accomplished (see Appendix C), specific interlaboratory requests have already been issued for the following additional data bases. It is urgently hoped that the transfers can be accomplished under the IWGDE exchange standard, thereby greatly reducing the programming effort of the data recipient in each case.

(BNL to LBL) Energy Model Data Base
(revised)

(BNL to LBL) NIH Mortality Data

(BNL to LBL) NCHS Morbidity Data

(BNL to LBL) U.S. Weather Bureau Data

(BNL to LBL) Data on Advanced Energy
Technologies

(ANL to LBL) Energy Supply and Demand
Data Base

(ANL to LBL) Nuclear Power Plant Data

(ANL to LBL) (from Argonne Code Library)
OMCST, a model for calculating operating
and maintenance costs of thermal power
plants

(LBL to LASL) Dun and Bradstreet data,
complete (partial data were sent earlier)

(LBL to ORNL) geographic data bases

(ORNL to LBL) Power Plant Siting Model

Still other data, not yet formally requested, are urgently required for Analysis and Assessment programs already in progress. These include additional Census data, employment data,

1.3 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 18 FUTURE PLANS

energy use data, input-output tables, emission inventory (air pollution) data, land use data, LANDSAT data, bibliographic data bases, energy supply and conversion technology data, and many others.

A partial list of socio-economic, environmental, demographic, and energy-related data bases presently available at the ERDA national laboratories is provided in the LLL Index of Energy Models and Data Bases mentioned earlier (Refs. LLL.2, LLL.4). In principle, any of these data bases can be converted to the standard exchange format, for convenient use in existing data base management systems at any of the ERDA installations.

The IWGDE will serve as a coordinating group for maintaining catalogs of data bases and modeling programs of interest to Analysis and Assessment program researchers in the ERDA laboratories. This group will facilitate ERDA's data acquisition and dissemination requirements and provide a focus for interlaboratory liaison.

DATA BASE MANAGEMENT SYSTEMS

Developmental work on BDMS, the Berkeley Data Base Management System, is being actively pursued at LBL. Present efforts include extension to lower-case characters, full Boolean and selective character-string search capabilities, commands to display keyword values, report generation capabilities, and adaptations for portability to other computers.

A BDMS-oriented teleconference has been initiated in PLANET, a program operating in BBN-TENEX, a computer in Cambridge, Mass. which is on the ARPANET. Using this medium, BDMS users at LBL and elsewhere have been communicating easily to each other their questions and comments about BDMS.

1.3 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 19 FUTURE PLANS

COMPUTER NETWORKS

As computer networks are further developed, a new medium for interlaboratory cooperation and communication is becoming available. Remote teleconferencing and computerized mail systems provide facilities for collaboration among the ERDA laboratories not possible earlier. In FY 1977, the combination of dial-up star networks, the government-sponsored ARPANET, and commercial networks such as TELENET and TYMNET will link all seven ERDA laboratories participating in the IWGDE. Further research into the feasibility of sharing data bases and modeling programs is required to exploit these new facilities and provide efficient and cost-effective resource sharing.

ON-LINE COMPUTER DOCUMENTATION

Adequate documentation of available resources, and effective dissemination and utilization of that information, is one of the most necessary, and one of the most difficult tasks facing research workers in ERDA and elsewhere. Within the ERDA national laboratories, on-line documentation of resources is being expanded in LBLIRI (LBL Interactive Resource Index) (Refs. LBL.21, LBL.22); and at ORNL, in RECON and ORLOOK (Refs. ORNL.1, ORNL.3). Further development of the LLL Index (Refs. LLL.2, LLL.4) has been suspended for lack of funding; it is hoped that new funding will allow this valuable resource to be brought up to date.

Within and outside the ERDA national laboratories, there exist literally dozens (possibly hundreds) of large computerized inventories of data bases and computational models, many of them currently maintained. Five of the largest and best known, plus additional information from survey questionnaires, were combined to form the LLL Index described above. The creation of such inventories, and their compilation into a single index, is an extremely

1.3 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 20
FUTURE PLANS

costly procedure, and the resulting compilation is out of date even before it is completed.

Periodically, it is desirable to assemble information from several major sources, as has been done in the LLL Index. However, to simply duplicate and combine existing inventories is not the best solution, because the user cannot benefit from updates and corrections of the component data bases. Rather, the preferred philosophy in future indices should be to include sufficient computer-oriented information (log on parameters, computer storage locations, format descriptions, etc.) to provide direct and automatic access to a multitude of owner-maintained data bases (of course in a read-only mode and under password control).

Such is the design philosophy of LBLIRI, the LBL Interactive Resource Index. (Refs. LBL.21, LBL.22). LBLIRI is, and will remain, a relatively small data base; however, it will contain 'pointers' to data bases much larger than itself, including the ARPANET NIC query system, the LLL Index, RECON, large EPA air quality data bases, and others. Gradually the interface to these data bases will be refined, so that the user will have, through a simple command language, interactive access to major data bases all over the country.

A distant goal is the establishment of an ERDA-wide (later, nation- or world-wide) information system, so linked that a user at any installation would have immediate and automatic access to any public resource in the system. Such a system is still years in the future, but the growth of computer networks and on-line resource directories is bringing that goal closer.

1.4 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 21
SUMMARY AND CONCLUSIONS

1.4 SUMMARY AND CONCLUSIONS

The achievements of the IWGDE during FY 1976 constitute an important step toward increased resource sharing among the ERDA laboratories. Face-to-face contacts enhanced mutual understanding of the resources available at all ERDA installations, and promoted a desire for interlaboratory collaboration and exchange of resources. A data exchange standard was defined, which will greatly aid in the interlaboratory exchange of data bases and models. Partial implementation of the standard at certain laboratories has been accomplished.

Full implementation of the standard at all laboratories is outside the scope of IWGDE funding during FY 1976. FY 1977 funding has been requested, in a Form 189 proposal submitted by LBL in March 1976. Ted Albert of ERDA's Office of Environmental Information Systems has recommended that a new Form 189 be submitted by each of the laboratories. The 189's should have a common title, but the content should reflect the particular capabilities of each laboratory. The IWGDE FY 1976 annual report will also serve to bolster requests for funding in FY 1977.

If funded, IWGDE activities during FY 1977 will include:

- (a) implementation of the IWGDE exchange standard at all ERDA laboratories;
- (b) further exchanges of data for use in Analysis and Assessment and other programs;
- (c) enhancement of interlaboratory communication through further implementation of the ARPANET, through an ERDA mail facility, and through teleconferencing;
- (d) improved on-line documentation, including centralized storage of documents, on-line help facilities, and further development and integration of directories presently at LLL, ORNL, and LBL;
- (e) improved data access through

1.4 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 22
SUMMARY AND CONCLUSIONS

development and utilization of interactive data base management systems.

The tasks necessary for for the creation of a smoothly functioning ERDA-wide information system are extraordinarily detailed and complex. A 'quick-and-dirty' approach to the solution of problems is not cost-effective where huge data bases are being used repeatedly and jointly at many installations. This fact is not always fully appreciated by administrators, who are necessarily more concerned with the end-result applications of research than with the tedious intermediate steps.

A double obligation is imposed upon ERDA, if the challenging energy problems facing the U.S. are to be effectively resolved. Investigators have an obligation to avoid plunging into sterile research projects having little or no possibility of practical application. At the same time, ERDA administrators have an obligation to continue to fund research (like IWGDE's) whose full benefits transcend individual programs and become evident only with the passage of time.

1.5 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 23
REFERENCES

1.5 REFERENCES

Documents relevant to IWGDE activities are listed here, arranged by division (or department) within each institution. Four of these references form appendices of this report; most of the others are directly cited in this report or in one of its appendices. Many are stored on-line in the LBL or ORNL computer system.

Because of the large number of references, detailed citations including authors, title, computer storage location, etc. are not included here. This information is provided in Ref. LBL.27, which may be requested from the authors or obtained directly at a computer terminal. To obtain Ref. LBL.27 at a terminal, consult the foreword of this report.

ERDA HEADQUARTERS

ERDA.1 A National Plan for
Energy Research

Administration for Environment and Safety

Division of Biomedical and Environmental
Research

Analysis and Assessment Programs

ERDA.2 Analysis and Assessment
Program draft

ERDA.3 Review of draft
'Guidelines' and Geocoding
'Strawman'

ARGONNE NATIONAL LABORATORY (ANL)

Energy and Environmental Systems Division

ANL.1 Division Report, 6/75

ANL.2 FY75 189 for Energy
Related Regional Studies Program

ANL.3 An Evaluation and Recommendation
of a Data Base Management System for
ANL and Fermilaboratory

1.5 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 25
REFERENCES

BROOKHAVEN NATIONAL LABORATORY (BNL)

Applied Mathematics Department

BNL.1 Heller-Nardi proposal

Department of Applied Science

BNL.2 FY75 Annual Report

National Center for Analysis of Energy
Systems

BNL.3 Memo about formation of
National Center

Atmospheric Sciences Division

BNL.4 Summary of BNL data bases

BNL.5 IWGDE minutes, LBL 11/75

BNL.6 IWGDE minutes, PNL 3/76

BNL.7 IWGDE minutes, LBL 5/76

Energy Data and Models Group

BNL.8 EMDB Users' Manual

Biomedical and Environmental
Assessment Group

BNL.9 U.S. Energetics Atlas

BNL.10 Recommendation for a DBMS for
the RESP

LOS ALAMOS SCIENTIFIC LABORATORY (LASL)

Analysis and Assessment Program

LASL.1 FY76 189's 2/75

Computer Science and Services Division

Applications Support and Research Group

LASL.2 ERDA Computing Facilities
(Appendix A)

1.5 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 27
REFERENCES

LAWRENCE BERKELEY LABORATORY (LBL)

Energy and Environment Division

LBL.1 FY74 Annual Report

Regional Energy Analysis Program

LBL.2 I/O Tables for California
and Rocky Mountain States

LBL.3 FY76 189, 3/75

LBL.4 Energy Research Program
(1975)

LBL.5 California Nuclear
Moratorium Study

Energy Use and Conservation Group

LBL.6 Electrical Energy
Consumption in California

Physics, Computer Science and Applied
Mathematics Division

Computer Center

LBL.7 PADRE writeup

Computer Science and Applied Mathematics
Department

LBL.8 Annual Report, FY75

LBL.9 Annual Report, FY76

LBL.10 SEEDIS Overview

LBL.11 SEEDIS References

LBL.12 Summary of Data Bases

LBL.13 BDMS Users' Manual

LBL.14 Report on Networking
Experiments

1.5 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 28
REFERENCES

LBL.15 Findings of Network
Users' Committee

LBL.16 Collaborative Exchange 189

LBL.17 IWGDE mailing list
(Appendix B)

LBL.18 Gill Ringland's revision
to ERDA/DBER Guidelines

LBL.19 Don Austin's letter to
John Wilson, 11/10/75

LBL.20 Dave Richards' paper about
hierarchical standard

LBL.21 Descriptive Summary of
LBLIRI

LBL.22 LBLIRI Users' Guide

LBL.23 Estimates of 1972 U.S.
Employment, by Input-Output
Industry Sector

LBL.24 CARTE Users' Guide

LBL.25 REAP Users' Guide

LBL.26 Summary of 1972 U.S.
Employment Estimates

LBL.27 Detailed IWGDE references,
from LBLIRI

LBL.28 Data Exchanges between
ERDA National Laboratories
(Appendix C)

LBL.29 IWGDE FY76 Annual Report
(this report)

LBL.30 MAPEDIT Users' Guide

LBL.31 Descriptive Summary of
CARTE

0 0 0 0 4 6 0 0 8 6 7

1.5 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 29
REFERENCES

LAWRENCE LIVERMORE LABORATORY (LLL)

Computer Research Division

Computation Department

Information Research Group

LLL.1 ERDA/DBER Guidelines

LLL.2 National Index of Data
Bases and Computational Models

LLL.3 MASTER CONTROL Users'
Manual

LLL.4 Summary Report on National
Index

1.5 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 30
REFERENCES

OAK RIDGE NATIONAL LABORATORY (ORNL)

ORNL.1 RECON Users' Manual

Computer Sciences Division

Computer Applications Department

ORNL.2 IWGDE Exchange Standard
(Appendix D)

ORNL.3 ORLOOK User's Manual

ORNL.4 ORCHIS Overview

ORNL.5 List of ORCHIS data bases

Energy Division

ORNL.6 Annual Report, period
ending 12/31/75

ORNL.7 ORRMIS Report, Part I

Analysis and Assessment Program

ORNL.8 FY76 189 to ERDA/DBER

Regional and Urban Studies Department

ORNL.9 Summary of RUSTIC Data
Bases

1.5 IWGDE ANNUAL REPORT FOR FISCAL YEAR 1976 31.
REFERENCES

PACIFIC NORTHWEST LABORATORY (PNL)

PNL.1 Letter to ERDA Headquarters
about PNL data bases

Energy Related Regional Assessment Program

PNL.2 FY76 189's to ERDA/DBER,
4/75

Systems Department

PNL.3 'Strawman' Standard for
Geographic Data Exchange

PNL.4 Remarks on Strawman

PNL.5 GISC minutes, LASL 7/75

PNL.6 GISC minutes, LBL 11/75

NON-ERDA INSTITUTIONS

Advanced Research Projects Agency

NE.1 ARPANET Directory

NE.2 ARPANET Resource Handbook

MRI Systems Corporation

NE.3 SYSTEM 2000 Users' Manual

American National Standards Institute

NE.4 ANSI Standard for
Bibliographic Data Exchange

Institute for the Future

NE.5 PLANET-2 User Manual

Teknekron, Inc.

Energy and Environmental Engineering
Division

NE.6 Data on New Energy
Technologies

0 0 0 0 4 6 0 0 8 6 9

APPENDIX A

HARDWARE & SOFTWARE CAPABILITIES
OF THE
LABORATORIES INVOLVED
IN THE
INTERLABORATORY WORKING GROUP ON DATA EXCHANGE
(with emphasis on data management)

Prepared by

Richard A. Wiley

APPLICATIONS SUPPORT AND RESEARCH GROUP
COMPUTER SCIENCE AND SERVICES DIVISION

LOS ALAMOS SCIENTIFIC LABORATORY

LOS ALAMOS, NEW MEXICO 87545

CONTENTS

INTRODUCTION.....1
ARGONNE NATIONAL LABORATORY (ANL).....1
BROOKHAVEN NATIONAL LABORATORY (BNL).....6
LAWRENCE BERKELEY LABORATORY (LBL).....9
LAWRENCE LIVERMORE LABORATORY (LLL).....14
LOS ALAMOS SCIENTIFIC LABORATORY (LASL).....17
OAK RIDGE NATIONAL LABORATORY (ORNL).....22
PACIFIC NORTHWEST LABORATORY (PNL).....26

INTRODUCTION

This report provides a brief overview of the hardware and software capabilities of the ERDA laboratories participating in the Interlaboratory Working Group on Data Exchange (IWGDE). It is intended for the use of IWGDE members and does not pretend to be an exhaustive treatment of the total hardware and software capabilities of the individual laboratories.

ARGONNE NATIONAL LABORATORY (ANL)

I. INTRODUCTION

The Applied Mathematics Division (AMD) operates the laboratory's central computing facility and provides programming support for it. Through its Systems and Applications Section, AMD maintains the operating system and engages in system development needed to provide required computing services. It also offers applied programming services.

II. COMPUTER SYSTEM DESCRIPTION

The AMD computer system consists of an IBM 370/195 computer and an IBM 360/75 computer. The model 195 computer is used to process jobs in the batch environment. The model 75 computer runs the Time-Sharing Option (TSO) of OS/360. The Asymmetric Processing System (ASP) is used to operate the batch system and to link the time-sharing system to it. This allows submission of batch jobs through the time-sharing system.

A. Hardware

The model 195 has 4 million bytes of main storage, 15 tape drives, and access to over 3.5 billion bytes of direct access storage. The maximum guaranteed main storage available for user jobs is slightly over 2 million bytes. The tape drives support 7-track recording at densities of 200, 556, or 800 bits/in. and 9-track recording at 800, 1600, or 6250 bits/in. There are three types of direct access devices available to users: over 40 IBM 3330 equivalent disk drives, each of which holds

approximately 100 million bytes and transfers data at 806,000 bytes/sec; 24 2314 disk drives each of which holds approximately 29 million bytes and transfers data at 312,000 bytes/sec; and 2 IBM 2321 data cells, each of which holds 400 million bytes and transfers data at 55,000 bytes/sec.

Other equipment available to model 195 users includes five 1100 line/min printers, two 300 card/min punches, three CALCOMP drum plotters, two microfilm recorders that write graphical or textual information on 35mm film, and a microfiche unit.

The model 75 has 1 million bytes of main storage, 2 million bytes of slow-core storage and access to approximately 1.3 billion bytes of the disk storage attached to the model 195. Approximately 90 dial-up lines are currently available; transmission speeds supported are 10, 15, 30, and 120 characters/sec.

B. Software

1. System Control Software

The model 195 is currently being run with OS/360. The 75 system includes both the Time-Sharing Option (TSO) of OS/360 and WYLBUR. Overall system control is handled by the Asymmetric Processing System (ASP), an operating system that provides a compatible extension of OS which controls the system job stream and provides job entry capabilities.

2. Software on the Model 195 (BATCH)

a. Computer Language Processors.

ALGOL
ASSEMBLER
COBOL (ANS)
FORTRAN-IV
PL/I
RPG
SPEAKEASY
WATFIV

b. Applications Oriented Packages.

| | |
|-------------|--|
| BMD, BMDP | - general purpose statistical programs |
| CSMP III | - continuous systems modeling program |
| CALFORM | - mapping program |
| ESP | - econometric statistical package |
| GPSS | - general purpose simulation system |
| MARK IV | - file management system |
| MPS | - mathematical programming system |
| PMS | - project management system |
| POLYVRT | - mapping system |
| SAS | - statistical analysis system |
| SPSS | - statistical package for the social sciences |
| SYMAP | - program which produces line printer representations of contoured data |
| SYMVU | - program for generating three-dimensional line drawing displays of data |
| SYSTEM 2000 | - data base management system |

c. Subroutine Libraries

| | |
|---------|--|
| AMDLIB | - collection of locally written mathematical and utility subroutines |
| DISSPLA | - integrated system of FORTRAN subroutines for graphical output |
| ESPLOT | - collection of subroutines for producing graphical output on printers |
| IMSL | - International mathematical and Statistical Library |
| SSP | - IBM's scientific subroutine package |

d. Utility Programs

| | |
|--------------|---|
| LIBRARIAN | - source program maintenance and retrieval system |
| OS Utilities | - IBM's multipurpose utility programs |
| SORT/MERGE | - IBM's generalized sorting program |

3. Software on the Model 75 (TSO)

a. Computer Language Processors

FORTRAN IV
PL/I
SPEAKEASY
WATFIV

b. Applications Oriented Packages

MARK IV - file management system
RASS - remote access statistical
system
SYSTEM 2000 - data base management system

c. Subroutine Libraries

AMDLIB - collection of locally
written mathematical and
utility subprograms
DISSPLA - integrated system of
FORTRAN subroutines for
graphical output
IMSL - International Mathematical
and Statistical Library

d. Utility Programs

LIBRARIAN - source program maintenance
and retrieval system
TSO Utilities - facilities for copying,
merging, and formatting TSO
data sets.

III. DATA BASE MANAGEMENT SYSTEMS

ANL has two data base management systems in general use,
MARK IV (INFORMATICS, Inc.) and SYSTEM 2000 (MRI, Inc.).

A. MARK IV

MARK IV provides extensive facilities for the
creation and maintenance of files and data bases,
as well as for report generation. It allows the
creation of highly structured master files. It
updates by passing a transaction file against the
master file, and is thus limited to batch-type
operation, although on-line, free-form input is
available.

B. SYSTEM 2000

SYSTEM 2000 is a full-scale generalized data base management system that uses partially inverted tree structures for efficient on-line and batch data processing. The extent to which the files are inverted is determined by the user through specification of data record fields as key or non-key. Interfaces to COBOL, FORTRAN, PL/I, and other end-user languages are available. A set of near-natural-language user commands is also available. The system is well suited for fully integrated data bases as well as simple, more specific applications on subsets of data. SYSTEM 2000 may be used in either batch or interactive mode.

BROOKHAVEN NATIONAL LABORATORY (BNL)

I. INTRODUCTION

The Central Scientific Computer Facility (CSCF) is operated by the Department of Applied Mathematics, which also provides programming support.

II. COMPUTER SYSTEM DESCRIPTION

The CSCF computer system includes a CDC CYBER 76-18 computer system and two CDC 6600 computers. The CYBER 76-18 computer processes jobs in a batch environment. The 6600 computers run the INTERCOM interactive and remote batch entry system and are connected to the CYBER 76-18, allowing remote submission of jobs for this machine also.

A. Hardware

The CYBER 76-18 has 65,536 60-bit words of Small Core Memory (SCM), 512,000 60-bit words of somewhat slower Large Core Memory (LCM), and 576 million 60-bit words of direct access storage. The CYBER 76-18 computer has no other peripherals since all magnetic tape handling, printing, card reading, and card punching are handled by the 6600's.

One of the 6600 computers has 98,304 60-bit words of main storage, 15 million 60-bit words of direct access storage, six tape drives, three line printers, one card reader, and one card punch. The direct access storage consists of a CDC 844 disk storage system with two drives, each of which holds 12 million 60-bit words and transfers data at a maximum rate of about 925,000 words/sec. The tape drives support 7-track recording at densities of 200, 556, or 800 bits/in. Two of the printers operate at 1000 lines/min. and the third prints at 1200 lines/min. The card reader reads 1200 cards/min. and the card punch punches 250 cards/min.

The second 6600 is identical to the first, except for having only 65,536 60-bit words of main storage and only one drive on the 844 disk storage system.

The two systems share 1.5 million 60-bit words of Extended Core Storage (ECS), 29 million words of

direct access storage, and six 9-track tape drives. The direct access storage is a CDC 841-8 multiple disk drive which transfers data at approximately 180,000 characters/sec. The 9-track tape drives support recording at 800 and 1600 bits/in.

Twenty-eight dialup lines and twenty-four local lines are currently available. Transmission speeds supported are 10, 30, and 200 characters/sec, as well as 4800, 9600, and 50,000 baud on the direct lines.

B. Software

1. System Control Software

The CYBER 76 is currently using SCOPE 2.1.3. The 6600's are using SCOPE 3.4.3.

2. Software on the CYBER 76

a. Language Processors

ALGOL
COBOL
COMPASS
FORTRAN

b. Applications Oriented Packages

There are many applications packages available in the areas of physics, engineering, mathematics, biology, and ecology.

c. Subroutine Libraries

DISSPLA - integrated system of
FORTRAN subroutines for
interactive graphics
IMSL - International Mathematical
and Statistical Library

d. Utility Programs

SORT/MERGE - generalized sorting program
UPDATE - program library maintenance
system

3. Software on the 6600's

a. Language Processors

ALGOL
 BASIC
 COBOL
 COMPASS
 FORTRAN

b. Applications Oriented Packages

QUERY/UPDATE - information retrieval
 language
 SYSTEM 2000 - data base management system

c. Subroutine Libraries

DISSPLA - integrated system of
 FORTRAN subroutines for
 interactive graphics
 IMSL - International Mathematical
 and Statistical Library
 8BIT - 8-bit subroutines

d. Utilities

FORM - file organizer and record
 manager utility
 SORT/MERGE - generalized sorting program
 UPDATE - program library maintenance
 system

III. DATA BASE MANAGEMENT SYSTEMS

The data base management systems currently in use at BNL are SYSTEM 2000 (MRI, Inc.) and QUERY/UPDATE (CDC).

A. SYSTEM 2000

SYSTEM 2000 is described in the ANL section of this appendix.

B. QUERY/UPDATE

QUERY/UPDATE allows the interrogation and maintenance of data files by clerical personnel. The system may be used in either batch or interactive mode. When used in interactive mode QUERY/UPDATE can provide a tutorial dialogue for the user.

LAWRENCE BERKELEY LABORATORY (LBL)

I. INTRODUCTION

The Lawrence Berkeley Laboratory computing facility is run by the Computer Center, part of the Physics, Computer Science and Mathematics Division of the laboratory. This group maintains the operating system, provides operating personnel, and furnishes a number of user services (programming consultants, a program library, a computing library, courses, and seminars).

The Computer Science and Applied Mathematics Department, also part of the Physics, Computer Science and Mathematics Division, undertakes research and development in a wide spectrum of computing related disciplines, including data management systems, information systems, computer networks, computational physics, computer graphics, and applied mathematics.

II. COMPUTER SYSTEM DESCRIPTION

The LBL computing facility is equipped with a CDC 7600, a CDC 6600, and a CDC 6400. The 7600 is used to process jobs in batch mode and is served by the 6400 as an I/O station. The 6600 can also serve as a 7600 I/O station along with its normal load. The 6400 normally provides job input and unit record output processing for both the 7600 and 6600. The 6600 is primarily used for timesharing, supporting interactive programming as well as interactive batch submittal.

A. Hardware

The 7600 has 65,536 60-bit words of Small Core Memory (SCM), 512,000 60-bit words of Large Core Memory (LCM), and 160 million 60-bit words of direct access storage. The direct access storage is provided by two CDC 7638 disk units, each of which holds approximately 80 million 60-bit words and transfers data at 6 million characters/sec. Up to about 25 million words may be used by any one job.

The 6600 has 131,072 60-bit words of memory and 121 million 60-bit words of direct access storage. There are two types of direct access devices available to users: eight CDC 844-2 disk storage units, each of which holds approximately 12 million 60-bit words and transfers data at approximately 925,000 characters/sec. and one CDC 841-8 disk

storage unit which will hold approximately 25 million 60-bit words and transfers data at 180,000 characters/sec. Up to about 3 million words may be used by any one job.

The 6400 has 131,072 60-bit words of main memory. The 6400 normally operates the large random access mass storage devices. These are the IBM 2321 Data Cell System, which holds 385 million words and transfers data at 55,000 characters/sec, and the IBM 1360 Photo-Digital Storage System (Chipstore), which holds 5 billion 60-bit words and transfers data at approximately 400,000 characters/sec. The two 6000's also share eight CDC 659 9-track tape drives, which record at densities of 800 and 1600 bytes/in., and twenty CDC 607 7-track tape drives which record at densities of 200, 556, or 800 bytes/in. Up to eight 607's and three 659's may be used by a single job. One IBM compatible 6250 bytes/in drive is currently being tested.

Also available are two CDC 512 1200 line/min. printers, five IBM 1403 1100 line/min. printers, a Stromberg DATAGRAPHIX 4460 microfilm/microfiche recorder, and four CALCOMP 565 plotters. Alphanumeric and graphic output may also be obtained on five CRT display consoles (CDC-252 Vista). The system supports a variety of graphics terminals in a device independent manner, including Tektronix 4012 and 4014, DEC GT40, GT42 and GT44, etc.

The Remote Equipment Control Computer (RECC) system is a collection of hardware and software which provides terminal service to LBL computer centers users. It gives access to two timesharing systems, permits users to write their own interactive or batch programs, and provides information on the status of the jobs running in the entire 7600/6600/6400 triplex.

The RECC system provides for the simultaneous handling of 224 terminals in full duplex mode. Any EIA compatible terminal, alphanumeric or graphics, is handled by RECC. Dial-up ports are available for 110, 300, and 1200 bits/sec terminals, and dedicated lines are available for up to 9600 bits/second.

The RECC system is available over the ARPANET. Communication with LBL's IMP allows computer center users to run on a variety of computer systems from Hawaii to London.

A-11

In TELNET mode, the system allows interactive access to other ARPANET hosts. In file transfer mode, files can be transferred between machines at an average rate of 14,000 bits/second.

In addition to interfacing with remote terminals for interactive computing and remote job entry, the Berkeley installation is also equipped with a remote batch interface system (University Computing Company COPE controller and a BKY-designed and implemented COKE system). Users who have remote batch stations at their own facilities may submit batch jobs and receive printed output directly.

Currently the remote batch system has 59 ports operating, covering users in 33 states, Washington, D.C., and Canada. Twenty-six ports are connected to COKE 200UT dedicated lines, and 12 to dedicated COPE lines. In addition there are 21 dial-up lines available for 200UT users, including eight 4800 bps lines, seven 2000 bps lines, and six 2000 bps lines connected to the Federal Telecommunication System (FTS) network.

B. Software

1. System Control Software

The operating systems on all three computers have evolved from original CDC systems which have been tailored to fit the particular needs of the Berkeley installation. Most programs run on any of three machines, except where special peripherals required.

2. Computer Language Processors

ALGOL
COBOL
COMPASS
FORTRAN
LISP
PASCAL
PL/I
SNOBOL

3. Applications Oriented Packages

| | |
|-----------|--|
| ALTRAN | - system for symbolic algebra manipulation |
| BMD | - general statistical package |
| BARB12 | - text formatting program |
| CARTE | - interactive mapping system with high quality graphics output |
| CHART | - interactive graphics analysis and report generator for decision making |
| CPMG | - critical path method for project management |
| ERGON | - high energy physics data analysis and display |
| GRASS | - interactive statistical package |
| GUMPS | - linear programming system |
| HEMP | - solves conservation equations for 2D elastic-plastic flow |
| JASON | - calculates electrostatic fields by finite element method |
| MAFCO | - calculates magnetic fields in 3D using elliptic integral method |
| MAPEDIT | - editing and data management system for digitized maps in polygonal form |
| MIMIC | - interactive graphics modeling program for digital simulation of analog systems |
| PICASSO | - general interactive graphics modeling program |
| POLYVRT | - map file translation program |
| SAP | - structural analysis program |
| SPICE | - electronic circuit analysis program |
| SPSS | - statistical package for the social sciences |
| SYMAP | - mapping program for printer plots |
| TRANSPORT | - interactive system for calculating particle beam transport systems to first and second order, including space charge effects |
| TRIM | - calculates non-linear magnetic fields in 2D using finite difference methods |
| ZING | - interactive map digitizing and editing program |

4. Subroutine Libraries

| | |
|----------|--|
| GRAFPAC | - device independent graphics system |
| IGS/NCAR | - high level graphics library |
| IMSL | - library of mathematical and statistical routines |
| LIBRY76 | - mathematical and utility subroutine library for the 7600 |
| LIBRARY | - mathematical and utility subroutine library for the 6000's |

5. Utility Programs

INTERAC - hierarchical file management system
SORT/MERGE - generalized sorting program
UPDATE - program library maintenance system

III. DATA BASE MANAGEMENT SYSTEMS

Several data base management systems are presently in use at LBL. Much of the socio-economic and demographic data is accessed via the BDMS and QWICK QUERY (CACI) systems.

A. BDMS(Berkeley Data Base Management System)

BDMS is a data management system which allows the storage and retrieval of data using keys. It may be accessed interactively or in batch mode. It may also be used from a FORTRAN program which itself may be either interactive or batch. BDMS makes use of a directory file containing the keys for searching.

B. QWICK QUERY

QWICK QUERY is a data analysis and report generation system that allows both programmers and non-programmers to selectively access and display information from existing data files.

C. RFMS(Remote File Management System)

RFMS is a complete interactive data management system, with a user language for the definition, loading, and updating of the data base, as well as for retrieval and report generation. It stores and retrieves strings, dates, integers, and decimal and exponential numbers. It allows hierarchical data structures and repeating groups. Inverted indices are constructed for all data elements, thus making every data record field a key.

D. STOFI(System TO Find Information)

STOFI is a system of subroutines for manipulating hierarchically related data blocks. It is intended for use with large data files in the BKY format.

LAWRENCE LIVERMORE LABORATORY (LLL)

The LLL computer facility is operated by the Computation Department. The Computer Operations Division operates the large and small computers plus the associated peripheral devices. System development and maintenance is done by the Computer Systems Division. Programming support and computer application expertise are provided to users by the Applications Programming Divisions.

II. COMPUTER SYSTEM DESCRIPTION

The major LLL computer system is the OCTOPUS network. This network includes two CDC STAR-100 computers and four CDC 7600 computers, plus two PDP-10 computers which operate the permanent file system. There is also an unclassified network containing a CDC 6600.

A. Hardware

Both of the STAR-100's have 1,240,000 64-bit words of main memory. Each has 80 million 64 bit words of direct access storage.

All four of the 7600's have 65,536 60-bit words of Small Core Memory (SCM), 512,000 60-bit words of Large Core Memory (LCM), and 80 million words of local direct access storage. Eighteen billion 60-bit words of direct access storage are available to users in the permanent file system. Three types of direct access devices are available in this system: an IBM 1360 Photo-Digital Storage system (photostore) which holds approximately 15 billion 60-bit words and transfers data at 200,000 characters/sec; 40 CDC 844-2 disk packs, each of which holds approximately 12 million 60-bit words and transfers data at about 450,000 characters/sec; and an IBM 2321 Data Cell which holds about 50 million 60-bit words and transfers data at 55,000 characters/sec.

Other equipment available to users of the OCTOPUS network includes two Information International, Inc., Model FR-80 microfilm recorders which record onto either 16mm, 35mm, or 105mm film. 16mm and 35mm film can be recorded in color as well as black

and white. Also available are two Data Display, Inc. DD80 microfilm recorders which record on 35mm film in black and white.

The OCTOPUS network currently supports 650 remote interactive terminals and twenty remote job entry terminals. It also supports 200 remote television display units.

The CDC 6600 in the unclassified network has 131,072 60-bit words of main memory and 17 million 60-bit words of direct access storage. This system supports about 80 remote interactive terminals and one remote job entry terminal.

B. Software

1. System Control Software

Both LLL time-sharing networks use the Livermore Time-Sharing System (LTSS). This system was developed at Livermore by the Computer Systems Division to serve the special needs of the laboratory's scientific users.

2. Software on LTSS

a. Language Processors

ALA76
COBOL
FTN-76
LRLTRAN

b. Applications Oriented Packages

MASTER CONTROL - data management system
ORDER - system for controlling the execution of programs in batch mode
RECOG - pattern recognition system

c. Subroutine Libraries

ORDERLIB - collection of locally written mathematical and utility subprograms
TV80LIB - graphics subroutine library

d. Utility Programs

Editors, file transfer utilities, file maintenance utilities.

III. DATA BASE MANAGEMENT SYSTEMS

LLL has a single locally developed data base management system, MASTER CONTROL. This system is designed to unify the storage, validation, manipulation, reorganization, retrieval, and display of information and data. It is open-ended, user-oriented, and applicable to bibliographic, administrative, and scientific purposes. The command language is English and allows the substitution of synonym expressions and abbreviations through user-defined macro commands. Numerous provisions exist for functional analysis of data, including text. The system is normally used interactively although batch operation is possible by making use of certain special features available in the Livermore Time Sharing System.

LOS ALAMOS SCIENTIFIC LABORATORY (LASL)

The Computer Science and Services Division (C Division) is responsible for the Central Computing Facility (CCF) at LASL. C Division provides operating system maintenance as well as the development of operating systems for new computers, programming support for users, and a wide variety of user services.

II. COMPUTER SYSTEM DESCRIPTION

The CCF contains four CDC 7600 computers, two CDC 6600 computers, and two CDC CYBER 73 computers, all tied to the CCF network. Two of the 7600's operate in batch mode only, the other two 7600's run the LTSS Time-Sharing System, and the two 6600's and one CYBER 73 run the NOS Time-Sharing System. The second CYBER 73 controls the Common File System.

A. Hardware

All four of the 7600's have 65,536 60-bit words of Small Core Memory and 512,000 60-bit words of Large Core Memory. Two of them have 120 million 60-bit words of direct access storage, one has 80 million words of direct access storage, and one has 152 million words of direct access storage. The direct access storage consists of CDC 7638 disk files, each of which holds 40 million 60-bit words and transfers data at 1 million characters/sec. One 7600 also has six CDC 844-2 disk packs, each of which holds 12 million 60-bit words and transfers data at 925,000 characters/sec.

One CDC 6600 (Machine 0) has 131,072 60-bit words of main memory, 1 million 60-bit words of Extended Core Storage (ECS), 125 million 60-bit words of direct access storage, and eight 7-track tape drives.

This direct access storage is made up of three CDC 6638 disk files, each of which holds about 18 million 60-bit words and transfers data at 1.5 million characters/sec, and six CDC 844-2 disk packs, each of which holds 12 million 60-bit words and transfers data at 925,000 characters/sec.

The second 6600 has 65,536 60-bit words of main memory, 13 million 60-bit words of direct access storage, and four 7-track tape drives.

One of the CYBER 73's has 131,072 60-bit words of main memory, 96 million 60-bit words of direct access storage, two 7-track tape drives, and two 9-track tape drives. The direct access storage is eight CDC 844-2 disk packs, each of which holds 12 million 60-bit words and transfers data at 925,000 characters/sec.

The second CYBER 73 has 65,536 60-bit words of main memory, 36 million words of direct access storage on 844-2 disk packs, and four 7-track tape drives. This computer presently manages the Common Filing System (CFS). The CFS has a total of about 15.3 billion 60-bit words of direct access storage, consisting of an IBM 1360 Photo-Digital Storage System which holds about 15 billion 60-bit words and transfers data at about 200,000 characters/sec., two CDC 821 Disk File Systems, each of which holds about 80 million 60-bit words and transfers data at about 180,000 characters/sec., and nine CDC 844-2 Disk Storage Units, each of which holds about 12 million 60-bit words and transfers data at about 450,000 characters/sec..

The following special equipment is available to CCF users: two Information International, Inc. FR-80 microfilm recorders which record on 16mm, 35mm, and 105mm film in black and white, plus color on 16mm and 35mm film; one Stromberg DATAGRAPHIX 4020 microfilm recorder which records in color and black and white on 16mm and 35mm film; two Sanders Graphics terminals which provide interactive graphics in black and white or color and are linked to the 7600 batch system(CROS); the Video Image Display System (VIDS), which provides both a high-quality TV monitor with four color and full gray scale display and a TV camera digitizer; and a DATAGRAFIX alphanumeric microfiche recorder.

The CCF network currently supports 218 remote interactive terminal ports, plus 72 ports that go only to 6600 #0. Twenty of these ports are dialup and six of them are remote job entry terminals (CDC 200 User Terminals).

B. Software

1. System Control Software

The two batch 7600's run the CROS operating system. The two time-sharing 7600's run the Livermore Time-Sharing System (LTSS). The two 6600's and the CYBER 73 run the CDC NOS Time-Sharing system.

2. Software on the 7600's Running CROS

a. Language Processors

COMPASS
FORTRAN

b. Applications Oriented Packages

PERT - project management system

c. Subroutine Libraries

LADIES - Los Alamos Digital Image
Enhancement System

d. Utility Programs

UPDATE - program library maintenance
system

3. Software on the 7600's Running LTSS

a. Language Processors

ALA76
COBOL
FORTRAN

b. Applications Oriented Packages

MASTER CONTROL - data base management system

c. Subroutine Libraries

- DISSPLA - integrated system of FORTRAN
subroutine for graphical output
- ORDERLIB - collection of mathematical
and utility subprograms

d. Utility Programs

Facilities for copying, listing, merging,
and formatting data files.

4. Software on the NOS Computers

a. Language Processors

ALGOL
BASIC
COBOL
FORTRAN
PASCAL
SIMULA
SIMSCRIPT
SNOBOL

b. Applications Oriented Packages

- APEX III - linear programming system
- GMAPS - general map analysis and
planning system
- GPSS - general purpose simulation
system
- PERT - project management system
- SIMSCRIPT II.5 - general purpose simulation
system
- SPSS - statistical packages for the
social science
- SYSTEM/2000 - data base management
system

c. Subroutine Libraries

- DISSPLA - integrated system of
FORTRAN subroutines for
graphical output
- IMSL - commercial library of
mathematical and statistical
subroutines
- LASLIB - collection of locally
written mathematical and
utility subprograms

d. Utility Programs

| | |
|------------|--------------------------------------|
| MODIFY | - program library maintenance system |
| SORT/MERGE | - generalized sorting program |
| UPDATE | - program library maintenance system |

III. DATA BASE MANAGEMENT SYSTEMS

LASL currently provides two data base management systems, MASTER CONTROL on LTSS and SYSTEM 2000 (MRI, Inc.) on NOS.

A. MASTER CONTROL

MASTER CONTROL is described in the LLL section of this appendix.

B. SYSTEM 2000

SYSTEM 2000 is described in the ANL section of this appendix.

OAK RIDGE NATIONAL LABORATORY (ORNL)

I. INTRODUCTION

The Computer Science Division (CSD) at Union Carbide Corporation - Nuclear Division (UCC-ND) operates a computer network which serves ORNL, the Oak Ridge Gaseous Diffusion Plant (ORGDP), the Oak Ridge Y-12 Plant, the ERDA Technical Information Center (TIC), and a number of remote users. CSD provides operating personnel, operating systems maintenance, and programming services.

II. COMPUTER SYSTEM DESCRIPTION

The CSD Computer Network consists of an IBM 360/91, an IBM 360/75, and a DEC PDP-10 at the X-10 site, an IBM 370/155 and an IBM 360/195 at the K-25 site, and an IBM 360/50 and a CDC 3300 at the Y-12 site. The computers at the Y-12 site are used primarily to support production at that site and will not be mentioned further.

A. Hardware

The IBM 360/195 has 1 million bytes of main storage, 8 million bytes of direct access storage, seven 9-track tape drives, and access to 3.3 billion bytes of direct access storage shared with the 370/155. The tape drives support 9-track recording at 800 and 1600 bits/in. This computer has two IBM 2301 Drum Units attached to it, each of which holds 4 million bytes and transfers data at a maximum rate of 1.2 million bytes/sec. This computer is connected to 14 hard wired remote terminals and has a direct connection to the 370/155. It has no unit record equipment attached to it, these services being provided by the 370/155.

The IBM 370/155 has 1 million bytes of main storage, 4 million bytes of direct access storage, 4 tape drives, and shares 3.3 billion bytes of direct access storage with the 360/195. The tape drives support 9-track recording at 800 and 1600 bits/in. The direct access storage is an IBM 2303 Drum Unit which holds 4 million bytes and has a maximum data transfer rate of 312,000 bytes/sec.

This computer is connected to six remote batch terminals as well as the IBM 360/50 at the Y-12 site. It also has an input-only link to the 360/91 and 360/95 at the X-10 site. It has five printers, two card readers, a card reader/punch, and a card punch.

The direct access storage shared by these two computers is made up of three different device types. Two IBM 3330 Disk Storage Systems each have a total capacity of 800 million bytes and a maximum data transfer rate of 312,000 bytes/sec. Three IBM 2321 Data Cells each have a capacity of 400 million bytes and a maximum data transfer rate of 55,000 bytes/sec. The two systems also share eight 3420-7 tape drives, which support 9-track recording at 800 and 1600 bits/inch.

The IBM 360/91 computer has 2 million bytes of main storage, 38 million bytes of direct access storage, and five 9-track tape drives. The direct access storage consists of two 2301 Drum Storage Units with a total capacity of 8 million bytes and a 1.2 million byte/sec maximum data transfer rate, and a 2314 Disk Storage Unit with a capacity of 30 million bytes and a maximum data transfer rate of 312,000 bytes/sec. It also shares 1.1 billion bytes of direct access storage with the IBM 360/75. It is connected to the 360/75 and to a DEC PDP-15 which controls remote terminals and dial-up access.

The IBM 360/75 computer has 512,000 bytes of main storage, 2 million bytes of Large Capacity Storage (LCS), 30 million bytes of direct access storage, and 4 tape drives. The direct access storage is five IBM 2311 Disk Storage Units with a total capacity of 29 million bytes and a maximum data transfer rate of 156,000 bytes/sec. It also is connected to the DEC PDP-15 which controls terminal access.

The shared direct access storage is composed of three device types. Three IBM 2314 Disk Storage Systems each hold 230 million bytes and have a 312,000 byte/sec maximum data transfer rate. Four IBM 2311 Disk Storage Systems each hold 7 million bytes and have a maximum data transfer rate of 156,000 bytes/sec. An IBM 2311 Data Cell System has a total capacity of 400 million bytes and a maximum data transfer rate of 55,000 bytes/sec.

B. Software

1. System Control Software

The 360/91 and the 360/75 computers operate under HASP-MVT. The 360/75 also runs the Time-Sharing Option (TSO). The 360/195 and the 370/155 run under ASP-MVT.

2. Software on the 360/370 Systems

With the exception of security constraints, these four computers can be used more or less interchangeably and therefore only one set of inclusive software will be described.

a. Computer Language Processors

ASSEMBLER
 COBOL
 FORTRAN IV
 PL/I

b. Applications Oriented Packages

There are an estimated 2000 applications packages in the areas of nuclear engineering, physics, engineering, mathematics, and information Processing. A partial inventory (900) of programs is available as well as a bibliography of computer oriented papers (1300) and CSD capabilities inventory. These include:

ADSEP - Automated Data Set Editing Program
 ORCHIS - Oak Ridge Computerized Hierarchical Information System
 ORRMIS - Oak Ridge Regional Modeling Information System
 RECON - Computerized Information Retrieval System

c. Subroutine Libraries

Several on-line libraries are available, including:

- IMSL - International Mathematical and Statistical Library
- SSP - IBM Scientific Subroutine Package

d. Utility Programs

- OS Utilities - IBM's multipurpose utility programs
- SORT/MERGE - IBM's generalized sorting program

III. DATA BASE MANAGEMENT SYSTEMS

ORNL has a locally developed data base management system, the Oak Ridge Computerized Hierarchical Information System (ORCHIS). A number of special purpose data management systems have been developed at ORNL (e. g. ORLOOK, ORRMIS, RECON), most of which use ORCHIS for the actual data management. ORCHIS 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 elements, and statistical analysis are included as well as the capacity to produce hard copy documents having comprehensive working indices and tabular data display. Although ORCHIS itself is limited to batch use, many of the systems which use it are themselves interactive.

PACIFIC NORTHWEST LABORATORY (PNL)

I. INTRODUCTION

The PNL Computer Facility is operated by the Systems Department. This department provides operating personnel, systems support, and applications support. The facility itself is small, and the majority of the PNL computing work is done remotely at the Boeing Computer Services (BCS) facility in Richland, Wash. PNL also has high-speed remote lines to the Battelle Columbus Laboratory (BCL), BNL, and LBL. Both the BCS and BCL facilities will be described briefly.

II. COMPUTER SYSTEM DESCRIPTION

The PNL interactive computer facility consists of two major systems. The main system is a DEC PDP-11/45 and the secondary system is a DEC PDP-9. The PDP-9 is interfaced with a CDC 6600 computer at BCS via a 40Kb line. The facility also has Remote Job Entry (RJE) terminals connected to BCL (CDC 6600 and 6400), BNL, and LBL.

A. Hardware

The DEC PDP-11/45 has 248,000 bytes of memory, 80 million bytes of direct access storage, two 7-track tape drives, two 9-track tape drives, a 300 line/min. printer, an 1100 line/min. printer, and a card reader. It also has plotters, terminals, and analog/digital conversion devices.

The DEC PDP-9 has 16,000 18-bit words of memory, one million 18-bit words of direct access storage, and two 7-track tape drives. It has an interactive graphics terminal, a digitizer, and a link to the PDP-11, thus allowing the two computers to share peripherals.

B. Software

1. Software at PNL

The software available at PNL is BASIC, COBOL, FORTRAN, and a number of applications packages.

2. Software at BCS

The CDC 6600 at BCS uses the CDC SCOPE 3.4 operating system and the CDC INTERCOM interactive and remote batch entry system.

a. Computer Language Processors

COBOL
FORTRAN
SIMSCRIPT

b. Applications Oriented Packages

MARS - Multi-Access Retrieval System
OMNITAB - data reduction system
QUERY/UPDATE - information retrieval language
SPSS - Statistical Package for the Social Sciences
SYSTEM 2000 - data base management system

c. Subroutine Libraries

DISSPLA - integrated system of FORTRAN subroutines for graphical output

d. Utility Programs

SORT/MERGE - generalized sorting program
UPDATE - program library maintenance system

3. Software at BCL

The CDC 6600 and CDC 6400 at BCL use the CDC SCOPE 3.4 operating system and the CDC INTERCOM interactive and remote job entry system.

a. Computer Language Processors

COBOL
FORTRAN

b. Applications Oriented Packages

A number of applications packages in both the physical and social sciences are available.

c. Subroutine Libraries

IMSL - International Mathematical
and Statistical Library

d. Utility Programs

SORT/MERGE - generalized sorting program
UPDATE - program library maintenance
system

III. DATA BASE MANAGEMENT SYSTEMS

PNL uses the MRI SYSTEM 2000 data base management system for general data base management. This system is available at BCS and BNL. For a description of SYSTEM 2000 refer to the ANL section of this appendix.

APPENDIX B

INTERLABORATORY WORKING GROUP FOR DATA EXCHANGE

Ted Albert
Office of Environmental Information Systems
Environment and Safety
ERDA
Washington, D. C. 20545

Don Austin (Chairman)
Computer Science and Applied Math Department
Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720
FTS 451-5313

Carmen Benkovitz (Secretary)
Department of Applied Science
Brookhaven National Laboratory
Upton, New York 11973
FTS 664-4135

Al Brooks
Computing Applications Department
Computer Sciences Division
Oak Ridge National Laboratory P.O. Box X
Oak Ridge, Tennessee 37830
FTS 850-0147

Paul Dionne
Systems Department
Pacific Northwest Laboratory
P.O. Box 999
Richland, Washington 99352
FTS 444-7411 ask for 946-2452

Murrey Goldberg
Department of Applied Science
Brookhaven National Laboratory
Upton, New York 11973
FTS 664-2028

Viktor Hampel
Computation Department
Lawrence Livermore Laboratory
University of California
P.O. Box 808
Livermore, California 94550
FTS 457-3406

E. Richard Hill
Systems Department
Pacific Northwest Laboratory
P.O. Box 999
Richland, Washington 99352
FTS 444-7411 ask for 946-2675

Robert Honea
Energy Division
Oak Ridge National Laboratory P.O. Box X
Oak Ridge, Tennessee 37830
FTS 850-1486

David Layton
Regional Studies Program
Lawrence Livermore Laboratory
University of California
P. O. Box 808
Livermore, California 94550
FTS 457-3630

Ellen Leonard
Regional Studies Program
Q-10, MS571
Los Alamos Scientific Laboratory
P.O. Box 1663
Los Alamos, New Mexico 87545
FTS 843-6722

B-3

Art Lindeman
Applied Mathematics Division
Argonne National Laboratory
9700 South Cass Avenue
Argonne, Illinois 60439
FTS 388-4247

Andrew Loebel
Energy Division
Oak Ridge National Laboratory
P.O. Box x
Oak Ridge, Tennessee 37830
FTS 850-0311

Ronald R. MacDonald
Energy and Environmental Systems Division
Argonne National Laboratory
9700 South Cass Avenue
Argonne, Illinois 60439
FTS 388-5258

James McClary
Computer Science and Services Division
C-3, MS265
Los Alamos Scientific Laboratory
P.O. Box 1663
Los Alamos, New Mexico 87545
FTS 843-7039

Deane Merrill
Computer Science and Applied Math Department
Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720
FTS 451-5063

Joseph Nardi
Applied Mathematics Department
Brookhaven, National Laboratory
Upton, New York 11973
FTS 664-4122

B-4

Lorraine Osterer
Applied Mathematics Department
Brookhaven National Laboratory
Upton, New York 11973
FTS 664-4156

David Richards
Computer Science and Applied Math Department
Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720
FTS 451-5766

Daniel Ries
Computation Department
Lawrence Livermore Laboratory
P.O. Box 808
Livermore, California 94550
FTS 457-3074

Dieter Tuerpe
Mail Stop L-142
Lawrence Livermore Laboratory
Livermore, California
FTS 457-3481

Phyllis Walker
Energy and Environmental Systems Division
Argonne National Laboratory
9700 South Cass Avenue
Argonne, Illinois 60439
FTS 388-2912

0 0 0 0 4 6 0 0 8 8 7

B-5

Richard Wiley
Computer Science and Services Division
C-3, MS265
Los Alamos Scientific Laboratory
P. O. Box 1663
Los Alamos, New Mexico 87545
FTS 843-7039

John Wilson
Information Systems Specialist
Division of Biomedical and Environmental Research
U. S. Energy Research and Development Administration
Washington, D. C. 20545
FTS 973-4086

STEERING COMMITTEE FOR ANALYSIS AND ASSESSMENT PROGRAMS

William Fulkerson
Director
Energy Division
Oak Ridge National Laboratory
P. O. Box X
Oak Ridge, Tennessee 37830
FTS 850-1519

John Hoover, Assistant Division Director
Energy and Environmental Systems Division
Argonne National Laboratory
9700 South Cass Avenue
Argonne, Illinois 60439
FTS 388-5137

Richard Malenfant
Group Leader, Analysis and Assessment Program
Q-10, MS571
Los Alamos Scientific Laboratory
University of California
P. O. Box 1663
Los Alamos, New Mexico 87545
FTS 843-6441

Phillip Palmedo, Head
Energy Policy Analysis Group
Building 475
Brookhaven National Laboratory
Upton, New York 11973
FTS 644-2316 or 2318

Ronald Ritschard
Head, Analysis and Assessment Program
Lawrence Livermore Laboratory
University of California
P. O. Box 808
Livermore, California 94550
FTS 457-3597

B-7

William E. Siri
Group Leader, Regional Studies Program
University of California
Lawrence Berkeley Laboratory
Berkeley, California 94720
FTS 451-5396

Ward Swift
Project Manager
Regional Studies Program
Battelle
Pacific Northwest Laboratories
P. O. Box 999
Richland, Washington 99352
FTS 444-7411 ask for 946-2229

GEOGRAPHIC INTERCHANGE STANDARDS COMMITTEE

Don Austin
Computer Science and Applied Math Department
Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720
FTS 451-5313

Carmen Benkovitz
Department of Applied Science
Brookhaven National Laboratory
Upton, New York 11973
FTS 664-4135

Al Brooks
Computing Applications Department
Computer Sciences Division
Oak Ridge National Laboratory P.O. Box X
Oak Ridge, Tennessee 37830
FTS 850-0147

Paul Dionne (Chairman)
Systems Department
Pacific Northwest Laboratory
P.O. Box 999
Richland, Washington 99352
FTS 444-7411 ask for 946-2452

E. Richard Hill
Systems Department
Pacific Northwest Laboratory
P.O. Box 999
Richland, Washington 99352
FTS 444-7411 ask for 946-2675

B-9

Harvard Holmes
Computer Science and Applied Math Department
Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720
FTS 451-5742

Ronald R. MacDonald
Energy and Environmental Systems Division
Argonne National Laboratory
9700 South Cass Avenue
Argonne, Illinois 60439
FTS 388-5258

Phyllis Walker
Energy and Environmental Systems Division
Argonne National Laboratory
9700 South Cass Avenue
Argonne, Illinois 60439
FTS 388-2912

Richard Wiley
Computer Science and Services Division
C-3, MS265
Los Alamos Scientific Laboratory
P. O. Box 1663
Los Alamos, New Mexico 87545
FTS 843-7039

John Wilson
Information Systems Specialist
Division of Biomedical and Environmental Research
U. S. Energy Research and Development Administration
Washington, D. C. 20545
FTS 973-4086

0 0 0 0 4 6 0 0 8 9 0

APPENDIX C

DATA EXCHANGES BETWEEN
ERDA NATIONAL LABORATORIES

Deane Merrill
Computer Science and Applied Mathematics
Department
Lawrence Berkeley Laboratory
Berkeley CA 94720

September 9, 1976
(Revised)

TABLE OF CONTENTS 09 SEP 76

| | |
|---|----|
| 1. APPENDIX C - INTERLABORATORY DATA EXCHANGES | 1 |
| 1.1 INTRODUCTION | 1 |
| 1.2 DATA FROM BNL | 2 |
| (BNL TO LASL) DATA FROM SAMPLING BUOYS | 2 |
| (BNL TO LASL, LBL) NASN AIR QUALITY | 2 |
| (BNL TO LBL) ENERGY MODEL DATA BASE | 3 |
| (BNL TO LBL, ORNL) COUNTY ENERGY DATA BASE | 4 |
| 1.3 DATA FROM LASL | 5 |
| (LASL TO LBL) ERDA COMPUTING FACILITIES | 5 |
| 1.4 DATA FROM LBL | 6 |
| (LBL TO ANL, BNL, LASL) OBERS PROJECTIONS | 6 |
| (LBL TO ANL, BNL, LASL, ORNL) CITY-COUNTY DATA BOOK | 6 |
| (LBL TO BNL) 1972 U.S. EMPLOYMENT DATA | 7 |
| (LBL TO BNL, LASL, ORNL) GEOGRAPHIC DATA BASES | 8 |
| (LBL TO LASL) INPUT-OUTPUT TABLES | 10 |
| (LBL TO LASL, LLL) COUNTY BUSINESS PATTERNS | 10 |
| 1.5 DATA FROM LLL | 12 |
| (LLS TO ORNL) INDEX OF DATA | 12 |

TABLE OF CONTENTS 09 SEP 76

BASES AND MODELS

| | |
|--|----|
| 1.6 DATA FROM ORNL | 13 |
| (ORNL TO ANL TO LBL TO LASL) DUN AND BRADSTREET | 13 |
| (ORNL TO ANL, BNL) U.S. CENSUS DATA | 14 |
| (ORNL TO LBL) IWGDE EXCHANGE STANDARD | 14 |
| (ORNL TO LBL) 1970 CENSUS DEFINITION FILE | 15 |
| (ORNL TO LBL) 1970 CENSUS, FOURTH COUNT HOUSING | 15 |
| (ORNL TO SRL) 1970 POPULATION DATA GRID | 16 |

1. APPENDIX C - INTERLABORATORY DATA EXCHANGES 1 INTRODUCTION

1.1 INTRODUCTION

During FY 1976 numerous socio-economic, environmental, demographic, and energy-related data bases were exchanged between two or more ERDA national laboratories. The data bases discussed in this report are those used in the Energy Analysis and Assessment Programs, or the Regional Studies Programs, at the ERDA multipurpose national laboratories. We use the following abbreviations:

| | |
|------|----------------------------------|
| ANL | Argonne National Laboratory |
| BNL | Brookhaven National Laboratory |
| LASL | Los Alamos Scientific Laboratory |
| LBL | Lawrence Berkeley Laboratory |
| LLL | Lawrence Livermore Laboratory |
| ORNL | Oak Ridge National Laboratory |
| PNL | Pacific Northwest Laboratory |
| SRL | Savannah River Laboratory |

This report was prepared with the assistance of many people in the ERDA IWGDE (Interlaboratory Working Group for Data Exchange) and others; in particular Andy Loebel (ORNL), Carmen Benkovitz (BNL), Dick Wiley (LASL), Don Austin (LBL), Paul Dionne (PNL), Phyllis Walker (ANL), Mark Horovitz (LBL), and Henry Ruderman (LBL).

1. APPENDIX C - INTERLABORATORY DATA EXCHANGES 2
DATA FROM BNL

1.2 DATA FROM BNL

(BNL TO LASL) DATA FROM SAMPLING BUOYS

In August 1976 Carmen Benkovitz of BNL sent to Dick Wiley at LASL a small data base containing measurements from sampling buoys in Long Island Sound. Although this is only a small sample data base, the exchange is significant because it is the first exchange successfully accomplished under the IWGDE data exchange standard.

(BNL TO LASL, LBL) NASN AIR QUALITY

This data base contains measurements of ambient air quality, made by NASN, the National Aerometric Surveillance Network, between 1957 and 1974, over the entire U.S. Not all pollutants were measured in all years. Data include geographic codes, site code, sampling dates and time intervals, pollutant type, method used, observed concentrations, wind speed and direction, and units of each parameter. The data were originally obtained from NADB, the National Aerometric Data Bank, Environmental Protection Agency (EPA), Research Triangle Park, North Carolina.

In early 1976, data for the western states were sent to LASL. In July 1976, data for the whole U.S. were sent to LBL, where they will be used in a two-year project funded by EPA, entitled 'Populations-at-Risk to Various Air Pollution Exposures'. This project is a collaboration between LBL's Energy and Environment Division and the Computer Science and Applied Mathematics Department.

1. APPENDIX C - INTERLABORATORY DATA EXCHANGES 3
DATA FROM BNL

Primary data on

- (a) socioeconomic and demographic characteristics
- (b) air pollution levels for several important pollutants
- (c) meteorological conditions
- (d) health, disease, and mortality

will be graphically displayed, interpolated as necessary, and statistically analyzed to investigate possible significant correlations.

(BNL TO LBL) ENERGY MODEL DATA BASE

These are the data used by the BNL Energy model known as ESNS (Energy Systems Network Simulator), described in Ref. BNL.8. Portions of this data base are known as the Reference Energy System, or the MERES Data Base. Detailed energy and pollution coefficients, costs, efficiencies, etc. are provided for about 600 energy supply processes and about 200 end-use energy utilization processes.

In late 1975 tapes containing the data in card and print format were sent by Paula Newhouse of BNL to Mark Horowitz at LBL. At LBL the data were used in a model to study the impacts of various California energy alternatives, including a proposed moratorium on nuclear power plant construction. (Ref. LBL.5). The same data were also used in a non-price energy conservation model being developed by the LBL Energy Use and Conservation group. (Ref. LBL.6). In December 1975 the same data were sent upon request from LBL to the California Energy Resources Conservation and Development Commission in Sacramento.

In a modified form, the BNL Energy Model Data Base arrived at LBL indirectly, without the participation (or even the previous knowledge) of LBL personnel. A combined energy I/O model,

1. APPENDIX C - INTERLABORATORY DATA EXCHANGES 4
DATA FROM BNL

including data from BNL, LBL, and other sources, has been stored and successfully run at LBL, from computer terminals in Illinois, by J.C. McMillan and Clark Bullard of the Center for Advanced Computation, University of Illinois at Urbana-Champaign.

Additional data for new energy technologies has been developed by Teknekron, Inc. for inclusion in the BNL Energy Model Data Base (Ref. NE.6). In the near future these data will be sent from BNL to Mark Horovitz of LBL.

(BNL TO LBL, ORNL) COUNTY ENERGY DATA BASE

This data base contains estimates of production and use of various fuels, and of emissions of various pollutants, for all counties of the U.S. for the year 1972 (approximately). In April 1976, Frank Drysdale and Charles Calef of BNL's Biomedical and Environmental Assessment Division brought the data base in tape form to LBL. With the assistance of Peter Wood of LBL's Computer Science and Applied Mathematics Department, they converted the data for input to CARTE, an LBL computer mapping program. (Refs. LBL.24, LBL.31). After several days of experimentation, they produced a series of 31 color maps of the U.S., showing production and use of various fuels, emissions of various pollutants, and other quantities. The maps are for publication in a forthcoming BNL report. (Ref. BNL.9).

A copy of the BNL data base is in the LBL Computer Center, in a format convenient for further mapping or modeling applications by local or remote users. The County Energy Data Base was also sent to ORNL, for use in modeling applications in ORNL's Energy Division.

1. APPENDIX C - INTERLABORATORY DATA EXCHANGES 5
DATA FROM LASL

1.3 DATA FROM LASL

(LASL TO LBL) ERDA COMPUTING FACILITIES

In July 1976, Dick Wiley of LASL prepared a descriptive summary of hardware and software capabilities of computer center installations at seven ERDA multipurpose laboratories. Using a cassette terminal at LASL, Dick Wiley transmitted the document directly to the LBL PSS (Program Storage System). Deane Merrill and Don Austin made some minor modifications, and included the writeup as Appendix A of the IWGDE FY76 Annual Report. From LBL, the document is directly accessible to a larger user community, including users at installations on the ARPANET. Updates and modifications received from LASL and elsewhere will be incorporated in the LBL version.

1. APPENDIX C - INTERLABORATORY DATA EXCHANGES 6
DATA FROM LBL

1.4 DATA FROM LBL

(LBL TO ANL, BNL, LASL) OBERS PROJECTIONS

The Series E OBERS Projections contain data for 1950, 1962, 1969, 1970, 1971, 1975, and projections for 1980, 1985, 1990, 2000, and 2020; of various socio-economic indicators, for states, SMSA's, BEA (Bureau of Economic Analysis) Areas, and Water Resource Areas. Data and projections presented include population, income, employment, earnings, and various ratios of these quantities. Earnings are disaggregated by 3-digit SIC (Standard Industrial Classification) Code. These data are in the LBL interactive retrieval system REAP, written by Fred Gey to access files in the QWICK QUERY data base management system. (Ref. LBL.25). In March 1976 these files, in QWICK QUERY format, were sent by Don Austin of LBL to Ellen Leonard at LASL, to Phyllis Walker at ANL, and to Carmen Benkovitz at BNL.

(LBL TO ANL, BNL, LASL, ORNL) CITY-COUNTY DATA BOOK

The City-County Data Book contains various socio-economic indicators, by county, for various years from 1940 through 1972. Not all data are presented for all years. Included are selected data from the U.S. Census of Population and Housing (1940, 1950, 1960, 1970); the Census of Agriculture (1949, 1959, 1964); the Census of Manufactures (1963); the Census of Business (1958, 1963, 1967); the Census of Governments (1962); and other data. Like the OBERS Projections, these data are in the REAP interactive retrieval system (Ref. LBL.25).

In March 1976 these files, in QWICK QUERY format, were sent by Don Austin of LBL to Ellen Leonard at LASL and to Phyllis Walker at ANL. The same files will also be sent to Carmen Benkovitz at BNL and Andy Loebl at ORNL in the IWGDE standard exchange format.

1. APPENDIX C - INTERLABORATORY DATA EXCHANGES 7
DATA FROM LBL

(LBL TO BNL) 1972 U.S. EMPLOYMENT DATA

Estimates of 1972 U.S. employment have been developed by Deane Merrill, for use in energy-related input-output and linear programming models being used at LBL, at BNL, at the University of Illinois Center for Advanced Computation (CAC), and elsewhere. The estimates are disaggregated by industry, by occupation, and by state; separate estimates are available for jobs, persons employed, and full-time equivalents. (Refs. LBL.23, LBL.26).

Regarding the disaggregation by industry, estimates are provided for

- (a) 368 input-output industry sectors defined by the Bureau of Economic Analysis (BEA);
- (b) 102 BEA input-output sectors (essentially 2-digit BEA codes);
- (c) 14 broad industry groups;
- (d) 2-digit 1967 SIC Codes;
- (e) 201 industry sectors defined in the 1970 Census of Population.

Regarding the disaggregation by occupation, estimates are provided for each of 422 detailed occupations defined in the 1970 Census of Population and in the annual Current Population Survey. These can be conveniently aggregated to about 40 broader occupational categories, or to about 12 still broader occupational groups. Disaggregation by occupation is provided at the national level only (not separately for each state).

In April 1976, some of these data (persons employed, national, for 102 BEA input-output industry sectors and 422 occupations) were sent by Deane Merrill to Richard Goettle in the Economic Analysis Division of BNL's Department of Applied Science. Initially, the data found application in a national nuclear moratorium analysis performed for ERDA's Assistant Director

I. APPENDIX C - INTERLABORATORY DATA EXCHANGES 8
DATA FROM LBL

for Systems Analysis, Richard Williamson. In August 1976 Robert Stern of BNL's Policy Analysis Division retrieved additional related data from LBL, using a computer terminal at BNL. Other applications of the data are envisaged.

(LBL TO BNL, LASL, ORNL) GEOGRAPHIC DATA BASES

At LBL, a computer-based mapping system has been developed to produce high-quality, low-cost maps for graphical display of statistical data by geographical and political area, and to provide an error-free geographical data base for spatial analysis applications. (See Appendix C of Ref. LBL.10 for a fuller description.) The project consists of three major components: a system for automatic digitizing of base maps; the MAPEDIT system (Ref. LBL.30) for editing, coding and retrieving the digitized maps; and the CARTE system (Refs. LBL.24, LBL.31) for correlating statistical data with geographic boundaries for the production of print-quality microfilm negatives by computer.

At its present state of development, the LBL geographic data base includes:

- (a) U.S. by state (50 boundaries)
- (b) U.S. by county (3200 boundaries)
- (c) 275 Standard Metropolitan Statistical Area (SMSA) boundaries
- (d) 241 SMSA's by census tract (35,000 boundaries)
- (e) point locations for some 500 cities
- (f) U.S. by BEA Area (173 boundaries)
- (g) U.S. by Air Quality Control Region (247 boundaries)
- (h) 10 EPA Federal Regions

1. APPENDIX C - INTERLABORATORY DATA EXCHANGES 9
DATA FROM LBL

(i) 3 EPA Special Regions

(j) 9 National Electric Reliability
Council (NERC) Regions.

Digitization of other elements, including air basin and watershed boundaries, and point locations of air quality monitoring stations, is contemplated.

Various subsets of the LBL geographic data base have been sent to BNL, LASL, and ORNL. For further details, contact Don Austin (LBL), Carmen Benkovitz (BNL), Dick Wiley (LASL), or Andy Loebel (ORNL).

One transfer was made in POLYVRT format, in order to test the utility of this format for interlaboratory exchange of geographic data bases. POLYVRT, which runs at LBL and ORNL (and elsewhere), is a program intended for mapping applications, available from the Harvard University Laboratory for Computer Graphics and Spatial Analysis. In early 1976, in connection with IWGDE (Interlaboratory Working Group for Data Exchange) activities, Don Austin converted a small cartographic data base containing census tract boundaries of the Reno, Nevada SMSA into POLYVRT format. A tape was mailed to Dick Durfee at ORNL, who successfully reproduced the polygonal tract boundaries, and went on to produce a contour map of population densities, through a spatial interpolation of enumeration district data on the MEDX tapes.

The POLYVRT exchange experiment revealed two serious deficiencies; namely, the inability to carry along textual identifiers associated with particular polygonal boundaries, and the inability to conveniently handle polygons nested within other polygons. The POLYVRT format appears to offer no advantages over those of the proposed IWGDE exchange standard (Ref. ORNL.2). For these reasons the IWGDE decided to abandon the use of POLYVRT as an exchange medium for geographic data bases.

1. APPENDIX C - INTERLABORATORY DATA EXCHANGES 10 DATA FROM LBL

(LBL TO LASL) INPUT-OUTPUT TABLES

The LBL Regional Studies Group has developed 404-sector 1967 input-output tables for California and the eight Rocky Mountain States. (Ref. LBL.2). These tables are based on the 367-order 1967 national input-output table published by the Bureau of Economic Analysis (BEA). Expansion to 404 sectors was achieved by additional disaggregation in the mineral industry sectors, of special interest for energy-related applications. Preliminary 404-sector 1972 I/O tables for the U.S., for California, and for selected California counties have been developed by Engineering Economics, Inc. and stored in the LBL computer system. An improved version of the U.S. I/O table, and a two-region model (California vs. the rest of the U.S.) are being developed by the LBL Regional Studies Group.

In the analysis of energy and other resources, I/O tables can be used to evaluate regional impacts of alternate development strategies. In late 1975, state I/O tables (1967, 404-sector) for the eight Rocky Mountain states, along with 87-sector summary tables, were sent by Henry Ruderman of LBL to Bill Byer of LASL.

(LBL TO LASL, LLL) COUNTY BUSINESS PATTERNS

County Business Patterns, published annually by the U.S. Bureau of the Census, contains data on the number of establishments, payrolls, and employment in U.S. industries, disaggregated by SIC (Standard Industrial Classification) code to the 4-digit level. The Bureau of the Census is obliged to withhold data whose disclosure would reveal confidential information about individual establishments.

Henry Ruderman of the LBL Regional Studies Group has 'filled in' the missing entries in County Business Patterns data, by using a maximum-likelihood method to estimate the most

0 0 0 0 4 6 0 0 8 9 7

I. APPENDIX C - INTERLABORATORY DATA EXCHANGES II
DATA FROM LBL

probable values of the missing data elements. The 'filled-in' data have been used at LBL to estimate 1972 U.S. employment by state (Ref. LBL.26), and to construct 1972 input-output tables for California and selected California counties. In Spring 1976 the 'filled-in' 1967 and 1972 national data by state were sent to Bill Beyer of LASL, and the 1972 California data by county were sent to Ken Haven at LLL.

1. APPENDIX C - INTERLABORATORY DATA EXCHANGES 12
DATA FROM LLL

1.5 DATA FROM LLL

(LLL TO ORNL) INDEX OF DATA BASES AND MODELS

In Spring 1975, the LLL Information Research Group produced a computerized compilation known as the National Index of Energy and Environmentally Related Data Bases. (Refs. LLL.2, LLL.4). The LLL Index, which contains detailed information on more than 8000 models and data bases, is available for interactive use on the unclassified LLL 6600 (G) computer. Alternatively, microfiche copies (about 15,000 pages including subject and author indices) are available from the LLL Information Research Group. In a typical recent application, users at PNL used a microfiche version of the Index to locate agricultural data required for the PNL Analysis and Assessment Program. The data base (which is at LBL) was accessed interactively, and selected data printed out on a terminal at PNL. The entire process took only a few hours.

In early 1976, a tape version of the LLL Index was sent to ORNL, where it will be made available on ORLOOK and on the nationwide RECON interactive search/retrieval system. (For more details, consult the IWGDE FY 1976 Annual Report, of which this report forms an appendix.)

1. APPENDIX C - INTERLABORATORY DATA EXCHANGES 13
DATA FROM ORNL

1.6 DATA FROM ORNL

(ORNL TO ANL TO LBL TO LASL) DUN AND BRADSTREET

The Dun and Bradstreet file at ORNL contains information on all of the industrial establishments listed in the U.S. and Canadian Dun and Bradstreet Reference Books as of 1972. (See Ref. ORNL.9). In early 1975, James Cavallo of ANL's Energy and Environmental Systems Division purchased the Dun and Bradstreet file from ORNL and subsequently sold a copy to the Regional Studies Group at LBL. (The Canadian data are not present on the file received at LBL.) The following items are included for each of approximately 1,100,000 establishments:

- Dun and Bradstreet Reference Number
- State Number
- County Number
- Place Code
- Business Name
- Address
- City
- State Name
- Chief Executive's Name and Title
- Phone Number
- 4-digit 1967 SIC Code
- ZIP Code
- Number of Employees

The file is sorted (a) by state; (b) within each state, by county; (c) within each county, by SIC Code. The data in the tapes do not represent full coverage of all establishments in the U.S.; the actual coverage is not known, and may vary from state to state or from industry to industry. Because the file contains confidential information on individual establishments, access is restricted to approved users, and only summary data may be released for publication.

At LBL, Deane Merrill experienced difficulty in reading the tapes received from ANL; the blocked binary records, which contain 15,000

1. APPENDIX C - INTERLABORATORY DATA EXCHANGES 14
DATA FROM ORNL

bytes each, are sufficiently long to produce occasional read errors on LBL's 9-track drives (for reasons related to I/O timing). Kiat Ang of LBL later unblocked the tapes at an IBM installation, and converted the tapes to a format suitable for printing on microfiche.

In addition, compressed binary tapes containing only SIC codes, county codes, and number of employees were produced. These tapes can be easily (and inexpensively) aggregated to yield 1972 employment estimates for any county and any SIC code. These data were used to produce 1972 input-output tables for California and for selected counties of California; in principle the same can be done for any other state(s) or county(s).

In October 1975 Deane Merrill sent one of the five reels received from ANL to Ellen Leonard at LASL. The tape was not used and was returned to LBL. Most likely LASL will prefer to access the LBL files directly, taking advantage of the pre-processing already done at LBL.

(ORNL TO ANL, BNL) U.S. CENSUS DATA

Data from the 1950, 1960, and 1970 U.S. Censuses, (including population by county, by sex, race, and age) were sent from ORNL to BNL. A subset of these data was sent from BNL to ANL. In addition, ORNL also sent directly to BNL a more complete set of data including housing data. A file containing adjusted 1970 county populations, by sex, age, and race, was sent from ORNL to BNL.

(ORNL TO LBL) IWGDE EXCHANGE STANDARD

In early 1976 Al Brooks of ORNL prepared a machine-readable document describing the proposed IWGDE Data Exchange Standard (Ref. ORNL.2). A current version is maintained on a permanent disk file of the ORNL PDP-10. In

1. APPENDIX C - INTERLABORATORY DATA EXCHANGES 15
DATA FROM ORNL

July 1976 a tape copy was sent by Karen Barry of ORNL to Deane Merrill at LBL. After conversion to ASCII print format, a copy was stored on the LBL PSS (Program Storage System), for inclusion as Appendix D of the IWGDE FY76 Annual Report. From LBL, the document may be directly accessed by a larger user community, including users of the ARPANET. The LBL version will be updated as new versions are received from ORNL.

As an ironic footnote, we shall describe how this document (the very essence of the IWGDE Exchange Standard!) was transmitted to LBL on tape. The first draft was written in ASCII, one record per line, and posed no problem. The second draft arrived with no format description and was assumed to be the same. After three days of frustrating effort by Deane Merrill and Bob Healey, Bob finally succeeded in 'cracking' the second version, which turned out to be blocked EBCDIC, with machine-dependent blocking factors written in binary. A more fitting testimonial to the need for the IWGDE exchange standard could not have been devised!

(ORNL TO LBL) 1970 CENSUS DEFINITION FILE

This file contains data base directories (dictionaries) for the 2nd, 3rd, 4th, and 5th counts (all files) of the 1970 U.S. Census. In Summer 1976 it was sent from Andy Loebel (ORNL) to Don Austin (LBL). For further information, contact Andy Loebel or Don Austin.

(ORNL TO LBL) 1970 CENSUS, FOURTH COUNT HOUSING

The Fourth Count Housing File is a portion of the 1970 U.S. Census not previously available at LBL. It contains data on housing facilities, number of appliances, etc. which are required by the Energy Use and Conservation group of the LBL Energy and Environment Division, for studies evaluating the impacts of proposed energy conservation alternatives. (Ref. LBL.6).

1. APPENDIX C - INTERLABORATORY DATA EXCHANGES 16
DATA FROM ORNL

Related studies are being performed by the LBL Regional Studies (Analysis and Assessment) Group for the California Energy Resources Conservation and Development Commission (CERCDC). The New York portion of the Fourth Count Housing file is also required for the Urban Atlas project being performed by LBL's Computer Science and Mathematics Department for the U.S. Bureau of the Census.

In June 1976, nine tapes containing the complete Fourth Count Housing Summary were sent from Andy Loebl of ORNL to Don Austin of LBL. They will be copied to the LBL MSS (Mass Storage System) for interactive random-access use, as was done earlier for other portions of the 1970 Census. (See Ref. LBL.10).

(ORNL TO SRL) 1970 POPULATION DATA GRID

This data base represents the reorganization of U.S. Census Bureau population data into arrays of rectangular sectors. Such reorganization facilitates data retrieval, simplifies calculation of population density and other analyses, and is essential for machine plotting of isopleths and isometric projections. (See Ref. ORNL.9). In 1975 these data were sent from ORNL to SRL. For further details, contact Andy Loebl (ORNL).

Draft Rev. 2, 6/10/76

DRAFT PROPOSAL
AN EXTENDABLE STANDARD FOR INFORMATION
INTERCHANGE ON MAGNETIC TAPE

Foreword

(This foreword is not part of the standard.)

The purpose of this standard is to establish machine independent formats for the transfer of a wide variety of hierarchically structured alphanumeric information on magnetic tape. The standard utilizes several of the ANSI standards and generalizes the ANSI Z39.2-1971 Standard for Bibliographic Interchange on Magnetic Tape as a vehicle for the interchange of a wider variety of information. The standard encompasses a model for structured fields which permits a definition of a different structure for each field, including an open-ended set of yet-to-be-defined data types and structures. The standard also anticipates the use of existing and future systems of the ISO 2709 type and provides the means to identify such systems and to include them as subsystems within this standard. It is inherent in the model for this standard that each logical record represent all of the information which is associated with an individual whose identity is persistent over the life of the data base and that this information be storable as defined

data elements in hierarchical storage structures. Each logical data record has a unique identifier field which has a prescribed tag (...001) and the field is located in a unique position. The higher levels of the structure are accessed through a system of tags and pointers. The structure can accept a planar relation and the non-cyclical directed subgraphs of chained linked data bases. If required, several planar relations can be collapsed on a common principal key to produce their hierarchical equivalent. Cyclical directed graphs expressing structure among individuals can be represented by including a successor (predecessor) field to contain identifier field entries which can be used to restore the original directed graphs.

The rationale of the standard is (1) to define a set of content-free standards for those portions of the standard which are solely computer-oriented, (2) to adopt a content-free version of ANSI Z39.2-1971 for the logical record structure, (3) to define a hierarchical structure as the storage structure for each logical record, (4) to define the means of codifying the type of data element stored in each field, and (5) to provide the means of transmitting a field-by-field description of the data base on the same magnetic tape.

The modus operandi to produce this standard was to use applicable portions of ANSI Z39.2-1971 verbatim deleting the unwanted restrictions and conventions for bibliographic information and then augmenting the remainder to produce the desired standard such that ANSI Z39.2-1971 and others of the ISO 2709 family can be a subset of this standard. This approach has resulted at times in unused control characters which are retained to maintain similarity to other systems.

An essential element of the standard is a data base description file that permits the designation of the defined subsystem which is to be invoked and if required, on a field-by-field basis, the designation of a specific occurrence of the many possible structural forms permitted.

The terminology of ANSI Z39.2-1971 has been retained wherever possible even though a more appropriate term for a general case might have been devised. In these cases the definition of the term has been augmented to reflect its broader meaning.

It is anticipated that this standard will result in automated, hardware-dependent software which will process all or part of the standard. It also assumes control card override or replacement of the data description file by the user when a data base requires said practice.

Draft Manuscript Note

The documentation for this standard is being maintained on a PDP-10/IBM 360 printer. Large brackets have been stylized by vertical bars and horizontal dashes.

1. Scope and Purpose

1.1. This standard defines a format which is intended for the interchange of information records on magnetic tape. It has not been designed as a record format for retention within the files of any specific organization, nor has it been the intent of the work group to define the content of individual records. The committee has attempted, instead, to define in this standard a generalized structure which can be used to transmit, between systems, records containing a wide and open-ended variety of information types and structures.

1.2. In designing the format, the committee has tried to achieve the goals listed in 1.2.1 through 1.2.4. It was recognized, however, that the goals were not completely compatible and that various trade-offs were required.

1.2.1. Hospitality. Hospitality to all kinds of information should be provided so that all types of information can be contained within the same record.

1.2.2. Hardware Independence. A format which can be used with a variety of digital computers should be defined.

1.2.3. Variety of Structure. The structure of all machine records should be basically identical. The data description should include such control information as may be required to specify unique characteristics of each field within a record. For any given class of records the components of the format may have specific meanings and unique characteristics.

1.2.4. Data Manipulation. The methods of recording and identifying data should provide for maximum manipulability at the field level leading to ease of automated conversion to other formats for various uses.

1.2.5. The standard should permit the recipient of a data base to read all records and to isolate data field contents at the highest level, to discern which data fields conform to those portions of the standard which have been implemented and to process those fields within his system.

1.3. Explanatory material which is not part of the standard but which will assist in its interpretation or implementation appears in footnotes.

1.4. Explanatory examples of the use of this standard are given in Appendix A. Commentary on the defined subsystem standards is given in Appendix B.

1.5. Standard Authority

This standard is open ended in its list of defined control parameters for:

- 1) Major Subsystem (3.2.3)
- 2) Character Code (3.2.4)
- 3) Field Control Character Concept (3.2.5)
- 4) DDF/Data Element Controls and Contents (3.5.3)

The authority for maintaining and extending these lists shall be the IWGDE and application for additions should be submitted to them. It is not the intent of this authority to proliferate subsystems, but rather to provide the capability to incorporate existing systems or new major systems established by other recognized standards authorities. Interim codes may be adopted by joint agreement among users, but such a method will not constitute a precedent.

1.6. This standard assumes the utilization of the following American National Standards:

American National Standard Code for Information Interchange, X3.4-1968

American National Standard Recorded Magnetic Tape for Information Interchange (800 CPI, NRZI) X3.22-1967

American National Standard Magnetic Tape Labels for Information Interchange, X3.27-1969, including its proposal revision, X3L5/506T (9-18-75) if adopted

1.6.1. Magnetic Tape Format. This standard adopts the ANSI X3.22-1967 as the magnetic tape format.

1.6.2. Magnetic Tape Label and Block Structure. This standard adopts the ANSI X3.27-1969, with its proposed revision X3L5/506T (9-18-75) if adopted, as the magnetic tape label and block (physical record) format. The maximum length of a block should be 2048 characters.

1.6.2.2. The physical block (record) format will be spanned variable records, type S, of the ANSI X3.27-1969 proposed revision X3L5/506T (9-18-75) Section 6.2.4 (see Appendix C).

1.6.2.3. In the event the ANSI X3.27-1969 is not revised per X3L5/506T, the spanning technique of Section 6.2 of the revision as described in Appendix C will be adopted.

1.6.3. Character Set and Codes

1.6.3.1. This standard adopts ANSI X3.4-1968 as its standard character set and codes. The control characters of this standard are reserved for their recommended uses.

1.6.3.2. It is anticipated that other character sets and codes will be required for special exchanges. Provision for the designation of such sets will be defined in the data description record.

2. Definitions

Definition of terms in common use has been considered unnecessary. (The experienced user may wish to omit this Section temporarily.) Terms which have a special meaning in the standard or which might be ambiguous are defined as follows:

base address of data. A data element whose value is equal to the character position of the first data field following the field terminator of the directory, where the specified origin (0) is the first character of the leader.

basic character. A character occurring in columns 2 through 7 of the Standard Code as specified in American National Standard Code for Information Interchange (ASCII), X3.4-1968.

Cartesian label. An array of identifiers formed by the Cartesian product of the elements of two (or more) label vectors and whose elements have the same order as the elements of the direct product such that if a and b are the label vectors, $a = (a(1) \dots, a(n))$ and $b = (b(1) \dots, b(m))$, then the Cartesian label, $a@b = (a(1)b(1), a(1)b(2), \dots, a(1)b(m), \dots, a(m)b(m))$ where $a(i)b(j)$ is an identifier of the i,j element of a corresponding data array.

character. See internal character.

communications format. See information interchange format.

data description file (DDF). A separate file which precedes the Data File (DF) and contains the

control parameters and data definitions necessary to interpret data records and data elements. The DDF is the first file other than labels and contains one record.

data element. A defined unit of information within a system.

data field. A variable field containing information.

data file. The second file which contains a collection of data records.

data subelement (subelement). A lower level data element contained within a hierarchically structured data element.

decimal digit. The decimal digits are the ASCII characters for 0-9 (i.e., 3/0-3/9).

delimited structure. A delimited structure is a storage structure composed of a collection of data strings separated by delimiters.

delimiter. A single character which serves as an initiator, a separator, or a terminator of individual data elements within a variable field. See ANSI X3.4-1968 for reserved delimiters (1/12, 1/13, 1/14, 1/15).

directory. An index to the location of the variable fields within an information record. The directory consists of entries.

entry. A fixed field within the directory which contains information about a variable field.

entry map. A data element which is used to indicate the structure of the entries in the directory.

external character. A graphic symbol which may be represented by one or a series of two or more internal characters.

field. A defined character string which may contain one or more data elements stored in a hierarchical structure. See also entry; variable field.

field terminator (FT). A character (1/14) used to terminate a variable field within a record. The

last variable field is terminated by a record terminator (1/13) and not a field terminator.

file. A set of related records denoted by a single name.

fill character. A character (space, ASCII character 2/0) used to fill areas in fixed fields which contain no data.

format. See structure.

format controls. A string of characters used to describe the format of a data field.

FT. See field terminator.

identifier field. A field containing a unique alphanumeric identification of a data record and its associated information content. This field must exist in each logical record and have a tag ...001.

information. A string of alphanumeric characters which when properly interpreted represents "intelligence." It includes both "words" as well as "numbers."

information interchange format. A format for the exchange, rather than the local processing, of records. (The terms "communications interchange format," "information interchange format," and "interchange format" are used interchangeably in this standard.)

interchange format. See information interchange format.

internal character. A pattern of bits of a predetermined length (depending on the system) treated as a meaningful unit. (The terms "internal character" and "character" are used interchangeably in this standard.)

label vector. A vector of identifying labels (i.e., "column" headings or "row" headings).

leader. A fixed field which occurs at the beginning of each record and provides parameters for the processing of the record.

logical record. A collection of fields terminated by a record terminator (1/13), including a leader,

a directory, and one or more data fields to be treated as a logical unit having a unique individual existence and identified by the identifier field.

numeric field. A numeric field is one which contains the ASCII characters 0-9, +, -, E or . in a combination representing an integer or real number. An integer-numeric field contains only the characters 0-9. A decimal-numeric field contains only non-integers. A mixed numeric field contains both integer and decimal representations.

padding character. A character (circumflex, ASCII character 5/14) used to pad out physical records on magnetic tape.

record. See logical record.

record length. A data element whose value is equal to the length (in characters) of the information record including the record terminator. It is used in some of the recognized subsystems (see Segment Control Word).

record terminator (RT). A character (1/13) used to terminate each record.

RT. See record terminator.

segment control word. Five characters which include a record continuation flag and segment lengths in character (see record length).

structure. The framework of fixed and variable fields within the information record.

subfield. A lower level field contained within a structured field.

tag. A series of characters in an entry used to specify the internal name or label of an associated variable field.

variable field. One in which the length of an occurrence of the field is determined by the length (in characters) required to contain the data stored in that occurrence. The length may vary from one occurrence to the next. This standard treats fixed length data fields as variable fields.

3. Information Interchange Format

3.1. Interchange Format

3.1.1. The magnetic tape format will consist of two files: (1) the data description file, (2) the data file, and (3) the required tape labels (see 1.6.2).

3.1.2. The data description file is the first file composed of a single record in the standard logical record format.

3.1.3. The data file is the second file composed of identified sequential records in the standard logical record format.

3.1.4. Logical Record Format. The logical record format consists of (1) a leader of 24 characters, (2) a directory of length $n \cdot l$ where n is the number of entries and l is the length of each entry, (3) a set of n variable length fields, each terminated by a field terminator (1/14). The last variable field is terminated by a record terminator (1/13). The interchange format of a logical record is schematically represented in Fig. 1.

3.1.5. Sequence of Logical Records. The data file is sorted into ascending collation sequence of the identifier field. When required for sorting, the field is padded on the right by spaces (2/0) if the field is alphanumeric and

on the left by spaces (2/0) if numeric characters only are used.

3.2. Leader

3.2.1. Schematic Representation. The leader is schematically represented in Fig. 2. The leader format of the DDF contains additional information in positions 5, 6, 11. The descriptions for these leader fields for the DDF are given above the line and those for the DF are given below the line. The DDF leader must be written in a character set for which the ANSI X3.4-1968 is a proper subset in order that the initialization of software can proceed properly.

3.2.2. Segment Control Word/Record Length.

(a) The segment control word is a five-character flag and character count in positions 0-4 as described in ANSI X3.27-1969 (X3L5/506T) Section 6.2 for spanned records.

(b) The record length is an alternative to the segment control word and is used in several existing systems (ANSI Z39.2-1971, UNISIST SC.74/WC/20, TID 4581 (R2), et al.). When the record length does not exceed 2048, it is compatible with the segment control word.

3.2.3. The major subsystem designator is a one-character code in position 5 for major defined

subsystems. This standard can conveniently accept as a subsystem any standard of the ISO 2709 family which conforms to the basic leader, directory, and variable data field format. When such a subsystem is invoked either by the DDF controls or by user control and override, the standard invoked takes precedence over the current standard. The currently defined codes are:

- E - this standard
- A - ANSI Z39.2 - 1971
- U - UNISIST WC.74/WS/20
- I - IAEA-INIS-9 (Rev. 2)
- T - TID 4581 (R2)
- M - MARC - MIF
- X - not defined, control cards required
 - others to be defined

3.2.4. The character set descriptor is a single character code in position 6 for the character set used as defined below:

- A - ANSI X3.4-1968
- T - TID 4581 (R2)
- I - IAEA-INIS-7 (Rev. 1)
- U - UNISIST WC.74/WS/20 ISO
- G - UNISIST WC.74/WS/20 GOST
 - others to be defined

Non-ASCII character sets must be processed by prior initialization by external controls.

3.2.5. The field control character count (in position 11) is the number of characters of the data field descriptor devoted to numeric data element type and structure codes and delimiters. (See Section 3.5.3.2.)

3.2.6. Base Address of Data. A data element in character positions 12-16 consisting of five decimal digits which give the starting position of the first data field and is equal to the combined length (in characters) of the leader and directory (including the field terminator at the end of the directory).

3.2.7. Entry Map. The entry map is shown in Fig. 3. The entry map describes the subfield lengths for directory entries (see 3.3.1).

The entry map is a data element in character positions 20-23 consisting of four decimal digits. Each decimal digit recorded corresponds sequentially to each portion of the entry. Character position 20 in the entry map indicates the length, m , (in characters) of the "length of field" portion of each entry in the directory; character position 21 indicates the length, n , (in characters) of the "starting character position" portion of each entry; character position 22 indicates the length, t , of the tag portion of each directory entry. Character position 23 is undefined and is available for future use.

3.3. Directory.

The directory consists of a series of fixed fields (hereafter referred to as entries). The directory ends with a field terminator (1/14). The directory must contain one entry for each subsequent variable field.

3.3.1. Entries. The structure of each entry in the directory is shown in Fig. 4. Each entry consists of $m+n+t$ characters. Each entry must contain a tag, length of field, and starting character position, and must correspond, unambiguously, to a specific variable length data field. The tag, length of field, and starting character position must be in that sequence.

3.3.1.1. Tag. The tag is a data element of $t < 7$ characters which does not lie below "...0..." in collation order.

3.3.1.2. Reserved tags. Tags ...000 through ...009 are reserved tags. Tag ...000 shall occur in the DDF only and shall designate the data base title. Tag ...001 is reserved for the record identifier field and must occur in each record. Tags ...002 through ...009 are reserved for future use.

3.3.1.3. Length of Field. The length of field in the entry is the length (in characters) of the variable field to which it corresponds. The length of field includes the terminator. It is expressed as a decimal number.

3.3.1.4. Starting Character Position. The starting character position is the character position of the first character in the variable field referenced by the entry. It is given relative to the base address of data (that is, the first character of the first variable field following the directory is numbered 0).

3.3.2. Sequence of Entries. The entries in the directory shall be recorded in ascending collation sequence for unstructured records. Multiple entries of the same tag are permitted.

3.4. Variable Fields - General.

3.4.1. Sequence of Variable Fields. The variable fields, except for the fields associated with tags ...000 and ...001, need not occur in the same sequence as the corresponding directory entries.

3.4.2. The variable field corresponding to tag ...000 shall always be the first field of the DDF.

3.4.3. The variable field corresponding to tag ...001 shall be the simplest structure which will permit the

meaningful identification of the logical record and shall be the first field of the DF.

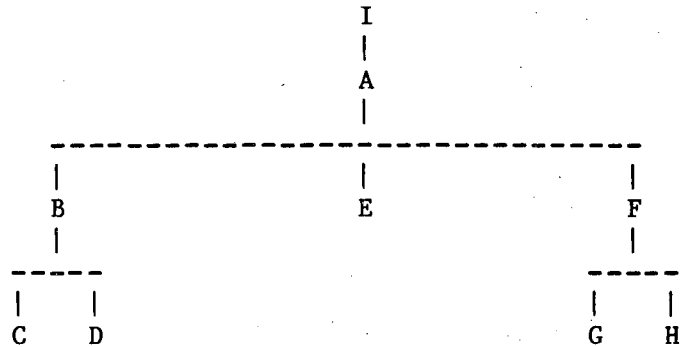
3.5. Variable Fields - Data Description File.

The variable fields of the DDF contain information which defines and describes the corresponding (i.e., same tag) data fields of the DF and provides control parameters for automated processing. These fields have a specified format modified only by the DDF leader parameters. All variable fields end with a field terminator (1/14) except the last, which ends with a record terminator (1/13).

3.5.1. Hierarchical Logical Records.

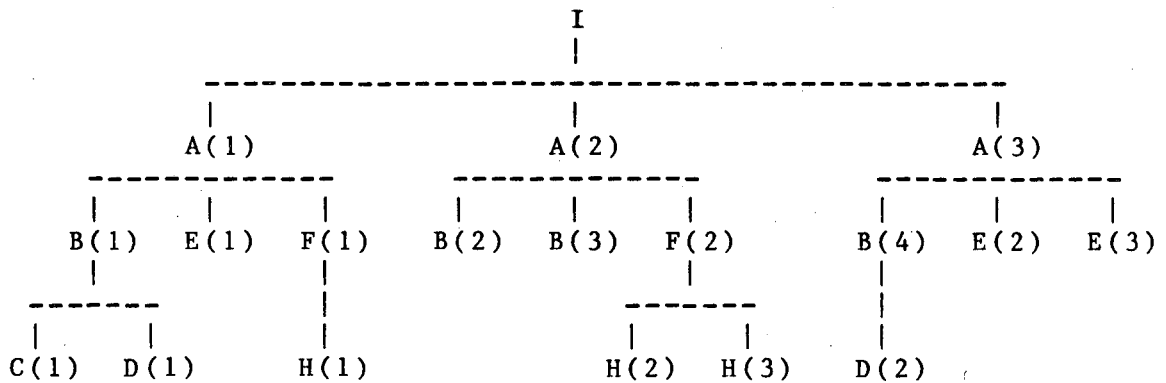
This section describes the means by which the inherently linear structure of the tagged variable fields can be reconstituted into a hierarchical structure. A hierarchy is to be designated by two pieces of information 1) the ordered pairs of node identifiers (i.e. tags) of the generic structure in the top down order of parent-offspring and 2) the list of node identifiers (i.e. tags) in the top down, left branch first (TDLBF) order.

The generic structure of a logical record



has the set of ordered pairs of identifiers (rather than tags for simplicity) IA, AB, BC, BD, AE, AF, FG, FH, where I is the unique record identifier. The TDLBF order of this structure is - IABCDEFHG (alphabetic order incidental) and each occurrence of a node is unique.

A specific instance of the previous generic structure might be:



where I is the unique record identifier. Its TDLBF order is IA(1)B(1)C(1)D(1)E(1)F(1)H(1)A(2)B(2)B(3)F(2)H(2)H(3)A(3)B(4)D(2)E(2)E(3). Note that only I is unique and that several instances of the field A, B, C etc. may occur in which case they are designated by repetition of the same tag.

3.5.1.1. Record Hierarchy Controls. The data base name field (tag ...000) shall be augmented by the delimiter, 1/15, followed by a succession of predecessor - successor tag pairs which determine the generic hierarchical structure of a logical record. All tags must be included and the root tag shall be tag ...001, the unique identifier tag.

3.5.1.2. Order of Directory Tags. The order of the directory tags shall be determined by tracing the original hierarchy using the top down left branch first rule starting with the root tag ...001.

The DDF directory contains only one instance of each tag and its variable field contains an instance of the defined data element controls. The order of the DDF directory is the TDLBF order of the generic hierarchy.

The DF directory may contain multiple identical tags as the specific instance of the generic structure requires and each variable field will contain an instance of the data element defined in the corresponding DDF field.

3.5.2. Unstructured Data Elements.

The following are specifications for DDF leader position 5 = "E" and position 11 = "0".

3.5.2.1. Field Controls. There are no field controls for this variation.

3.5.2.2. Data Field Description. The data field description starts in character 0 of the data field and is a simple text string constituting the name of the data field.

3.5.2.3. Data Field Formats. The data field format for this variant is unstructured alphanumeric text. An unstructured field may contain external delimiters specified in some "style" authority, but they are not recognized as delimiters in this variant.

3.5.3. Structured Data Elements.

The following are specifications for DDF leader position 5 = 'E' and position 11 = '8'.

The data structures defined are simple, vector and array containing data types text, integer, real, and mixed (text, integer, real). The data description field contains control information, delimiters, field name, subfield labels (vectors, arrays) and data field format information.

3.5.3.1. Data Field Names, Labels, and Format. The conventions for the names, labels, and formats for the defined data structure/types are given in the following tables and further defined in this section. Table I defines the use of delimiters. Table II defines the format of the data description field.

Table I. Delimiters and Their Uses.

| Delimiter | Symbol* | File | Use |
|-----------|---------|--------|---|
| 1/12 | % | DDF | To delimit data element labels within a vector label. |
| 1/13 | \$ | DDF,DF | Record terminator. |
| 1/14 | # | DDF,DF | Field terminator. |
| 1/15 | @ | DF | To delimit subfields in a data field. |
| | | DDF | To pre- and post-delimit an optional field name and a vector label. |
| | | DDF | To delimit a vector label within a Cartesian label. |
| | | DDF | To predelimit the hierarchy controls in field ...000. |
| | | DDF | To predelimit format controls. |

*The ASCII character is not printable; the symbol given here is used throughout the manuscript for readability.

Table II

| Structure/Type | Control | Name/Label/Mode/Format* |
|----------------|---------|---|
| Simple Text | 00 00 | -- |
| Simple Integer | 01 00 | - 'Label' # |
| Simple Real | 02 00 | -- |
| Vector Text | 10 00 | -- |
| Vector Integer | 11 00 | - @ Name @ 'Vector Label' @ 'fctr' # |
| Vector Real | 12 00 | -- |
| Vector Mixed | 13 00 | @ Name @ 'Vector Label' @ 'fctr' # |
| Array Text | 20 00 | -- |
| Array Integer | 21 00 | - @ Name @ 'Cartesian Label' @ 'fctr' # |
| Array Real | 22 00 | -- |
| Array Mixed | 23 00 | @ Name @ 'Cartesian Label' @fctr# |

*For vector and array descriptions, null names, labels, vector labels and format controls must be used to insure automated deciphering; the entire vector or Cartesian label may be null, i.e., @name@@fctr#.

where

signifies optional presence,

'Label' is a character string (not delimited),

'Name' is a character string, pre and post delimited by @,

'Vector label' is an internally delimited (%) set of simple labels corresponding to the set of subfields in the data file and takes the form, label1%label2%... @ labela%labelb%... @....,

'Cartesian label' is a set of labels designated by vector labels forming an appropriately delimited (@) Cartesian product having an order corresponding to the set of subfields in the data file,

'fctr' is a string of characters defining the format of the data field.

3.5.3.2. Field Controls and Delimiters. The field controls will consist of four numeric characters followed by up to four printable delimiters in positions 0-7. Positions 0-1 will be interpreted as a structure-type code and position 2-3 is reserved as an auxiliary control for later definition and use. The four printable delimiters are reserved for later definition and use with delimited structures.

3.5.3.3. Names, Labels and Format Controls

The use of names, labels and format controls are described below:

- a) Name - an optional collective title for structured data (vector or array).
- b) Label - an optional title for simple atomic data. Where the atomic fields form a vector, the label is in the

form of a corresponding vector label. Where the atomic fields form an array, the label is in the form of a corresponding Cartesian label which when expanded by the defining conventions forms an array of labels corresponding to the data array. The vector labels of a Cartesian label provide row and column headings for the appropriate cross section of the array.

- c) Format Controls - the format controls designate the character by character structure of the data field. The format controls are optional for data types -- text, integer and real -- where the lack indicates delimited data. The format string is required to determine the sequence and type of the subfields in a mixed data field and to determine the field width in non-delimited fields.

The format controls are delimited by @(1/15) and the field terminator #(1/14) and takes the form

(X|X(*)|X(n)|mX(n)|k(mX(n),...),... ,...)

where

A signifying text,
 X = I signifying integer,
 R signifying real,

* = a subfield terminal delimiter,

n is a positive integer specifying field width,

m,k are positive integers signifying repetition,

... implies repetition with optional nesting.

| implies an alternate choice.

The use of the format controls will be governed by the following rules:

- a) The order of fields and their types specified with format controls shall correspond to the data field when the format controls are traversed from left to right expanding the nested terms from the left. If the data field is not exhausted the format will repeat from the left hand parenthesis.
- b) X - implies delimiting of the DF fields by the delimiter, @.

- c) X(*) - implies the presence of the character, *, as a terminating delimiter for the corresponding data field.
- d) Data fields for R-type will completely specify the real number in the form

$$+|- \dots ddd.dd\dots E|E+|E- dd$$

and the decimal point if missing will be assumed at the right of the least significant digit.

3.5.3.4. Dimension, Extent and Order of Arrays. The extent (number of elements) of a vector is known through its label, its format, or by the delimiting of a specific instance. The dimension and extent of an array is known from the Cartesian label. In the absence of the Cartesian label, the data field will be proceeded by 1) a positive integer giving the dimension of the array and 2) a series of positive integers giving the extent of each dimension. The order of an array will be row-major order. This dope vector will be delimited by @.

3.6. Variable Fields - Data File

3.6.1. The variable fields of the Data File contain the actual information to be transmitted. Each variable data field is an instance of the structure defined by the DDF data field with the corresponding tag. The subfields contain the data subelements corresponding to subtitles contained in the corresponding DDF subfields. In the

delimited mode, null subelements (i.e., successive delimiters) must be used to indicate missing subelements.

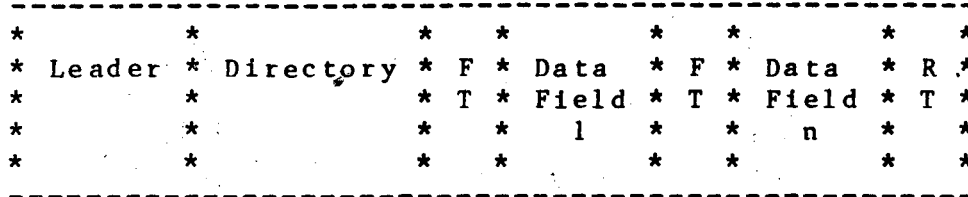
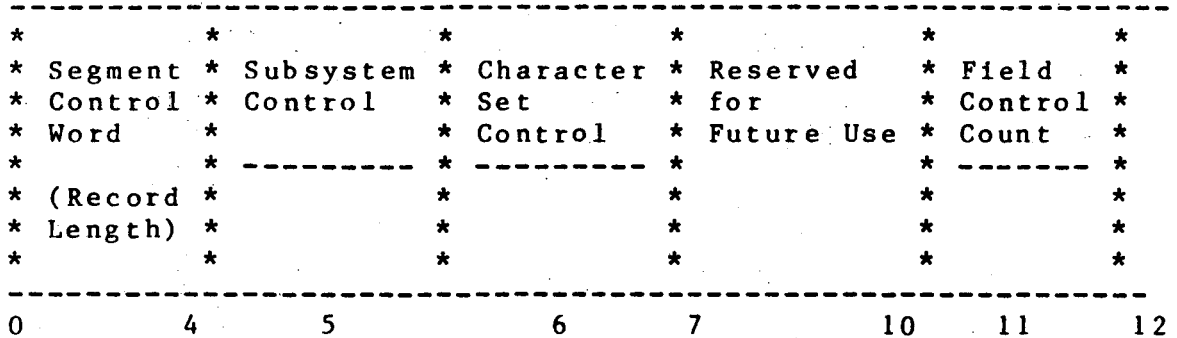


Fig. 1. Schematic Representation of the Logical Record of the Interchange Format.



continued

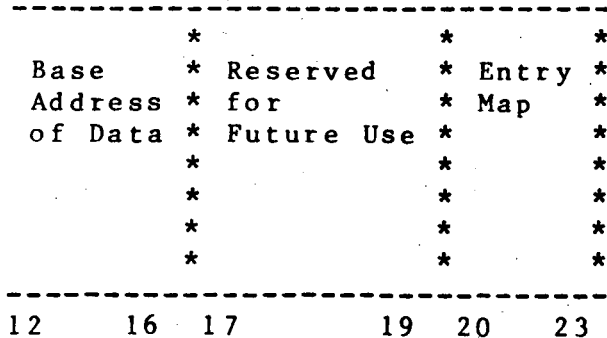


Fig. 2. Schematic Representation of the Leader

- c) X(*) - implies the presence of the character, *, as a terminating delimiter for the corresponding data field.
- d) Data fields for R-type will completely specify the real number in the form

+ |- ...ddd.dd... E|E+|E- dd

and the decimal point if missing will be assumed at the right of the least significant digit.

3.5.3.4. Dimension, Extent and Order of Arrays. The extent (number of elements) of a vector is known through its label, its format, or by the delimiting of a specific instance. The dimension and extent of an array is known from the Cartesian label. In the absence of the Cartesian label, the data field will be proceeded by 1) a positive integer giving the dimension of the array and 2) a series of positive integers giving the extent of each dimension. The order of an array will be row-major order. This dope vector will be delimited by @.

3.6. Variable Fields - Data File

3.6.1. The variable fields of the Data File contain the actual information to be transmitted. Each variable data field is an instance of the structure defined by the DDF data field with the corresponding tag. The subfields contain the data subelements corresponding to subtitles contained in the corresponding DDF subfields. In the

delimited mode, null subelements (i.e., successive delimiters) must be used to indicate missing subelements.

```

-----
*           *           *   *           *   *           *   *
* Leader * Directory * F * Data * F * Data * R *
*         *         * T * Field * T * Field * T *
*         *         *   *   l   *   *   n   *   *
*         *         *   *           *   *           *   *
-----

```

Fig. 1. Schematic Representation of the Logical Record of the Interchange Format.

```

-----
*           *           *           *           *           *
* Segment * Subsystem * Character * Reserved * Field *
* Control * Control   * Set       * for     * Control *
* Word    *           * Control  * Future * Count  *
*         * ----- * ----- *         * ----- *
* (Record *           *           *           *           *
* Length) *           *           *           *           *
*         *           *           *           *           *
-----
0         4         5         6         7         10        11        12

```

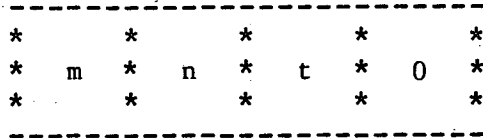
continued

```

-----
Base      * Reserved * Entry *
Address   * for     * Map  *
of Data  * Future *
*         * Use   *
*         *     *
*         *     *
*         *     *
*         *     *
-----
12       16  17       19  20  23

```

Fig. 2. Schematic Representation of the Leader



m = length (in characters) of the "length of field" portion of each entry in the directory
n = length (in characters) of the "starting character position" portion of each entry in the directory
t = length (in characters) of the "tag" portion of each entry in the directory
0 = undefined; reserved for future use

Fig. 3. Entry Map.

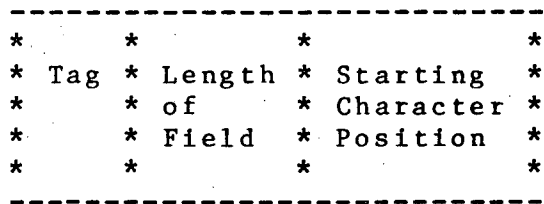


Fig. 4. Structure of Each Entry in the Directory

Appendix A

Explanatory Examples

A large portion of technical information subject to interchange can be transmitted as data elements in simple hierarchical structures - single items, simple lists (vectors) or multi-dimensional arrays. More complicated relationships inherent in the data can almost always be expressed in hierarchical form for the purpose of exchange. The simpler fields have been defined to permit progressive implementation and to provide simple more efficient processing paths for frequently used, simple structures.

Hierarchical structures can represent collections of information elements which at any level can represent either (1) an information vector in which the order of occurrence of elements is of importance and each element is a distinct and named entity or (2) an information set in which order is not important (and replication must not occur) and a name is applicable to the entire set but not to elements (except for the purpose of enumeration). A set (or vector) may consist in turn of a collection of vectors or sets. In either case the storage structure may be hierarchical.

A.1. Data Element Examples

This section contains several examples of data description fields and the corresponding data fields. Since

the permitted combination of structure, type, names, labels and modes is very large only a representation sample of each variation is given. Throughout the examples the following printable characters have been substituted for reserved delimiters or characters:

% - Sublabel delimiter (1/12)

- Field terminator (1/14)

\$ - Record terminator (1/13)

@ - Subfield delimiter (1/15)

& - Fill character (1/15) for unused delimiters

A.1.1. Unstructured Data Elements (DDF Controls 5/11=E/0)

| File | Variable Field Contents |
|------|-------------------------|
| DDF | AUTHOR# |
| DF | DOE,JANE# |

A.1.2. Structured Data Elements (DDF Controls 5/11 = E/8)

A.1.2.1. Simple Fields

| Example | Type | Variable Field Contents | |
|---------|---------|-------------------------|-----------|
| | | DDF | DF |
| 1 | Text | 0000&&&&NAME# | DOE,JANE# |
| 2 | Integer | 0100&&&&AGE# | 18# |
| 3 | Real | 0200&&&&GPA# | 3.46# |

A.1.2.2 Vector Fields

1) Delimited Text Vector with Name and Subfield Labels

```
DDF      1000&&&&@MAILING ADDRESS@NAME%STREET%CITY%
          STATE%ZIP@#
DF       DOE, JOHN@1010 MAPLE ST.@OSHGOSH@OHIO@12345#
```

2) Formatted Integer Vector with Name and Subfield Labels

```
DDF      1100&&&&@POPULATION@CY60%CY65%CY70%CY75@(4I6)#
DF       765432987345903231897654#
```

3) Delimited Real Vector with Name (no labels)

```
DDF      1200&&&&@GRAINS@@#
DF       3.46@2.47@11.94E-4#
```

4) Formatted Mixed Vector with Name (no labels)

```
DDF      1300&&&&@LIVE STOCK@@(A(,),I5,R5)#
DF       PIGS,0274437.46STEERS,1776447.84#
```

A.1.2.3. Array Fields

1) Delimited Text Array with Name and Cartesian Label

```
DDF      2000&&&&@PROPERTIES@GOLD%SODIUM% COPPER@DENSITY
          %COLOR%ACTIVITY@#
DF       HIGH@YELLOW@INERT@LOW@GREY@HIGH@MEDIUM@REDDISH
          @LOW#
```

Equivalent Table

| | DENSITY | PROPERTIES COLOR | ACTIVITY |
|--------|---------|---------------------|----------|
| GOLD | HIGH | YELLOW | INERT |
| SODIUM | LOW | GREY | HIGH |
| COPPER | MEDIUM | REDDISH | LOW |

2) Formatted Integer Array (no name, no label)

```
DDF      2100&&&&@@@ (9I2) #
DF       2@3@3@124788334672441921#
```

3) Delimited Real Array with Name

```
DDF      2200&&&&@UNIT MATRIX@@#  
DF       2@3@3@1.@0.@0.@0.@1.@0.@0.@0.@1.#
```

4) Formatted Mixed Array with Name and Column Label Vector

```
DDF      2300&&&&@TABLE II@@METAL%DENSITY%COLOR%  
ACTIVITY@(A($),R4,A(,),R4)#  
DF       GOLD$14.8YELLOW,-1.3SODIUM$0.63GREY,4.81  
COPPER$10.6REDDISH,.043#
```

Appendix B

Comments on Major Subsystems
(to be completed)

Appendix C

Section 6.2.4. ANSI X3.27-1969
Proposed Revision X3L5/506T (9-18-75)6. STRUCTURE OF BLOCKS
6.1 Character Code in Data

The data in each record is recorded using only characters of the code table specified in ANSI X3.4-1968. Either a seven- or eight-bit code version structured as specified in the ANSI Recorded Magnetic Tape for Information Interchange standard may be used.

6.2 Block Formats

6.2.1 Blocking Records

No explicit indication of the boundaries between records is required. There is an integral number of records in a block for fixed-length records (P) and variable-length records (D). There is an integral number of segments in a block for spanned records (S). Truncated and varying length blocks are permitted. On input, it is assumed that the actual number of characters that have been read in is known to the system.

6.2.3 Not Applicable

6.2.3 Not Applicable

6.2.4 Spanned Records (S)

A segment control word (SCW) precedes each segment. The SCW consists of five characters. The first character of the SCW is called the Spanning Indicator. It has the following meanings:

- 0 record begins and ends in this segment
- 1 record begins but does not end in this segment
- 2 record neither begins nor ends in this segment
- 3 record ends but does not begin in this segment

The segment length (i.e., the number of characters it contains) is expressed as a decimal numeral occupying the next four character positions of the SCW. The segment length includes the length of the SCW.

- 6.2.4.1 No explicit record control words are defined for spanned records.
- 6.2.4.2 Records may span volumes.
- 6.2.4.3 Record length is unbounded in that there is no limit to the number of segments in one record.
- 6.2.4.4 There is only one segment of the same record in a block.
- 6.2.4.5 The segments of a record are written in consecutive order (i.e., segment n+1 is written immediately following segment n) and do not have segments of other records interspersed.
- 6.2.4.6 Examples of spanned records appear in Figures 9 through 12.

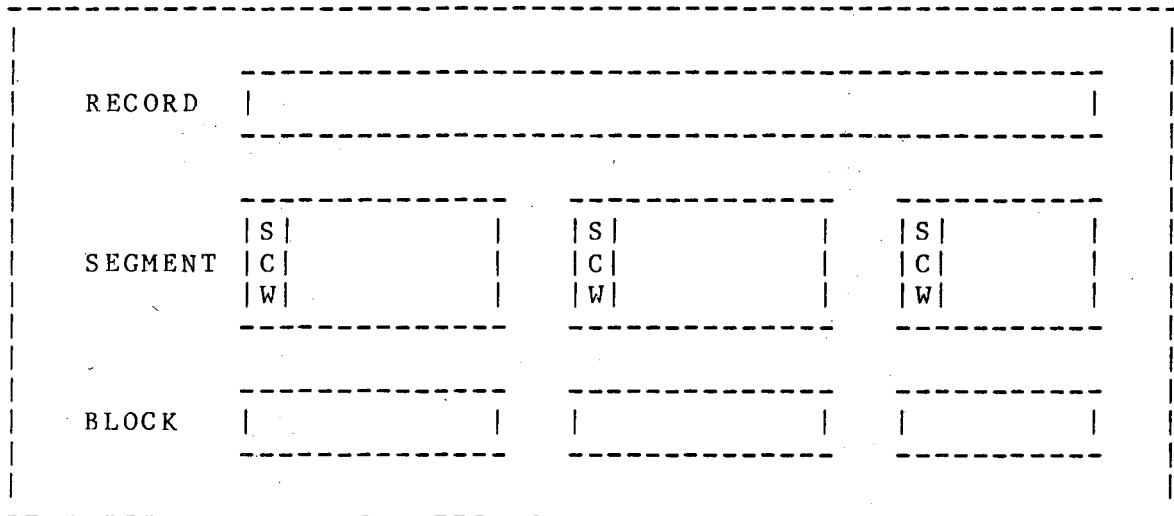


Figure 9. Spanned Record (S), Unblocked

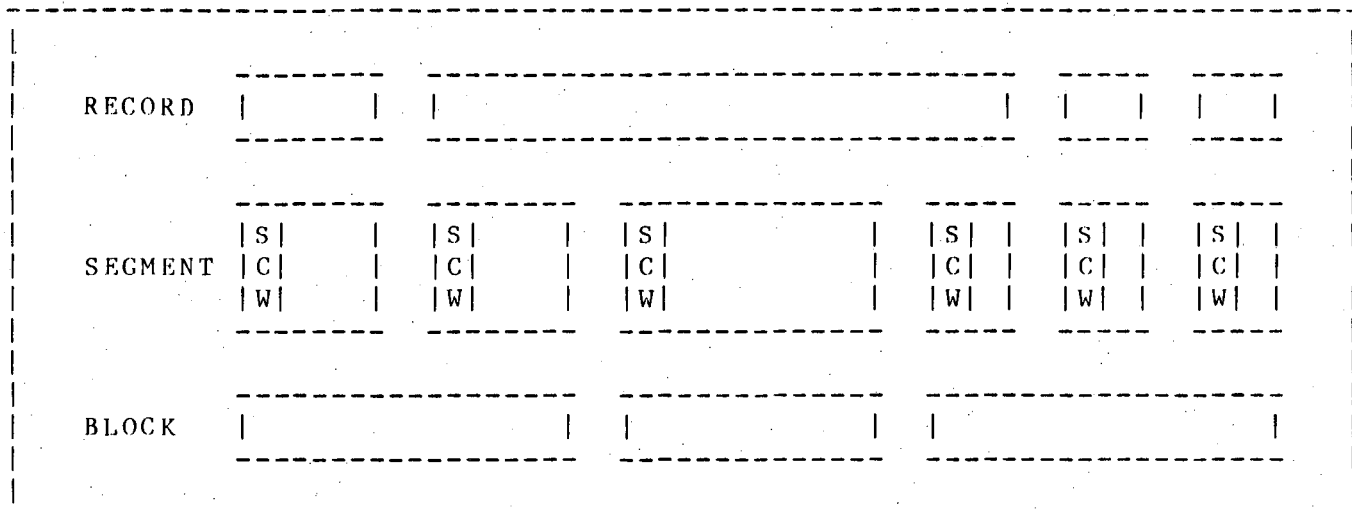
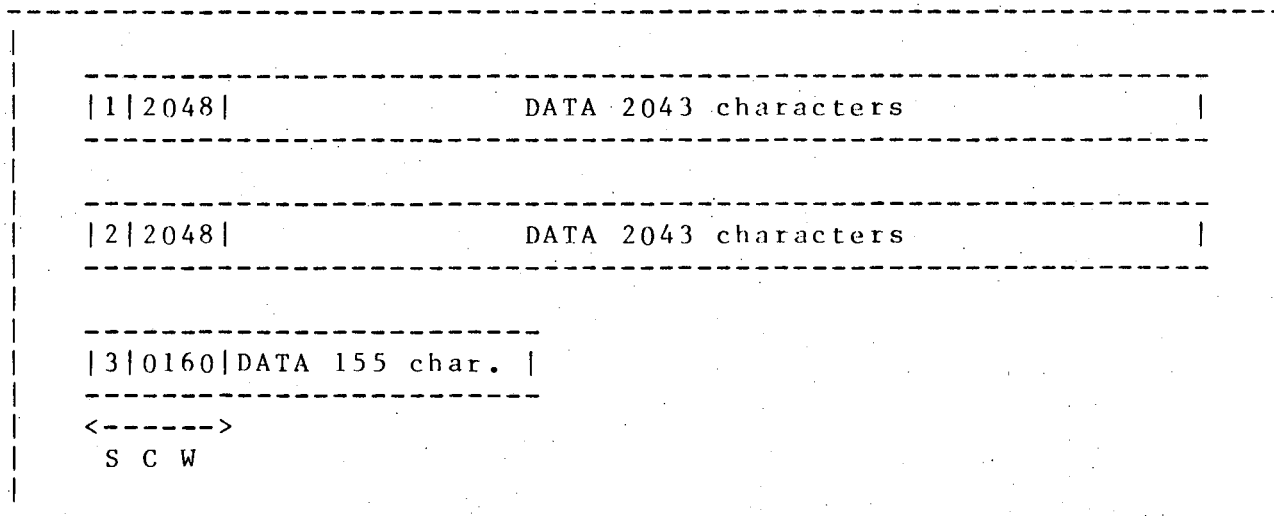


Figure 10. Spanned Records (S), Blocked.

Figure 11. Example of Unblocked Spanned Record
Length of record is 4241 characters. Each
line represents a block.

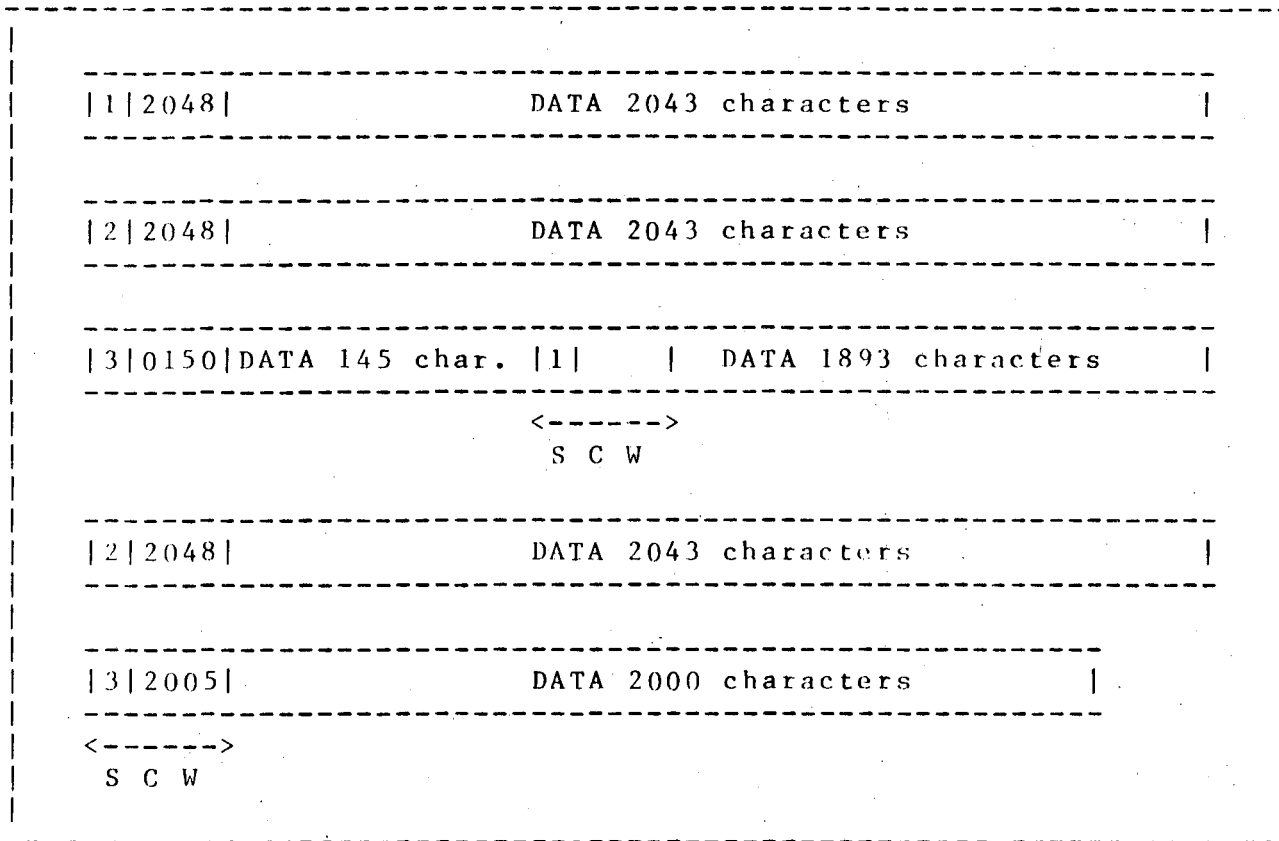


Figure 12. Example of Blocked Spanned Records
 Record #1: Length of record is 4231 characters.
 Record #2: Length of record is 5936 characters.
 Each line represents a block.

This report was done with support from the United States Energy Research and Development Administration. Any conclusions or opinions expressed in this report represent solely those of the author(s) and not necessarily those of The Regents of the University of California, the Lawrence Berkeley Laboratory or the United States Energy Research and Development Administration.

TECHNICAL INFORMATION DIVISION
LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720