
The Electronic Triumvirate: The Archives; The Data Processors; the New York State Department of Correctional Services Inmate Files. A Case Study

by Hugh W. Shinn¹

New York State Archives and Records Administration

Abstract:

The New York State Archives and Records Administration (SARA) was one of the first state archives in the United States to accession electronic records into its holdings and make them available to the public. SARA worked with the New York State Education Department's (SED) electronic data processing division (EDP) to obtain main frame computing services. SARA's dependency on SED EDP for data processing services required the development of a positive relationship with SED EDP.

This case study examines the relationship between a government archival institution working with electronic records and a centralized data processing unit that is completely unfamiliar with the operations and requirements of a data archive. There are two levels of a successful relationship: formal, which includes agreements on hardware, disk space, training, etc.; and informal, including the development of creative solutions to technical or procedural problems. These levels of interaction were necessary because SARA (unlike other SED divisions) developed and executed its own applications rather than use the traditional EDP services.

SARA

The New York State Archives and Records Administration's (SARA) program for the archival preservation of and research services for electronic records had its origins in 1988 with the release of the Special Media Records Project report: Strategic Plan for Managing and Preserving Electronic Records in New York State Government. This plan gave birth to SARA's Center for Electronic Records (CER) in 1990, which is to be the focal point of electronic records program development in SARA. Among its many charges, CER was the unit assigned to bring electronic records transferred from agencies to SARA under archival control and to provide reference services for those data files. In this instance, archival control refers to variable-level descriptions of the data set, explanations of the data set's technical aspects, the arrangement and description of the data set as a records series, and verification of the data and the documentation. In archival terminology, the process of bringing electronic (or paper) records under intellectual and physical control is termed

"accessioning".

Archival Administration

The accessioning procedure for electronic (and paper records) is preceded by the records retention and disposition scheduling process in which agencies establish minimum retention requirements for records and determine their final disposition: either destruction or transfer to SARA. Most records are destroyed after they are no longer useful to the agency because they do not have enduring legal, administrative, evidential, or research value.

Before electronic records are accessioned by SARA, they are appraised for archival value. Appraisal involves the examination of electronic records from both a content and a technical point of view. Once the records have been determined to be of archival value, the data set's technical aspects are examined to determine whether SARA has the skill and equipment to preserve the data. Technical appraisal is based on discussions with agency personnel and careful examination of the technical and data documentation of the data set.

Data Processing

Before electronic records accessioning operations could begin in 1990, SARA had to obtain computer and technical services from the State Department of Education (SED) Division of Electronic Data Processing (EDP). SARA initially developed an informal plan that specified CER's requirements and contributions to the joint venture of accessioning electronic records.

Requirements of the Archives

CER operates differently from other units in SED with respect to data processing requirements. Most units work with a specified type of data and a specific set of data files. Conversely, SARA collects data from all Executive Branch agencies and must contend with a wide array of data files, types, and formats. The result is that diagnostic programs must be individually designed for specific data files.

Generally, SED EDP customers have informational products such as reports and publications that must be produced. Unfortunately, the user may lack the skills,

time, or equipment to accomplish the unit's tasks. Typically, at the customer's request, EDP will carry out an examination of the unit's particular information requirements, conduct a needs assessment, and where appropriate, assist in the development of goals for the automated system. After developing these plans, SED EDP designs, produces, and tests the programs that accomplish the project's stated objectives. The data processing shop is also responsible for system upgrades and major changes. The customer determines when the data system will run, EDP determines how it will operate. This type of arrangement is practical for systems that are totally dependent upon EDP for development and support.

SARA contends with data from different systems that contain few (if any) common denominators, designed by personnel with varying skill levels. Each data set requires unique levels of handling and effort to document the data fully. In order to provide this level of service, the SARA analyst must become a hybrid of programmer and end user, successfully combining attributes from both worlds. It is this requirement for multi-faceted operation that sets SARA apart from traditional SED EDP customers.

One aspect of commonality among the data sets that SARA has already accessioned is that they were produced by automated systems that are now largely defunct. Electronic record systems are suspended for many reasons including: loss of funding, migration to new equipment, and the completion of a temporary commission's task. When this is the case, the source code is generally missing, and the original programmers have long since departed. Under these conditions, SARA must become an investigative body to determine the scope and function of aging or superseded electronic records systems. These investigations seldom conform to predictable schedules such as annual reviews or a decennial census. Rather, they may appear at inopportune moments in an astonishing array of formats. SARA is currently working with agencies to replace this practice of sporadic submission of out-dated material with the more predictable method of scheduled transfers of current data.

CER's role as an investigator and the necessity for multi-faceted operations, produced a number of data processing requirements. The services required by CER for screening purposes are as follows:

1. Access to magnetic tape drives on a non-scheduled basis.
2. Programs to check the physical condition of magnetic tapes for problems such as parity errors.

3. Non-standard training for SED's Unisys mainframe software tools. (The goal is to develop a degree of independence from the EDP staff.)

4. Training in the use of software packages (such as SPSS and INSYTE) that can be used to analyze agency data sets.

5. On-line access to data files.

6. Ability to manipulate data files from the programmer's point of view. This includes variable-level operations involving the verification of data values.

7. The ability to carry out data verification procedures by comparing data values with existing documentation. (Generally carried out with frequency distributions or displays of value ranges.)

8. The ability to isolate and examine specific columns and records within a data set from the operating system level.

9. Production of copies and extracts of data sets, and a method for distributing the data to customers.

Armed with these requirements, CER met with representatives of SED EDP in search of agreements and assistance. For the most part, the data processing staff were helpful and available for consultation.

SED EDP's Operating Procedures

SED is an enormous and complex organization. It has an excess of 3,000 employees, many of whom use or maintain the Department's fifty diverse mainframe applications or its state-wide data network. The size and complexity of the Department required that SED EDP develop standard operating procedures for most types of data and technical functions. These procedures are as follows:

1. All requests for services will be directed to an EDP staff member in the unit assigned to the customer's organization.

2. Customers will not have direct contact with the technical support staff.

3. Customers that have programs requiring the use of magnetic tape drives will request that their EDP contact run the program and submit a 'run sheet' to the EDP operations staff.

4. Formal training above the level of technical manuals and vendor programs is not available for

software packages or higher-level languages.

Most of these procedures reflect EDP's customer service orientation. In fact, even the lack of formal training relieves the customer of the tedium of writing code and deciphering arcane documentation. SED EDP's view is that the customer should simply have to make a request, and the data processing professionals will provide the user with the required product. It is an efficient method for managing requests from a large number of customers who require routine services.

Combined Procedures:

The first attempt to combine SARA's data processing requirements with EDP's existing procedures occurred when CER began to examine records for the Department of Correctional Services (DOCS).

The DOCS Inmate Under Custody data files were the testing vehicle for the electronic records accessioning program at SARA. DOCS began collecting data on punch cards in 1956. The data were used to produce reports on the numbers and characteristics of inmates under custody from 1956-1974. The information contains incarceration data including facility name and date received; crime and sentencing data; detailed demographic data including race, religion, nativity, occupation, and education; criminal history data. Each annual file is a reflection of the prison population at a given point in time.

These files are unique in that they contain relatively complete inmate-level demographic data on the population of New York State's correctional facilities².

The initial foray into the unknown and the semi-known began when the first data files from the DOCS were examined. The technical information did not coincide with the data sets, nor did the data documentation match the actual data values. It became apparent that additional tools were necessary if the proper file formats were to be discovered.

The Mystery File

The mystery file appeared with the second shipment of data files from the Department of Correctional Services. This data file refused to coincide with any of the printed documentation. It was difficult to determine the structure, size, record lengths and other important characteristics of this data set.

The tools and skills that could be used to augment SARA's level of training and experience and bring the mystery file under archival control were distributed throughout SED EDP's sections. While individual sections often cooperated with each other, staff in one section had only a limited need to understand the activi-

ties in the other sections. In this instance, the single contact system introduced an additional level of bureaucracy into the arena. The contact had to transmit SARA's requests accurately to the appropriate EDP section, and accurately transmit that section's response. Unfortunately, the mystery file stretched the single contact system beyond its limitations, and the file had to be abandoned.

The mystery file situation demonstrated that CER's data processing requirements were not routine and new procedures had to be developed by EDP to assist CER in its attempt to accession electronic records. SARA had the ability to perform much of the work normally provided by SED EDP. With this in mind, SARA requested access to the tools, software, and training that would lead to self-sufficiency.

EDP's Response to SARA's Requests

SED EDP's response to SARA's requests for services was to set up a liaison system where CER would contact a programmer in the applications section assigned to SARA. The liaison would respond to SARA's requests for services, or attempt to find someone who could. Negotiations between SARA and SED EDP resulted in the modification of EDP's procedures with respect to non-routine operations.

1. On-line access to data files from a programmer's point of view was always available.
2. SED EDP responded rapidly to requests for service.
3. SED EDP willingly participated in the development of tools and procedures to solve unforeseen problems.
4. SED EDP allowed limited access to the technical support staff.

These modifications allowed CER to make use the old, inadequate documentation, examine the data, and determine the true nature of the DOCS files. In addition, SARA was able produce more useful documentation for the DOCS Inmate Under Custody files.

Attempts at the formal level to resolve the issues and problems raised by the mystery file situation were not entirely successful; and it became apparent that an additional and different type of relationship with SED EDP was required. Fortunately, strong working relationships were developing between CER and several SED EDP sections. This led to the development of informal relationships with specific programmers from selected sections of the data processing shop.

The basis of the informal relationship was the realization that personal cooperation between specific SED EDP programmers and the SARA analyst would be a more effective than an organizational approach for solving unique problems. One of the mystery file issues was resolved when the programmers and the analyst devised a method for examining specific columns and records in a data set. Admittedly, this was one CER's formal data processing requirements. It became an informal item when it could only be operationalized by cooperation between the analyst and the SED EDP programmers.

The informal relationship was designed to augment the liaison system by developing contacts in other sections of EDP. The liaison served as a facilitator, providing the SARA analyst with the initial introduction to the appropriate sections. From that point, the analyst and the programmer in that section would work together as needed to resolve the problem at hand. The strength of the informal relationship is its flexibility.

The combination of the informal and formal relationships formats results in the creation of a *Relationship Network*. This network is based on the premise that formal relationships are effective for repetitive operations; while informal relationships are more suited to solving unique problems.

There are formal connections between SARA and the EDP liaison and between the EDP liaison and EDP's individual sections. Through these connections flow procedural recommendations, written procedures, and formal requests for assistance. The formal relationship is particularly useful for contending with repetitive activities such as transferring data files. In addition, it provides a structure for communication between CER and SED EDP.

Informal relationships also exist between SARA and the EDP liaison and among SED EDP's various sections.

Additionally, informal relationships exist between SARA and the individual SED EDP sections. The informal relationship is useful for the resolution of unique problems such as accessioning new types of electronic records. In short, the informal relationship is more effective in resolving cutting-edge problems rather than completing routine tasks. Informal relationships are legitimized by the existence of formal relationships, and cannot function effectively in the long-term without them.

Conclusion

The flexible aspects of accessioning electronic records require that the archivists and the programmers have the ability to adjust to the variable demands of differing electronic records formats. Flexible operations spring from adjustable, non-stagnant relationships.

The development of a *Relationship Network* between SARA and SED EDP provided the necessary structure for a flexible electronic records accessioning program. The *Relationship Network* combines the aspects the formal and the informal relationship formats. As a result, the electronic records program at SARA is flexible, efficient, and has the capability to address technical and logistical problems (e.g. reading data dictionaries and transferring data files).

The development of a strong *Relationship Network* has provided a stable foundation for cooperation between SARA and SED EDP.

**** a diagram follows (see p.12 of text) ****

1. Presented at the IASSIST 92 Conference held in Madison, Wisconsin, U.S.A. May 26 - 29, 1992.

2. New York State Archives and Records Administration, Bureau of Records Analysis and Disposition. Appraisal report # 87-28N; May, 1987; p. 3.

