The Electric Archive

By **HUGH W. SHINN***

The methods and tools for administering archival records have evolved significantly during the last two decades in order to meet the requirements and exploit the advantages of the computer as a practical research and storage tool. However, the evolution of archival practice to accommodate electronic records has been, and continues to be met with some resistance and confusion. A common argument is that rapidly-changing technology makes it impossible to administer and reference electronic records. I shall not address the more theoretical issues relating to acquiring and preserving electronic records. Rather, I shall focus my discussion on how traditional archival principles and techniques can successfully be adapted to electronic records. I shall use as my frame of reference the New York State Archives and Records Administration (SARA), which has successfully accessioned electronic records using modified, general archival techniques.

SARA makes a clear delineation among the processes of archival appraisal, accessioning, and description of paper records. Although all three lead to an accessioned, preserved, and available set of records, they exist as separate procedural stages in the archival processing system. The same processes are used for electronic records, but in a format that synthesizes these separate activities into a procedure that adheres to the particular requirements of electronic records accessioning, description, and reference. A successful accessioning programme also depends on a number of other considerations, such as: the type and level of service provided to clients, the background and skills of the archival staff, and the provision of data processing services.

Appraisal

Initially, SARA appraises records for long-term value regardless of the medium on which the information is stored. When electronic records are appraised, an additional, technical evaluation is conducted. This technical evaluation, which determines whether the archive has the skill and technology to accession the records series fully, is based on careful examination of the series documentation, and discussions with the creators/custodians of the records. The technical evaluation is part of the standard appraisal process for electronic records and appears as a section in a detailed appraisal report, which forms the basis of the decision to accept or reject a records series. A records series may be accepted for informational or evidential value but rejected for technical reasons. For SARA, these reasons include the following: data sets that are not or cannot be easily placed in a software-independent format; data sets for which there is

missing documentation or none at all; and data sets stored on obsolete media that cannot be transferred to current formats. An electronic records series that cannot be accessioned for technical reasons, but contains important archival information, may be stored in its existing condition in the hope that the archive will obtain the necessary skill, technology, or documentation to accession the records in the future.

Electronic records accepted by the appraisal process are passed to the pre-accessioning evaluation stage in order to determine the veracity of the documentation and the data. In order to examine the data closely, copies of the data sets and documentation are transferred to the archive.

Pre-accessioning

Pre-accessioning examines the preliminary conclusions of the technical evaluation. This procedure combines the data with the documentation by constructing computer programmes or routines in software packages that display the values or value ranges for variables (e.g., age, name, sex) found in the data set. These results are compared to the expected values found in the documentation. Discrepancies are noted, and their cause(s) investigated. Errors can arise from two sources: programming mistakes and inaccurate documentation. The programming errors can be identified and corrected with relative ease. They are characterized by complete mismatches of variable values with respect to the anticipated values indicated by the documentation. An examination of the data definition section of the computer programme, moreover, will indicate errors in the positional specification of the variables.

Erroneous-documentation problems are characterized by accurate positioning of variables in the data definition section, while the variables described by the documentation are not the variables found in the data (e.g., the documentation indicated a name variable, but the data in that position represent a date).

After it has been determined that the position of the variables is correct and that the documentation is accurate, the programmes that display actual values of the variables are run. These programmes (e.g., frequency distributions) will indicate the number and percentage of actual values that match the expected values. When there are more actual values than expected values, the number and percentage of these additional values are recorded and used to compile the pre-accessioning report. This report indicates the true condition of the data with respect to its accuracy, the accuracy of the documentation, and the number of missing or invalid values for each variable. Based on these elements, the records series is again accepted or rejected. Rejection may occur because of the unreliability of the data itself, significant numbers of missing or invalid values for important variables, or gross inaccuracies in the documentation. Electronic records series that are accepted by the pre-accessioning process are then transferred to the archive, which assumes legal responsibility for their custody.

The next phase of the accessioning procedure for records that have progressed to this point is description. Until they have been described, the records are not completely accessioned.

Description

Electronic records must be described in greater detail than paper records. In addition to explaining the function, use, type, scope, and content of the records, the description must contain the data sets in block-size, the record's length, the locations of variables, their type, valid values, and necessary appendices. The combination of these elements form a user's guide that serves as the primary user tool for the data series. When the user's guide has been written and security copies of the data sets have been made, the records are considered to be fully accessioned.

Appraisal, accessioning, and description, when combined with the technical evaluation and pre-accessioning procedures, form the core functions of an archival electronic records programme. The inclusion of the core functions among the additional aspects of type and level of service provided to clients, personnel issues, and data processing services will result in a coherent and productive programme.

To achieve the goal of a smoothly running electronic records archival programme, the repository must first answer basic questions relating to the type and level of service that it will deliver to its clients. The "type of service" questions are as follows:

1. Is the repository a data library in which the staff have physical control of the data? *or*
2. Is the repository a data archive in which the staff have physical and intellectual control of the data?

The "level of service" questions are as follows:

1. How much effort can and should the institution expend in examining and verifying the data?
2. What quantity of electronic records would the institution prefer to accession during an operational year?
3. How should the institution distribute its resources in order to meet the competing goals of these two service levels—rigorously verified data versus a larger quantity of available data?

Type and Level of Service

The major difference between a data library and data archive is that the data archive is concerned with the reliability of its data whereas the data library is a distribution point for existing data and does not vouch for the data's veracity. The practical difference is that because it does not carry out extensive examinations of the data in its holdings, the library will be able to process and disseminate data sets more quickly than will the archive. Furthermore, the library will have fewer unique demands for data processing services.

Conversely, the data archive, since it engages in the core functions described above, is very likely to have higher quality data in its holdings. The disadvantage is that the core functions are time-consuming, with the result that fewer electronic records will be available from the data archive.

Neither of these choices is correct or incorrect; they simply represent the endpoints of the continuum involving the provision of access to electronic records to various research communities and the public. It is likely that each archive and library will adopt some of each other's practices. This hybrid will evolve independently, based on the requirements of the organization and the number of electronic records series to be accessioned by an archive or acquired by a library.

There may be a point at which the archive will be unable to carry out the core procedures on all of the data sets for a particular records series. When this is the case, as in annual or longitudinal data files, the decision to use sampling techniques to determine which data sets to process is both necessary and practical. The data sets selected in the sample are processed with respect to the veracity of the data and the adequacy of the documentation. If the sample is accepted, both it and the non-selected data sets from the series will be described and accessioned. As a result, the archive can accession a larger number of electronic records while maintaining a respectable level of data reliability.

Personnel

Two other factors to be addressed in the developmental stage of an electronic records programme are the skills and background of the archive or library staff, who require some familiarity with data and their use. Generally, social scientists such as sociologists, political scientists, and, to some extent, historians have training in quantitative analysis. Persons in these professions are likely to have a "feel" for data and may be more comfortable with electronic records than their more traditionally-trained archival and library colleagues. The social scientists, moreover, will be able to communicate with the research communities and the data processing professionals without the use of an interpreter. Accordingly, the electronic records programme does not have to be staffed by hard-core data processing professionals. Indeed, such an arrangement could limit the effectiveness of the programme.

Data Processing Services

The arrangement of the core procedures and determination of the type and level of services, coupled with the selection of technically-oriented archive or library personnel, will help to determine the type and amount of data processing services required by the repository.

Among the most critical functions of an archival electronic records programme is who will provide data processing services. A wealthy archive or library may well purchase its own computer and handle all data processing requests internally. However, most archives and libraries will have to contract with outside organizations for data processing services. The level of interaction will depend on whether the repository has an archival or a library science orientation.

Routine operations, such as copying data sets from one tape to another, can be handled by the data processing organization with little input from the library. However, the diagnostic work related to the core procedures varies in amount and intensity, and does not conform to time schedules. On this account, the archival staff must cooperate with the data processing staff to produce an effective product. This condition does not require complete surrender by the archive to the data processing organization's agenda. Archival records are best dealt with by professionals from an archival rather than a technical point of view.

The archive or library has a set of procedures for dealing with electronic records. The data processing organization has its procedures as well. The two must be married in a way that preserves the independence of the archive while not destroying its links to the data processors. One method is through cooperation. The technically-oriented archive or library staff can work as independently as possible so that they do not become a burden to the data processing organization. However, technical queries from the archive or library to the data processing organization should be answered quickly. Communication is the key factor with respect to data processing services. A network of relationships, formal for routine matters (e.g., tape copying) and informal for non-routine operations (e.g., investigative work on complex electronic records) will greatly enhance the archival electronic records programme (**see figure**).

Relationship Network

Formal Relationship

Informal Relationship

Data Processing Organization (DPO)

	DPO
	Department
Archive	
or	DPO
Library	Liaison
	DPO
	Department

Conclusion

The situation that I have described is a real one. By integrating the processes of appraisal, accessioning, and description it is possible to ensure that data is of sufficient quality to avoid the pitfalls that may occur if an archive or library takes a wait-and-see approach. I propose that in order to accession electronic records successfully, it is vital to invest the time and resources at the beginning of the process, rather than paying the price later on—and potentially disappointing the users.

Notes

* Paper presented at the 56th Annual Meeting of the Society of American Archivists, Montreal, 14 September 1992.