AN AUTOMATED SYSTEM FOR RESPONDING TO DATA SERVICE REQUESTS

Tina G. Bixby
and
Janet K. Vavra

Inter-university Consortium for Political and Social Research
Institute for Social Research
University of Michigan

Introduction

During the 1970s, there was a steady decline in the cost and size of computing hardware with a corresponding phenomenal growth in computing capability. Computers now help store, manage, duplicate and interpret vast quantities of data with an ease and relative economy undreamed of in the past. These developments have, over the years, fostered the growth of new research methods in a variety of fields, including the social sciences. Large and more complex bodies of quantitative data have been collected as social scientists seek ways to understand human behavior with empirical research methods and scientific sampling techniques. In addition to collecting their own data, researchers have also utilized vast amounts of machine-readable data that have been prepared by other researchers, governmental agencies, and private organizations.

The changes in the computing industry combined with the increased demand for services have made it feasible for organizations to consider automating as many tasks as possible. FAST (Facility to Aid Servicing Transactions) is one system that was created in response to these conditions. This paper describes FAST and its impact on the organization which developed it.

The Inter-university Consortium for Political and Social Research

In the late 1950s a group of political scientists saw the need for an organization that would further research in political science by maintaining machine-readable data for redissemination and providing a training center in empirical research. To meet this

need, the Inter-university Consortium for Political Research (ICPR), known since 1975 as the Inter-university Consortium for Political and Social Research (ICPSR), was formed in 1962 and housed at the University of Michigan. The ICPSR remains there today in the Center for Political Studies in the Institute for Social Research. Over the years, the ICPSR has collected machine-readable data files on social phenomena, both contemporary and historical, and made them conveniently and economically available to Consortium members all over the world.

The ICPSR is supported by the merged financial resources of the over 240 academic institutions which are presently members. Each member pays an annual membership fee that entitles it to services provided by the ICPSR. Users at member institutions coordinate their requests for ICPSR services through a locally designated Official Representative (OR). ORs serve as the liaison between their users and the ICPSR. Although ORs were originally from the departments of political science, there now are representatives from the fields of sociology, library science, computer science and history.

The different fields which current ORs represent reflect both the growing interest in quantitative research in a variety of disciplines and the diversification of ICPSR holdings over the years. These developments have resulted in a growing demand for ICPSR services. From 1975 to 1980 the ICPSR experienced a 324% increase in the amount of data distributed, with 1981 continuing this growth pattern.

This increase in demand for data has put corresponding pressures upon the ICPSR staff and resources. The automation of various portions of the data dissemination activities began in 1975 when distribution records were converted to machine-readable form. In 1979 the ICPSR acquired a Prime minicomputer and eight Ontel Op-1 video display terminals. This FORTRAN-based time-sharing facility made the automation of a variety of activities feasible, including the development of a prototype for a system that would process user requests. In the summer of 1980, with the aid of a grant from the National Science

Foundation, the ICPSR upgraded to a Prime 750 and purchased several Zenith microcomputers. It was with this acquisition, a powerful but straightforward computing system, that the ICPSR was able to begin the full implementation of FAST.

Facility to Aid Servicing Transactions

FAST began operating on a limited scale in the Fall of 1980. This Spring it assumed most responsibilities for the processing of requests for data. Designed to help the ICPSR prepare and manage requests, it handles orders from the time they are received until they are completed.

The development of FAST was undertaken with a number of goals in mind, including the automation of the handling of requests, the continued effective control over the growing volume of data and auxiliary materials, the management of increasing costs, and the maintenance of a high level of service. It was further decided that the system be designed and written so that new features could easily be incorporated as they were developed.

FAST grew from a small prototype. In the early stages of development, the staff was able to use the system to perform limited tasks. At this time, staff reactions and suggestions were encouraged. Additional capabilities were added as they became available, again with staff interaction and support. This approach to system development has been successful and continues to be used as new features are added.

FAST builds and updates information in several databases. The largest database contains sixty-three variables for over 800 data collections and more than 12,000 files. Included are the names, the principal investigators and the technical characteristics of the different data files and formats available for each data collection. Thirty-one variables about ICPSR members, including the name and address of the OR and technical specifications, are in another database. FAST additionally stores information about orders in progress, including such items as the date received, the data ordered, and the request status.

FAST generates setups for the programs that copy data to users´ tapes and those which verify the contents of the tapes after they have been written. If a tape has to be rewritten or new files added, the appropriate changes can be made through FAST.

Every user requesting data from the ICPSR receives printed documentation in addition to the data tape. The documentation consists of a letter, a tape information form, partial listings of the files on the tape and the appropriate codebooks. The letter advises the user that a tape has been sent, identifies the data supplied and provides any other specific information appropriate to the order. The tape information form identifies the tape, its contents and provides appropriate technical details.

In the past letters and tape information forms were typed. Now FAST generates files containing both the letter and the tape information form for each request. Letters that were first drafted by the staff, typed, and then proofed are now composed by issuing instructions to FAST from a terminal located in each staff member´s office. Tape information forms that were originally prepared from computer printout, typed, checked first for technical problems and checked a second time for typographical errors, are now generated by FAST and checked once. The files are then routed to a hardcopy printer. A copy is sent to the user and another is retained for ICPSR records.

The Implementation of FAST

The interaction between FAST and the staff is designed to prevent errors and allow for the processing of an order in a systematic, uncomplicated manner. The user begins by indicating the "command" (or "task") that is to be performed. These commands include such operations as entering a new request or a new institution, updating information about a data collection, writing a letter or checking on the status of a request. Once the command is given, FAST issues "prompts" to obtain the information it needs to accomplish the task. When appropriate, each prompt also lists the acceptable responses. The response is checked for validity as soon as it is given. There also is an on-line help facility which the user can invoke whenever in doubt about how to respond to a prompt. At any point in the prompting sequence, all the prompts and responses given up to that point can be listed, a response just given can be changed, or the whole prompting sequence can be restarted. Further, there are special procedures for interrogating any of the databases (e.g. a data collection or a request) in the midst of a prompting sequence, and then returning to the prompt in progress. The prompting flow proceeds in logical order and with phraseology familiar to the user, allowing errors to be easily identified and corrected.

Each institution that is either a member or has utilized ICPSR resources in the past is assigned an unique number. Also there are unique numbers assigned to each data collection and file in the holdings. These codes are used to identify and access entries in the different databases that FAST uses. Previously this information was available to the ICPSR staff in the form of paper records and lists.

When a request is entered into the system, the institution number is used to identify the requestor. Each data collection requested is uniquely specified by the three identifiers:

the data collection number, the dataset number within the data collection, and the format of the dataset. The system automatically assigns the request a number and all parts of the request are assigned subrequest numbers. In this fashion, the staff can give instructions for any separate file that a given request involves.

Tapes that will be used to fill the order are identified by institution number, external label and request number. The ICPSR staff can also at this time modify any of the tape specifications that are on file for the user to accommodate any temporary changes requested for a particular order. Similar modifications can be made for data formats that a user may request for any given data collection. Should there be a need to modify either the data or tape specifications for an order that has already been input, an update function is available. By updating either the tape or the request, appropriate changes can be made in the entire order or a portion thereof.

After a request and tape have been entered, the user issues a command for FAST to generate a "work form". This is a report that summarizes the request by institution, data ordered, and tapes to be used. This information can be displayed at the terminal or can be routed to the printer for hardcopy. The work form allows the staff member to check for problems with the request and to make changes if a request was incorrectly entered. It can also be a checkpoint for a supervisor to determine whether work on a request should progress or whether other information is needed from the requestor before any other work can be done.

The setup command creates the computer job that will copy or subset the ICPSR files to the user's tape. Working from the information it accumulated when the request was entered, FAST builds the program setup that will write the tape according to the user's specifications. This setup is stored in a file on the Prime. This file, and all similar files generated during the day, are transferred to the University of Michigan's AMDAHL 470 for execution overnight. After the tape has been written, the verify command prepares the job that will check the tape for problems, normally the following day. Among the problems this procedure can detect are those of a technical nature, such as parity errors, or human errors such as mispunching a data collection number and retrieving the wrong data file.

When a tape has been written and verified and is ready for shipment, it is closed out of the system. At this stage, the staff member filling the order has another opportunity to catch any problems with the tape or request. The system requires the staff member to approve the closeout of each subrequest on the tape. After the tape is closed out, the system creates the tape information form. This form is generated from the information

that accumulated as the request was processed. Stored in a file, the form can be listed for checking either at the terminal or routed to the printer. Final finished copies are then made to accompany the tape and to go into the ICPSR files.

After the completion of the tape information form, FAST creates records summarizing the request. Included are institution name, data collections supplied, data format, and date sent. These records are used to generate reports on data distribution activities.

When all of the subrequests have been filled, the request is closed out. At this time, the request is also assigned a date when it should be purged from the system. If for any reason the request must be reactivated, it will be available up to its purge date. If all portions of an order have not been filled, the request is retained in the system to be finished at a later date.

By issuing the letter command, one can write a letter to accompany a tape or in response to any other inquiry that has been received. The command prompts for the institution and addressee. Certain defaults are presented thereby reducing the amount of information that the staff member needs to enter. If a tape is associated with a letter, FAST will look up the tape and indicate the data that were written on the tape, incorporating this information into the letter. The letter writing capability is an extension of the system that was designed for a magnetic card typewriter. This approach allows staff to incorporate standardized text with individual responses to specific requests and inquiries.

By giving either the purge request or purge tape command, the database administrator can remove all records of completed orders for which the purge dates have been reached. This feature not only keeps the database from overflowing with completed transactions and eventually impacting on system response time, it makes the maintenance of the request database a simple procedure.

The Impact of FAST on the Staff

Although the FAST system has not been in full operation for an extended period of time, it has already provided many benefits for the staff. The handling of a variety of paper records in order to fill an order has been eliminated since all information needed is as close as one's terminal. The transposing of information as it is transferred from records to work sheets has been eliminated. The need to leave one's desk to seek information from centrally located records has also ended. The hours that were spent preparing, typing and proofreading the supporting documentation associated with each order have been replaced by several commands issued to FAST.

The acceptance of the system by the staff has been excellent. A good portion of that can be

attributed to the fact that the system did not totally change everything that they were familiar with, but instead incorporated this information into its operations. Data collection and institution codes remained the same as did many of the steps needed to fill data requests. Now, information was centrally stored by FAST and at one's fingertips. Many of the timeconsuming tasks, such as preparing tape information forms, which required accuracy could now be handled quickly by FAST. Futhermore, FAST was brought up in stages and, therefore, it was neither overwhelming nor threatening. As each stage was implemented, they were able to adjust to new approaches to performing familiar tasks. As a result, they could become comfortable with only a few changes at a time, before they had to make other adjustments that the next stage required.

As FAST assumes routine clerical chores, the staff can concentrate on those tasks that require their attention and which cannot be handled by a computer. Their productivity has increased and this has helped their morale as they confront steadily increasing demands on limited resources. As costs rise and demand for services continues to grow, FAST helps the ICPSR to continue to respond to requests for services in an economical, efficient and timely manner and with a modest sized staff.

****