COMPUTERIZED DATA SYSTEMS

UNIVERSITY OF MINNESOTA

TASK FORCE NEWSLETTER

Vol. I, No. 1
February 1971

* * * * * * * * * * * * * * * * * * * *

1. The Task Force on Computerized Data Systems

A. Organization

The Task Force on Computerized Data Systems is a sub-group of the University Advisory Committee on University Computer Services (UACUCS).

In order to deal with a number of important short-and long-range problems facing University Computer Services: some related to the University's own needs; and some related to the recently developed Statewide planning guidelines in Minnesota Higher Education, UACUCS established a number of sub-committees and Task Forces.

The general purposes of task forces are:

1. To develop proposals for specific facilities, services, staffing, and funding for computing activities which need to be established at the University for internal or statewide use.

2. To monitor ongoing computer activities of a particular kind at the University; to identify needs and problems concerned with that kind of computing, and communicate these to sub-committees, the Advisory Committee, the Director of UCS, or to an appropriate operations manager; and to consider matters of short-range operating policy, consistent with overall policies established by University Computer Services.

B. Mission

The Task Force on Computerized Data Systems has been assigned the following mission:

"to study and recommend policies and activities for instruction, research, and public service which involve the use or development of non-archival data bases and information systems on all-University computing facilities. Many such activities may be statewide or national in scope."

C. Organization and Sub-committees

The Task Force on Computerized Data Systems is comprised of the following individuals:

John S. Hoyt, Jr., Chairman	373-1225
Frank Boddy	373-2966 or 3583
Thomas Anding	373-3724
Bruce Boraas	373-7714
Bill Craig	376-7232
Van Mueller	373-5836
Luther Pickrel	373-3001
Glenn Brudvig	373-5581
Gene Eidenberg, ex officio	
Frank Verbrugge, ex officio	

In order to respond to the mission with which it has been charged the Task Force has established three sub-committees--Data Bases, Users, and Systems. It has also determined, based on a recommendation of Dr. Frank Verbrugge, that a periodic newsletter circulated within the University Community has a potential value as a communications link with all University faculty and staff whose interests are related to the use of computer services data systems.

The remainder of this newsletter is comprised of the initial reports of the sub-committees, some items of general news, and a form to be returned which is designed to add additional names to the newsletter mailing list. Please make a special effort to return that form with a list of additional names--we will eliminate duplicate names that are received.

2. Report of Users Sub-committee

D. The Users Sub-committee has established as its objectives:

1. The identification and establishment of guidelines for present and future utilization of University computer facilities; and

2. To carry out an inventory of users of University computer facilities which will serve as a basis for classification of types of users showing areas of complementarity and conflicts in their objectives.

It has also tentatively established the following general Classification of Computer Users:

1. Internal (University) users where software programs and/or data files are developed and are open (available to other users) for general use.

2. Internal users where programs and data are closed (restricted due to confidential nature of files or the limited applicability) to general use.

3. External (Non-University) users with open programs and/or files.

4. External users with closed programs and/or files.

3. Report of Data Base Sub-committees

E. The sub-committee on Data Bases has outlined the following tasks for itself.

1. Inventory users on campus to determine what machine readable (tape, card or disk only) data bases exist on campus and what data bases are needed. We will send a questionnaire to each center, department and division head on campus with a covering letter from the Moos appointed committee. I would hope that we can generate a list of other individuals who should receive a questionnaire at the TF meeting on Wednesday.

2. Define what is meant by "a non-archival data base."

3. In cooperation with the systems sub-committee, define a set of data standards for the University.

4. Report of Systems Sub-committee

F. The primary objective of the Systems Sub-committee is to evaluate and propose the type(s) of software and hardware required for a standard, all University, data base handling system. It is felt that some standardized method of file accessing should be implemented to:

1. Allow cross file studies

2. Provide more efficient data handling capabilities

We note that preliminary guidelines for the development of such a system have been produced by the Data Base Task Group of the CODASYL (Conference on Data Systems Languages) Programming Language Committee. The role of this group is specifically to define extensions to COBOL to handle data bases. The recommendations are general, however, and might be incorporated using any host language.

With this in mind, the Systems Sub-committee defined the following four tasks for itself.

1. Review the CODASYL standards (as outlined in the CODASYL Data Base Task Group Report) for applicability.

2. Recommend features to be incorporated in any all-University data management system.

3. Investigate and recommend use of a standardized data structure(s). (Based on generality and efficiency.)

4. Investigate and recommend development and/or purchase at once of more access software systems to implement the above.

5. Items of General Interest

G. CODASYL Summary

Summary: October 1969 Report of the CODASYL Data Base Task Group: This report (available for only $4.00 from Department C, Association for Computing Machinery, 1133 Avenue of the Americas, New York, New York 10036) consists of a proposal for a data description language and a data manipulation language. The success which the CODASYL group has had in promoting COBOL as a standardized business language augurs well for the future of a standardized data base system recommended by the Committee.

The data description language provides a means for describing data bases. The data manipulation language is a language which, when associated with a host language (such as COBOL, PL/1, ALGOL, FORTRAN....) allows manipulation of data bases described by the data description language. The specification of separate data description and data manipulation languages is significant in that it allows the data bases described to be independent of the host languages used in processing the data. Of course, for this to be possible, the host language processors must be able to interface with such independent descriptions of data.

Two general concepts presented in the report appear to be of prime importance. First: the separation of data description and data manipulation languages should allow a given program to handle data residing in any given type of file structure. Thus, a given file would be designated as a sequential structure, a tree structure, or a network structure. Hopefully, the data manipulation language (or body of sub-routines) would make it possible to provide efficient retrieval of data in each type of file structure. By declaring one file to be sequential and a second to be a network structure, a user could perform cross-file studies on vastly different data sets. In addition, it would be possible to use the most efficient file structure for each data base.

A second recommendation is that the data manipulation facilities be embedded in a host language (probably FORTRAN and COBOL at the U of M). We feel it essential that data base handling capabilities be embedded in a language which is suitable for sophisticated analysis procedures. The other alternative of writing a special language which allows a few simple statistical or arithmetical operations to be performed on the data seems a poor alternative.

The objective of the Data Base Group (of CODASYL) in developing its proposals was to make it easier and more efficient for programmers to store and retrieve data in secondary storage by providing features which:

- allow data to be structured in the manner most suitable to each application, regardless of the fact that some or all of that data may be used by other applications--such flexibility to be achieved without requiring data redundancy.

- allows more than one run-unit to concurrently retrieve or update the data in the data base.

- provide and permit the use of a variety of access methods against an entire data base or portions of a data base.

- provide protection of the data base against unauthorized access of data and from untoward interaction of programs.

- allow the user to plan and implement his system as if he had a virtual memory at his disposal.

- provide the Data Base Manager with the capability to control the physical placement of data.

- allow the declaration of a variety of data structures ranging from those in which no connection exists between data elements to network structures.

- allow the user to interact with the data while being relieved of all of the mechanics of maintaining the structural associations which have been declared.

- allow programs to be as independent of the data as current techniques will permit.

These features, then, provide both generality and flexibility and allow the building and manipulation of structures as complex as necessary for a given application.

Such complexity is not spurious. It allows concise and efficient problem-modeling in a way that no single technique could. In any one application, many different techniques can be used effectively to help in solving the various aspects of the overall problem. The Data Base Task Group approach allows the use of appropriate techniques for each aspect, while providing the discipline for interlocking the various parts into a unified whole.

It is important to note that the Data Base Task Group's proposals are oriented to the programmer. Specifically, the DML is a language for programmers. It is not an inquiry language intended for the non-programmer.

The Data Base Task Group recognized the need for an inquiry language, but considers it essential that such a language be capable of interacting with the same data bases that are developed for routine, day to day processing. Because of this, specification of an inquiry language has been deferred in favor of providing the tools required by programmers to build and maintain "data bases that are available to, and suitable for, processing by multiple applications." The proposals in this report, however, provide a solid foundation for building an inquiry system.

While the feasibility of implementing the detailed specifications (as outlined in this preliminary report or in any future final reports of the CODASYL group) is open to question, we feel that the spirit of the CODASYL data base task group must be followed in the development of any general data handling system. The Systems Sub-committee of the Data Systems Task Force strongly recommend that interested individuals obtain the report and read the first two sections (about 50 pages).*

H. 1970 Population Census Tapes Status Report

1. First Count Summary Tapes

The first count summary tapes for the states of Minnesota, Montana, North Dakota and South Dakota are now available in the MAPS data base. These files contain the data items of complete-count population and housing data, summarized for enumeration districts within the census tract and for the state, counties, congressional districts, places, and minor civil divisions. (MCD).

MAPS has ordered the first count summary tapes for the states of Michigan and Wisconsin. These files will be processed by the Census Bureau after the Minnesota tapes.

2. Master Enumeration District List (MEDList)

The preliminary version of MEDList is available. MEDList furnishes area and places names corresponding to the geographic identification codes in the summary tapes. The preliminary version covers the entire United States. MEDList for Minnesota has been ordered and should arrive in mid-February.

With MEDList we will add state, county, and area names to the 1970 census summary tapes. In addition, we plan to make the MEDList file Qwick Qwery accessible. This will enable a user to generate a printout of the geographic area codes and names on a selected basis; i.e., MCD code and names for all MCD's in Hennepin County. Present legislative districts and school districts are currently being coded for use with the Summary tapes.

3. DAUList

In addition to the first count summary tapes, the DAUList computer program is now available for use with the third count summary tapes. DAUList reads the first or third count summary tapes and prints the geographic identification and the respective population and housing counts with description.

4. Dane County Test File

Copies of the first, third, and fourth count summary tapes of Dane County are now available. These files are identical in

*10 copies have been ordered and will be available for check-out from Dr. Verbrugge's office.

format to the actual summary tapes and serve as an excellent test file to be used before receipt of the final summary tape for any state counts.

5. Minor Civil Division (MCD) Maps

MCD maps for all counties and the Minneapolis-St. Paul (7-county), Duluth, and Moorhead SMSA's are available for use at the MAPS office (Lauderdale, UCC Bldg.) and in 246 Coffey Hall (John Hoyt's office).

Duplicate copies can be ordered for individual use from the Census Bureau. Their cost is $1.00 per copy and delivery is about 6 weeks. MAPS is attempting to arrange for local copying and shorter delivery time. There are 180+ maps in the available sets and several hundred more can be ordered for individual Enumeration District boundaries within incorporated areas throughout the state. (For example, Thief River Falls has seven ED's within its corporate limits and maps can be ordered for such areas).

6. Early Requests

Many requests for population and housing data reports from the first count summary tapes have been received. The necessary control cards are being prepared now and requests will be satisfied shortly after receipt of the appropriate summary tape.

Most of the early requests will be completed with the use of DAUList or Qwick Qwery. Some typical requests include displaying the entire data base or some tables for selected counties, MCD's within counties, aggregating for project areas and displaying selected data items, and displaying ratios. If you have specific requests at this time, contact the MAPS staff at 376-7003.

I. Data Base and Computer Systems Questionnaire

The Task Force has devised several questionnaires designed to inventory existing and needed data files; computer system users; and computer systems used for data systems support within the University. Recipients of this newsletter and many others will be asked to respond to the surveys. We urge your full cooperation in order that the resultant inventories will be both complete and useful.

J. Request for Newsletter Subscribers

By virtue of your receipt of this first newsletter you can be assured of receipt of all future issues unless you indicate a desire to be removed from the current list.

However, despite an initial effort to circulate the newsletter widely we are sure we have missed faculty, staff, and graduate students who would have an interest in the newsletter. Would you therefore take 5 or 10 minutes now to fill out the attached form and drop it in the campus mail?

TO: Professor John S. Hoyt, Jr., Chairman
Task Force on Computerized Data Systems
246 Coffey Hall
University of Minnesota
St. Paul Campus

Please add the following individuals to your Newsletter mailing list:

Name	Office Address	Department or Unit

FROM: _____

_____

_____

(use reverse side for additional names)