Commentaries & Case Studies

DEAN DeBOLT *and* JOEL WURL, *Editors*

The Commentaries and Case Studies department is a forum for sharply focused archival topics that may not require full-length articles. Commentaries and Case Studies articles generally take the form of analyses of archivists' experiences implementing archival principles and techniques within specific institutional settings, or short discussions of common theoretical, methodological, or professional issues. Members of the Society and others knowledgeable in areas of archival interest are encouraged to submit papers for consideration. Papers should be sent to Managing Editor, *The American Archivist*, Society of American Archivists, 600 S. Federal, Suite 504, Chicago, IL 60605.

Optical Character Recognition: Technology with New Relevance for Archival Automation Projects

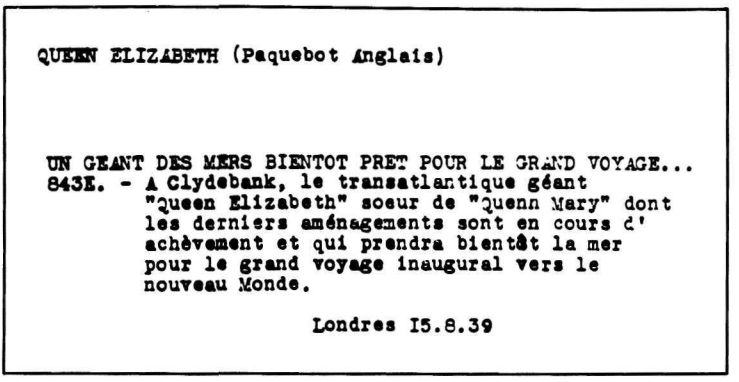
MARIE ALLEN

Optical character recognition (OCR), the automated conversion of written data to machine-readable form, has for many years been a technology with more potential than practical application for archives. The breakthroughs in the area have generally related to a few special typefonts, to large and extremely expensive machines, or to machine-readable codes such as bar codes. Although these developments had a major impact on financial and retail industry applications, they had less relevance for archives with smaller budgets and large quantities of older documents, often distinguished by poor resolution type, deteriorating ink or paper, and the cursive handwriting of previous centuries. Now, however, given the OCR developments of the last three years, the time has come for the relevance of the technology to be reexamined. The cursive handwriting barrier has been broken by at least one OCR service bureau. Inexpensive desktop OCR machines, which can be attached to microcomputers at archival work stations, can take the place of some manual data entry tasks. What follows is a report

Marie Allen has held a number of supervisory archival positions at the National Archives and is currently on detail to the Archival Research and Evaluation Staff. *She expresses appreciation to Archives officials James O'Neill, Charles Dollar, and William Holmes, who provided her the opportunity to work in the area of optical character recognition technology.*

on those new OCR applications presently underway at the National Archives and Records Administration.

In 1983, the National Archives selected OCR as one of three technologies that the newly-formed Archival Research and Evaluation Staff (NSZ) should examine. The personnel of this office, professional archivists and computer specialists, were committed to identifying and monitoring the development of new and emerging technologies, and evaluating their potential for application to archival programs and operations. Besides OCR, other conversion technologies targeted for priority examination included speech pattern recognition and digital raster scanning. As part of the OCR investigations, sample archival documents were converted by leading OCR companies in the United States, and the results were compared. The conclusions of this study were published in October 1984 in the *Technology Assessment Report*; a copy of one document converted, with disappointing results, is shown in Figure 1.

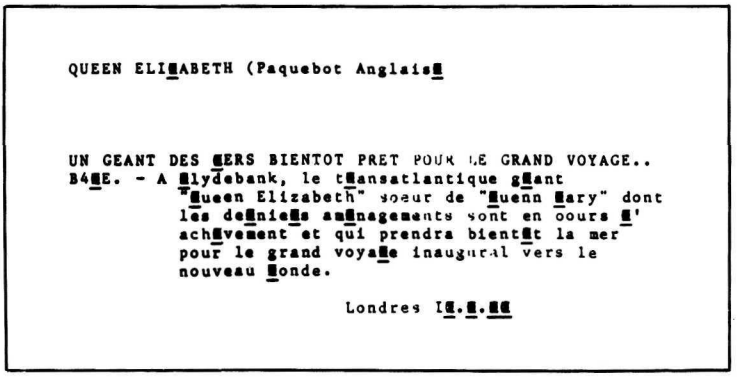


QUEEN ELIZABETH (Paquebot Anglais)

UN GEANT DES MERS BIENTOT PRET POUR LE GRAND VOYAGE...
843E. - A Clydebank, le transatlantique géant
"Queen Elizabeth" soeur de "Quenn Mary" dont
les derniers aménagements sont en cours d'
achèvement et qui prendra bientôt la mer
pour le grand voyage inaugural vers le
nouveau Monde.

Londres I5.8.39

Index card from the *New York Times*, Paris Bureau, Photograph Series; part of the holdings of the National Archives.



QUEEN ELI■ABETH (Paquebot Anglais■

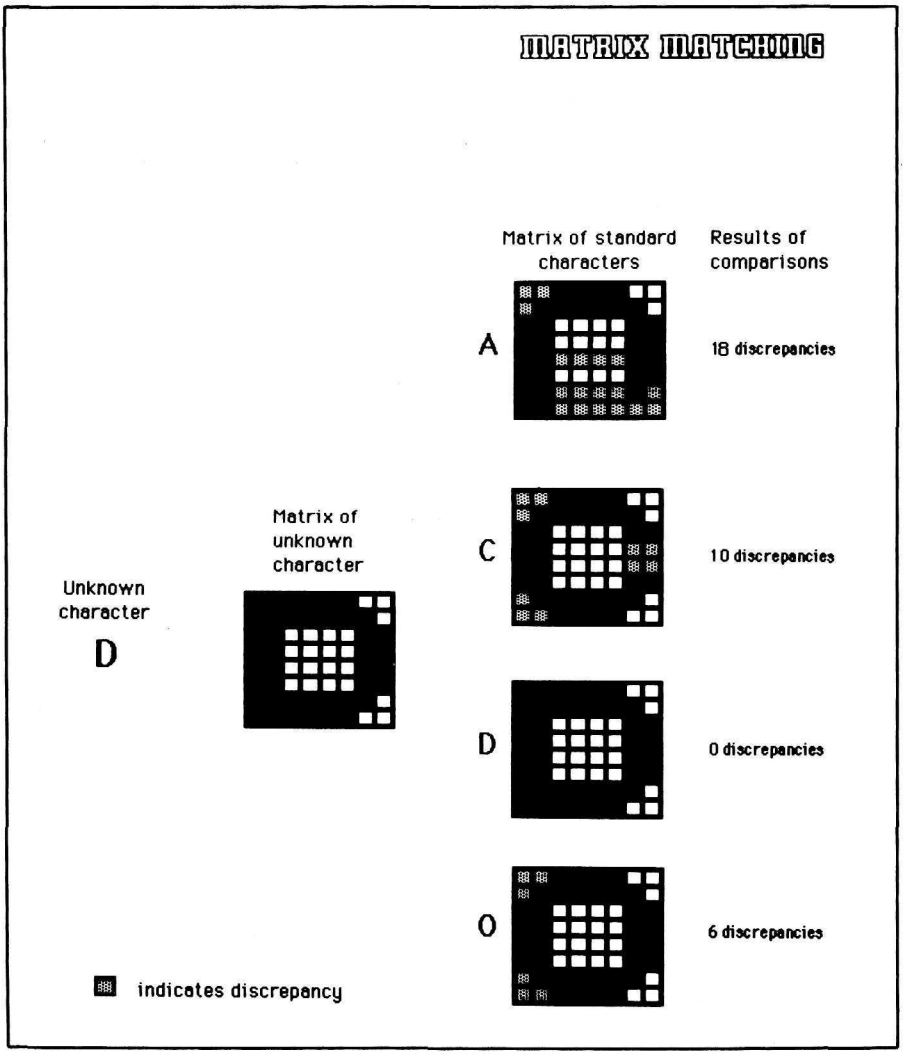
UN GEANT DES ■ERS BIENTOT PRET POUR LE GRAND VOYAGE..
B4■E. - A ■lydebank, le t■ansatlantique g■ant
"■ueen Elizabeth" soeur de "■uenn ■ary" dont
les de■nie■s am■nagements sont en oours ■'
ach■vement et qui prendra bient■t la mer
pour le grand voya■e inaugural vers le
nouveau ■onde.

Londres I■.■.■■

Printout of Index Card after OCR conversion. Black rectangles indicate characters that the OCR process could not read. This printout is typical of the disappointing results achieved by standard OCR processes for poor resolution typed archival material.

Figure 1

The *Report* identified three types of systems in the OCR industry: character readers for alphanumeric text; bar code readers for black lines and bars representing characters; and optical mark readers for reading black marks on forms. Although a typical OCR system consists of three functions—detection, recognition, and output—the *Report* defined recognition as the most critical activity. There were, at that time, two primary methods of recognition: matrix matching and feature analysis. Matrix matching consists of converting an alphanumeric character into digital signals that are compared with digital signals of a known

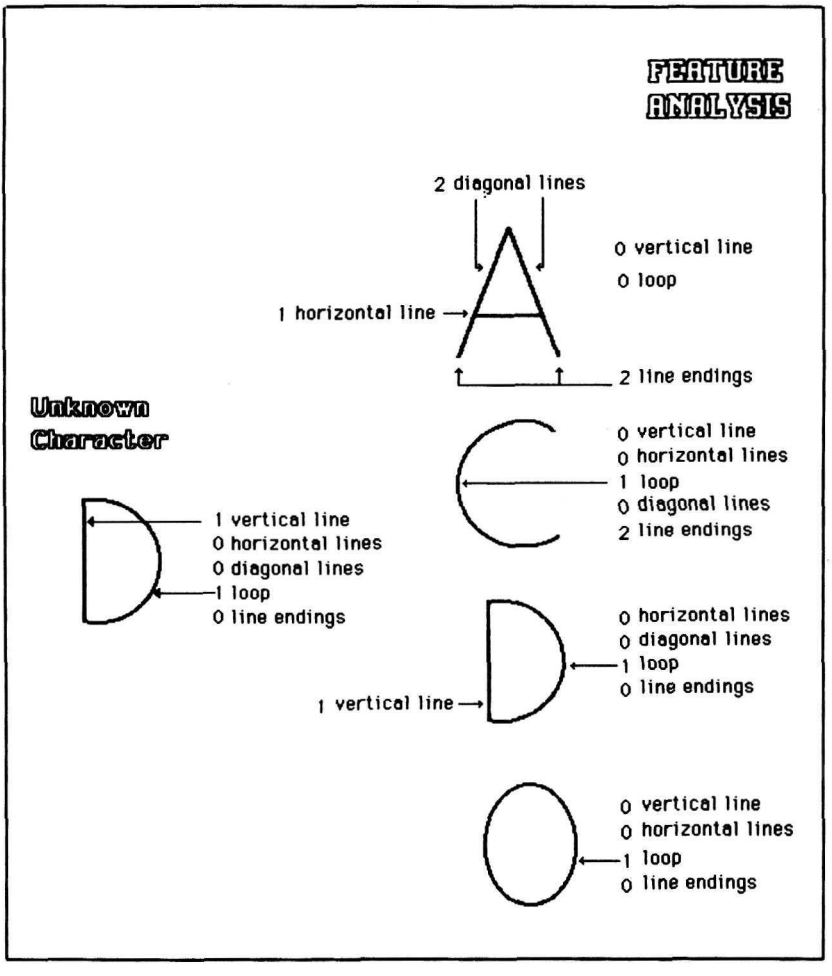


Matrix Matching is one of the standard character recognition methods used by OCR machines. This illustrative chart was created by William Hooton of the Archival Research and Evaluation Staff.

Figure 2

standard alphanumeric character set. The conversion is done by dividing each line or page into a graph in which each small square is referred to as a pixel, or picture element. The pattern of black and white pixels is used to create digital signals. Although feature analysis also uses the pixel element method, the black pixels are analyzed in terms of vertical, horizontal, and diagonal lines, loops, and ending lines. The pattern of loops and lines is then compared with a standard set for alphanumeric characters. Pictoral representations of the two primary methods of recognition are shown in Figures 2 and 3.

The *Technology Assessment Report* described current advanced OCR machines as having a throughput rate of between 85 and 150 characters per second, reading the equivalent of 300 pages an hour with established error rates of



Feature Analysis is another standard character recognition method used by OCR machines. This illustrative chart was created by William Hooton of the Archival Research and Evaluation Staff.

Figure 3

one in 300,000 characters; by comparison, a good typist can produce about six to ten pages per hour, with an error rate of about one in 3,000 characters. Nevertheless, the archival documents scanned by various vendors revealed an unacceptably high error rate; the time required by an operator to intervene manually and correct errors eradicated most of the time advantage of OCR over manual keying. The rate of correct character recognition by OCR machines is closely related to the clarity of the image on the document, and archival holdings typically include a high percentage of older documents with poor resolution images on faded paper. None of the OCR machines examined could read cursive handwriting, although some could convert printed handwritten characters on special forms.

The *Report* concluded that, if the technology were able to convert archival documents successfully, significant benefits could be realized through the conversion of item level finding aids to machine-readable form. Many of these finding aids are in typed or printed form, and manipulation as automated data bases could significantly enhance their availability, usefulness, and cross reference capability. Through pilot projects, a selection of finding aids should be tested against current OCR machines. The findings from such pilot projects would determine the feasibility of OCR technology at the National Archives for the immediate future.

Building on the recommendations of the *Report*, the National Archives established two pilot projects in 1984 and 1985 to continue the examination of OCR technology[1]: (1) a practical test of the usefulness of the new desktop OCR machines, in converting a few selected finding aids from the Still Picture Branch of the National Archives; and (2) the development of a contract with an OCR service bureau, for state-of-the-art conversion services for more than a dozen larger series of finding aids.

The first project involved the purchase of a desktop OCR reader. A relatively new product, this generation of small, portable OCR machines attached to microcomputers or minicomputers was, at the time, available for under $10,000. (Similar machines now are available for less than $5000.) The portable machines came equipped to read only programmed typefonts, without the capability of larger and more expensive OCR machines for learning other typefonts. After looking at several vendors, NARA selected the "Workless Station 213-A" manufactured by the DEST Corporation of Milpitas, California, and equipped to read a maximum of 12 typefonts. The Workless Station operated in conjunction with an IBM-PC-AT microcomputer, using Crosstalk communication software and Word Perfect 4.0 word processing software.

Archives staff members surveyed finding aids to identify those documents with a good type quality and a high level of research use, appropriate for a pilot project involving the Workless Station. Because of the limitations of microcomputer storage, the series selected were relatively small in data size, compared to many of the Archives' finding aids. Five series from the Still Picture Branch were selected and the Archival Research and Evaluation Staff began the conversion process for the first finding aid—descriptions of record groups—consisting of approximately 200,000 characters. The record group descriptions had been typed on one electric and many manual

[1]An earlier project was also underway, examining the usefulness of bar codes for the management of storage and retrieval of cans of motion picture film.

typewriters over a period of several years. It was not known whether all of the typefonts were among those included in the programming of the Workless Station.

For this first segment of finding aids, the Workless Station was able to read successfully, with less than 10 percent error rate, 40 percent of the documents. All of the typefonts included were among those in the OCR machine's programming. The speed with which the machine scanned and converted the data varied according to the quality and clarity of the typed image: on the best quality typed images, the Workless Station could convert 2621 characters per minute with only one or two errors per page, as compared to 708 characters and more errors for a relatively poor image. By comparison, manual keying of the same data could only be accomplished at an average rate of 223 characters per minute. The Workless Station exhibited an "unreadable page" message whenever the error rate for a particular page reached or exceeded 20 percent; however, work experience suggested that it was more efficient to use manual keying whenever the OCR error rate exceeded 10 percent.

The final stage in the desktop OCR project was the loading of the converted data into a data base management system (dbms). Most dbms packages require that data be manually keyed into fields and they are not appropriate when the data is already in machine-readable form. After some investigation, a software package was identified which did allow the automatic loading of machine-readable data into fields. The Group L Corporation's Textbank program provides a list of machine codes through which the user establishes field or zone identification rules; the loading of the data and the preparation of the file for full text retrieval searches is then accomplished automatically.

Textbank runs on a variety of microcomputers and minicomputers, requires 514K (preferably 640K) resident memory, and provides a full range of text search and retrieval capabilities—including the retrieval of words, phrases, dates, and numbers, using the Boolean logic connectors as well as adjacency, proximity, truncation, and numeric range searches.[2] The simplicity of the search commands and procedures makes it possible to use the program successfully with minimal training.

By the middle of 1986, NSZ had completed the pilot phase of the desktop OCR project and had begun to demonstrate the new capabilities for the manipulation and cross-referencing of

[2]The Boolean Operators used by Textbank enable a searcher to qualify a search for two or more words, phrases, dates, etc., with the connecting words "or, and, not, or xor." The meaning of the connectors is as follows:

or at least one of the terms must occur; both may occur but one is sufficient
and both terms must occur
not the first term must occur but not the second
xor only one or the other of the terms may occur, not both.

Proximity searches require that two terms sought be located in the text within a specified number of words, sentences, or paragraphs of each other. The searcher may also specify that two terms be retrieved only when located directly adjacent to each other and in the stated order. Truncation or wild card searches enable a retrieval to include variants of a particular search term with wild card characters as prefix, infix, or suffix. The Textbank wild cards are as follows:

* any number of characters or no character
? one character
[] encloses specific alternate characters.

Numeric range searches enable an inquiry to ask for those fields where a specific numeric value is greater than or less than a value stated in the retrieval request.

the Still Picture Branch's record group descriptions. Branch personnel will continue the data entry process for other selected finding aids, using the OCR machine whenever warranted by the clarity of the typed image.

There were several lessons learned from this desktop OCR project. The Workless Station did provide a significant savings in data entry time for a major segment of a record series, but only for documents with clearly defined type. The errors made by the OCR machine were relatively easy to locate and correct because the Workless Station operated in conjunction with an excellent word processing program on a microcomputer. Another value of the word processing connection was that branch personnel were able to obtain formatted copy of the record group descriptions for an Archives publication without additional data entry. Finally, the product of this project was a highly sophisticated text retrieval system created through relatively inexpensive desktop machines and off-the-shelf software packages, indicating that optical character recognition was, in this instance, both effective and inexpensive.

The second project undertaken by NARA is larger and more significant. When the *Technology Assessment Report* was issued, the National Archives had not located a single OCR vendor who was able, or who expected soon to be able, to read cursive handwriting. There were a few special applications in which hand-printed alphanumeric characters had been read by OCR processes. In late 1984, however, there were reports that an English company had broken the handwriting barrier, with accuracy rates of 99.9 percent and fee schedules less than those for manual data entry.

In the fall of 1985, Archives staff members visited an Optiram Inc. facility outside London, England, taking sample archival documents for OCR conversion, including poor resolution type and handwritten documents. Copies of some of these documents and the printouts after conversion are shown in Figures 4 and 5. Optiram founder John Jenkins had developed OCR software that exceeded the capabilities of matrix matching and feature analysis, apparently through the use of linguistic mathematical probability rules, and extended analyses of word shapes. By early 1986, there were published reports that the handwriting barrier had been broken by other organizations as well.

To test the capabilities of this new OCR technology and determine the extent to which it was available, Archives management officials decided to proceed with the development of a formal research test solicitation. Published in the summer of 1986, the purpose of the solicitation was "to obtain information on the application of optical character recognition to the conversion of archival documents," including documents that were typed, handwritten, printed, and on microfilm. The specific technology questions addressed in the research test included the following:

1. Can a variety of forms of information display, including but not limited to handwriting, print, poor resolution type, and microfilm, be converted to machine-readable format using OCR technology?
2. What is the impact on the OCR conversion of the various forms of information display described above of factors such as the size or type of paper, color or clarity of ink or paper, or quality of microfilm image or stock?
3. What is the accuracy, throughput rate and cost of OCR conversion for each of the types of information display?

4. How effective and accurate is the preparation of the data by the vendor after conversion for loading in data base or word processing systems (insertion of field delimiters, identification of field location on document, etc.)?

The series of documents from the Archives selected for conversion included records that were significant for their reference value, preservation value, scholarly interest, genealogical interest, or viability as a discrete record series, and included (but were not limited to) the following:

Muster Rolls of U.S. Navy Ships, 1880-91 (18,400 pages in bound volumes)

Presently arranged chronologically by year and quarter, and thereunder alphabetically by name of ship, the muster rolls contain unique service information for enlisted men in oversized handwritten registers. Conversion to machine-readable form would enable the various entries for each enlisted man to be combined (a subject of great genealogical interest), enable historians of social history to pursue many forms of inquiry into the lives and service histories of individuals, and reduce the deterioration through use of the original volumes.

Master Abstracts of Enrollments and Registers of Vessels, 1815-66 (17,000 pages in bound registers)

The unique information in these handwritten registers is arranged chronologically and by port, making it difficult and time-consuming for archivists and researchers conducting vessel histories.

Compiled Military Service Records, Civil War

A small segment of this extensive record series (more than a million items) was selected for inclusion in the OCR project to ascertain whether a name index could be automatically created from the information on the jacket covers of these records. There are plans to create an expensive manual index to the entire series.

Catalogue of the Combined British-American records of the Mediterranean Theater, 1941–45 (3300 pages)

This oversized bound catalog, in three volumes, contains poor quality typewritten information which is a unique and valuable index to a large microfilm publication. The catalog is poorly arranged in its present form; conversion to machine-readable form for text retrieval would significantly enhance the research value of the information.

Captions for Documerica color slides, 1972–77

The 16,500 typed captions for a significant series of slides are available to researchers only on microfiche while the slides are published in a separate series of microfiche. Converting the captions would make them more useful as an index tool and would greatly enhance the informational value of the data.

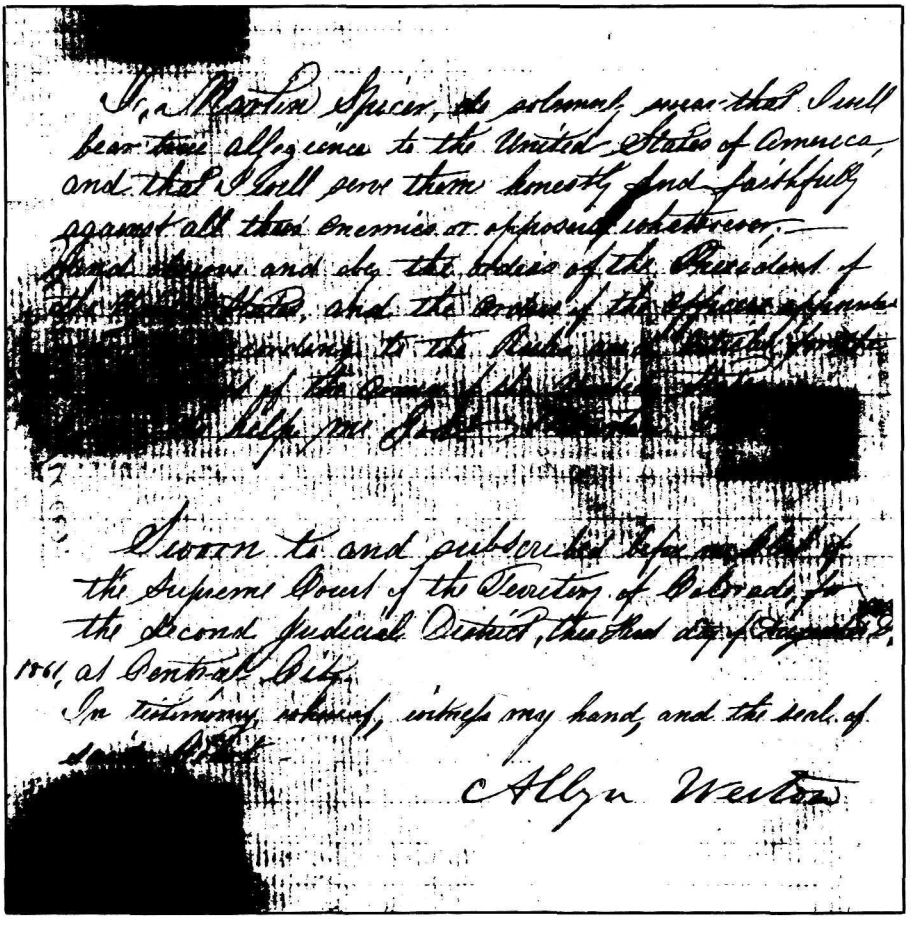
Naval Photograph Center Documentary Film Catalog, 1941–60s (19,000 cards)

The card index for this important film series is cumbersome to use because of the complicated subject classification system under which the cards are arranged. Conversion and manipulation of the cards would resolve problems with the subject system and facilitate reference within and outside the Archives building.

Responding to interest and inquiries from other federal agencies, the Archival

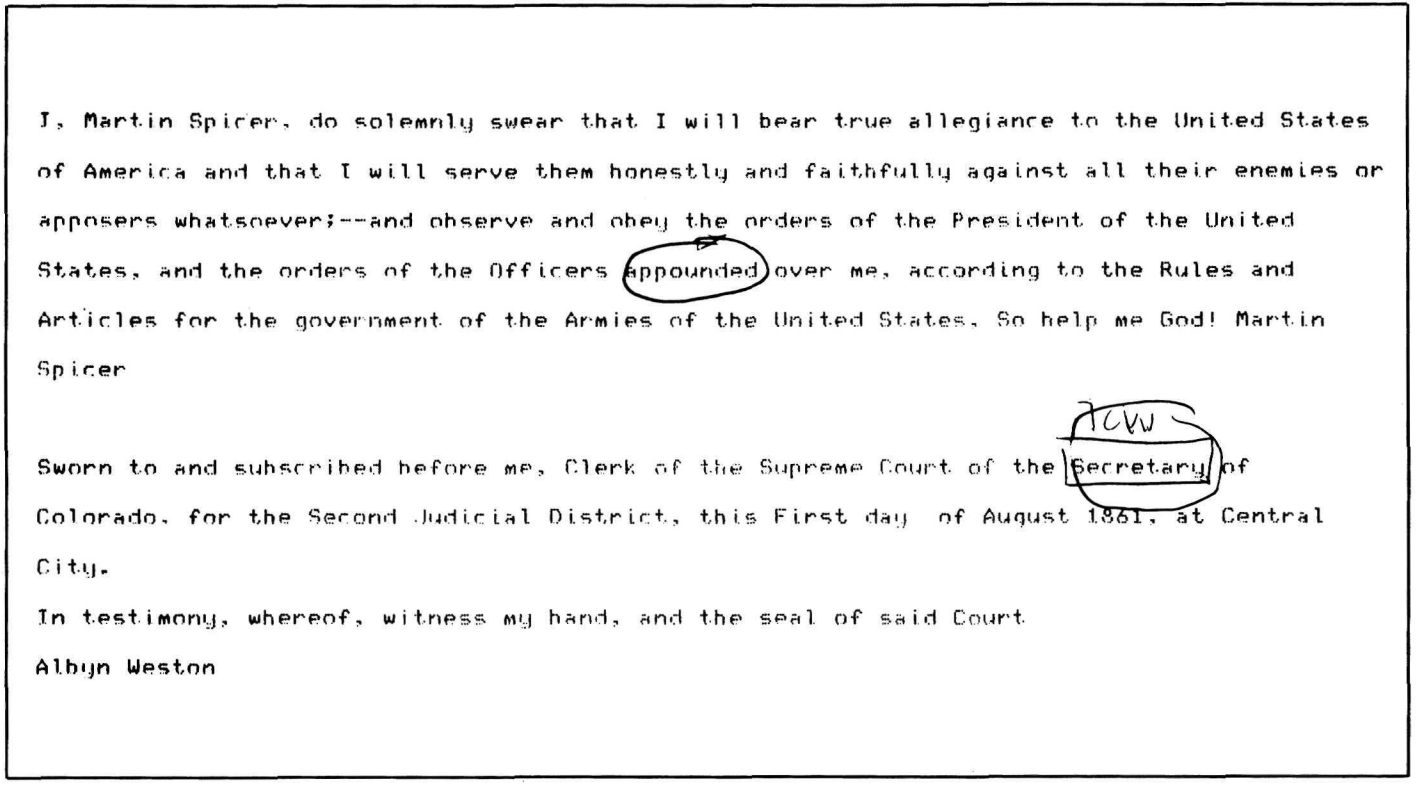
Research and Evaluation Staff conducted briefings in October and November 1985 concerning the recent OCR developments and the proposed research test solicitation, offering other agencies the opportunity to join the solicitation on a reimbursable basis. The Archives research test now includes not only Archives projects, but also ones from the Library of Congress, Smithsonian Institution, Justice Department, and Department of Defense (Library Component). In each case, a significant series of records will be converted and an even more significant collection of statistics on conversion successes and failures will be compiled.

A technology of limited interest to archivists only a few years ago, optical character recognition has come into the forefront of the National Archives' examinations of significant new data processing developments. Whether the subject is bar codes or desktop OCR machines or even the conversion of cursive handwriting, OCR has new relevance for archivists.



This handwritten nineteenth-century document is part of the holdings of the National Archives.

Figure 4A



I, Martin Spicer, do solemnly swear that I will bear true allegiance to the United States of America and that I will serve them honestly and faithfully against all their enemies or opposers whatsoever;--and observe and obey the orders of the President of the United States, and the orders of the Officers appounded over me, according to the Rules and Articles for the government of the Armies of the United States. So help me God! Martin Spicer

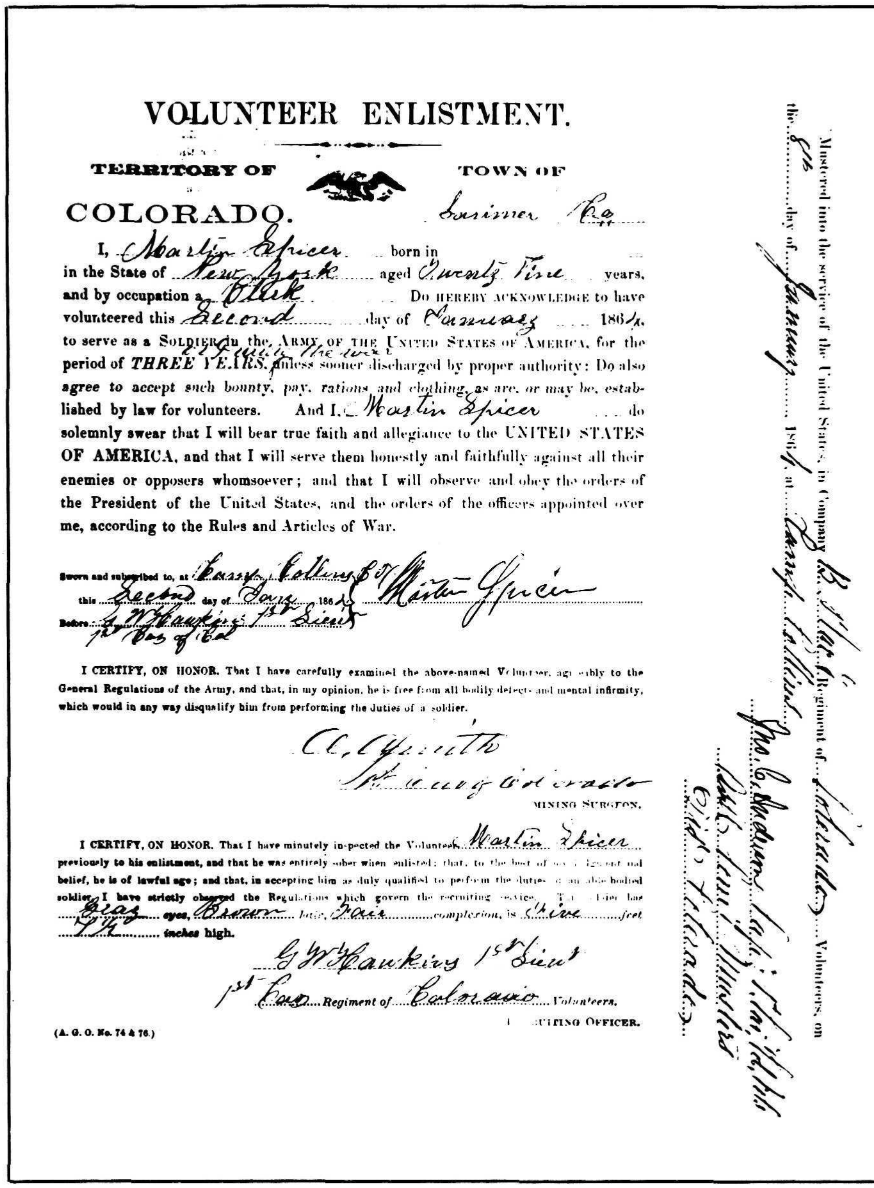
Sworn to and subscribed before me, Clerk of the Supreme Court of the Secretary of Colorado, for the Second Judicial District, this First day of August 1861, at Central City.

In testimony, whereof, witness my hand, and the seal of said Court

Albyn Weston

Printout after OCR conversion of the document. The two errors are circled.

Figure 4B



VOLUNTEER ENLISTMENT.

TERRITORY OF COLORADO. TOWN OF Larimer Co

I, Martin Spicer born in in the State of New York aged Twenty Five years, and by occupation a Clerk Do HEREBY ACKNOWLEDGE to have volunteered this Second day of January 1864, to serve as a SOLDIER in the ARMY OF THE UNITED STATES OF AMERICA, for the period of THREE YEARS, unless sooner discharged by proper authority: Do also agree to accept such bounty, pay, rations and clothing, as are, or may be, established by law for volunteers. And I, Martin Spicer do solemnly swear that I will bear true faith and allegiance to the UNITED STATES OF AMERICA, and that I will serve them honestly and faithfully against all their enemies or opposers whomsoever; and that I will observe and obey the orders of the President of the United States, and the orders of the officers appointed over me, according to the Rules and Articles of War.

Sworn and subscribed to, at Camp Collins C.T.
this Second day of Jany 1864 Martin Spicer
Before G W Hawkins 1st Lieut 1st Cav of Col

I CERTIFY, ON HONOR, That I have carefully examined the above-named Volunteer, agreeably to the General Regulations of the Army, and that, in my opinion, he is free from all bodily defects and mental infirmity, which would in any way disqualify him from performing the duties of a soldier.

MINING SURGEON.

I CERTIFY, ON HONOR, That I have minutely inspected the Volunteer Martin Spicer previously to his enlistment, and that he was entirely sober when enlisted; that, to the best of my knowledge and belief, he is of lawful age; and that, in accepting him as duly qualified to perform the duties of an able-bodied soldier, I have strictly observed the Regulations which govern the recruiting service. This soldier has Gray eyes, Brown hair, Fair complexion, is Five feet 7½ inches high.

G W Hawkins 1st Lieut
1st Cav Regiment of Colorado Volunteers.
RECRUITING OFFICER.

(A. G. O. No. 74 & 76.)

Mustered into the service of the United States, in Company B 1st Cav Regiment of Colorado Volunteers, on the 8th day of January 1864, at Camp Collins

This handwritten nineteenth-century document is part of the holdings of the National Archives.

Figure 5A

VOLUNTEER ENLISTMENT.
TERRITORY OF TOWN OF
COLORADO.
Sarimen Co.
I, Martin Spicer born in
in the State of New York aged Twenty-Nine years,
and by occupation a Clerk DO HEREBY ACKNOWLEDGE to have
volunteered this Second day of January 1864,
to serve as a SOLDIER in the ARMY OF THE UNITED STATES OF AMERICA, for the
period of THREE YEARS, unless sooner discharged by proper authority: Do also
agree to accept such bounty, pay, rations and clothing, as are, or may be, estab-
lished by law for volunteers. And I, Martin Spicer do
solemnly swear that I will bear true faith and allegiance to the UNITED STATES
OF AMERICA, and that I will serve them honestly and faithfully against all their
enemies or opposers whomsoever; and that I will observe and obey the orders of
the President of the United States, and the orders of the officers appointed over
me, according to the Rules and Articles of War.
Sworn and subscribed to, at ◇
this Second day of June 1864, Martin Spicer
Before ◇
1st ◇
I CERTIFY, ON HONOR, That I have carefully examined the above-named Volunteer, agreeably
to the
General Regulations of the Army, and that, in my opinion, he is free from all bodily
defects and mental infirmity,
which would in any way disqualify him from performing the duties of a soldier.

EXAMINING SURGEON.
I CERTIFY, ON HONOR, That I have minutely inspected the Volunteere, Martin Spicer
previously to his enlistment, and that he was entirely sober when enlisted; that, to the
best of my judgment and
belief, he is of lawful age; and that, in accepting him as duly qualified to perform the
duties of an able bodied
solider, I have strictly observed the Regulations which govern the recruiting service.
This soldier has
Grey eyes, Brown hair, Fair complexion, is Five feet 71/2 inches high.
GW Hawkins 1st Lieut
1st ◇ Regiment of Colorado Volunteers.
RECRUITING OFFICER.
(A. G. O. No. 74 & 76.)
Mustered into the service of the United States, in Company B ◇ Regiment of Colorado
Volunteers, on
the 8th day of January 1864, at Camp Collins

Printout after OCR conversion of the document. Errors are indicated by diamond-shaped symbols. Note that the information written in the right margin of the original document has also been successfully converted.

Figure 5B

Variations in the Processing Rates on the Magnuson and Jackson Senatorial Papers

ULI HALLER

Senators Warren G. Magnuson (WGM) and Henry M. Jackson (HMJ) jointly represented the state of Washington in the U.S. Senate for a total of twenty-eight years, Magnuson serving from 1944–1980 and Jackson from 1952–1983. Their combined 3,000 linear feet of personal papers were transferred to the University of Washington Libraries Manuscripts Division following Magnuson's defeat in 1980 and Jackson's death in 1983. In 1984 the state legislature passed a supplemental appropriation to fund processing of these materials.

Because the Magnuson/Jackson Congressional Papers Project was funded, staffed, and housed as a separate unit within the Manuscripts Division, it was possible to closely monitor the progress of the processing on the papers. As might be expected, the rate of progress varied depending upon the type of material being processed and the level of detail to which that material was processed. The work rate varied not only between different series (e.g., general correspondence vs. press releases), but also between similar series (e.g., early WGM personal correspondence vs. later WGM personal correspondence).

Study Methodology

The data reported below were not the product of a special study. They were derived from the project's weekly payroll sheets. Both processors and supervisors were asked to log the actual time spent working on particular series. Although the supervisors tried to verify the information on the time sheets, the voluntary nature of the reporting and the fact that work was reported only to the nearest half hour, may have allowed some distortions to creep into the data.

Because of staffing and funding changes, collection of this data was stopped shortly before the end of the first year of the project. Both the Magnuson and Jackson papers were processed in segments, each accession covering approximately twelve years. The figures reported below cover the following accessions: WGM-3 (1945–1956), WGM-4 (1957–1968), HMJ-2 (1940–1951), and HMJ-3 (1952–1963). The accessions were processed in the listed order. No detailed records of the processing were kept prior to WGM-3 or after HMJ-2.

The completed processing was assessed using two different measures: first, by the traditional time/volume ratio, in this case, hours/cubic foot; second, by the number of inventory entries generated. Inventory entries were measured by inventory lines (IL) and index terms (IT). Counting the number of inventory lines (excluding phrases that run two or more lines) gives a rough measure of the level of detail provided by the processing. While such a count is not a precise measure of the access provided to the archival materials by the inventory, it does give a reliable estimate. For example, a twenty-line inventory of one cubic foot of materials certainly provides less description, and presumably less access, than a 200-line inventory of those same materials. Just how much more access the 200 lines would provide is an open question, one that could be answered only through empirical testing or recall and precision. The IL and IT counts referred

Haller was formerly the Congressional Papers Librarian at the University of Washington.

to throughout this article should thus not be regarded as quantitative measures of access, but simply as measures of the relative levels of detail of the processing, meant only to add depth to the hourly work rate measures. To assess processing one needs to know both how fast the archivist processed a cubic foot of material and what was accomplished during the processing; the IL and IT counts provide one type of quantitative measure of accomplishment. Since the IL figures also equal the folder heading counts—minus duplicates—the IL counts also offer a close approximation of the number of folders handled during processing.

One of the reasons IL counts alone cannot measure access is because many IL contain only generic information, "P Miscellaneous." Such lines do provide meaningful access to users browsing through a particular inventory, but they do not include terms that can be posted to cumulative indexes to help users find inventories of interest. For this reason, a separate count of the indexable terms in each inventory was also made. Figure 1 is a sample inventory page; it includes 32 IL and 17 IT.

Average Processing Rate and Cross-Repository Comparisons

The processing of the individual Magnuson and Jackson record series varied from a low of 1 hr./cu. ft. to a high of 40 hrs./cu.ft. To put these processing rates into perspective, it is necessary to compare them to other reported rates. While the primary purpose of this article is to examine variations in the processing rates for individual series, for cross-repository comparison purposes the Magnuson/Jackson project data must be aggregated into an overall average processing rate. Table 1 provides such an overview.

Figure 1

Sample Inventory Page (WGM-3)

PERSONAL CORRESPONDENCE
(cont.)

Box/Folder	Series
1946	
168/69	R
168/70	Schenck, Joseph
168/71	Stern, William
168/72-74	S
168/75	Teeter, John H.
168/76	T
168/77	University of Washington
168/78	Von Herberg, J.
168/79	Wasmer, Louis
168/80-82	W
168/83	Yoris, E. W.
168/84	Z
168/85-86	Christmas Greetings
168/87-94	Invitations
169/1-13	Invitations
169/14	Thank Yous
1947	
169/15	Allen, Edward W.
169/16-17	A
169/18-20	Bartley, Bruce
169/21	Brady, Thomas
169/22	Brand, Harry
169/23	Broderick, Henry
169/24-27	B
169/28	Cohn, Al
169/29-30	C
169/31	Drumheller, Joseph
169/32	D
169/33-34	Edris, William
169/35	E (incl. Einstein, Albert)
169/36	F
169/37	Golden, Bill
169/38	G

Table 1

Processing Summary

ACC.	CU.FT.	HRS.	IT	IL	HRS./ CU.FT.	IT/ CU.FT.	IL/ CU.FT.	IT/ HR.	IL/ HR.
WGM-3	288	1858	3525	6024	6.5	12.2	20.9	1.9	3.2
WGM-4	396	1173	3491	5314	3.0	8.8	13.4	3.0	4.5
HMJ-2	72	319	1297	1528	4.4	18.0	21.2	4.1	4.8
HMJ-3	294	625	3323	4178	2.1	11.3	14.2	5.3	6.7
Overall	1050	3975	11,636	17,044	3.8	11.1	16.2	3.0	4.3

The information in the table is straightforward. For example, the 288 cu.ft. of WGM-3 were processed in 1858 hours, a rate of 6.5 hrs./cu.ft. In that time processors listed 3,525 index terms in 6,024 inventory lines, a rate of 1.9 IT/hr. and 3.2 IL/hr. By volume there were 12.2 IT/cu.ft. and 20.9 IL/cu.ft.

Table 1 does show a weak correlation between IL/cu.ft. and Hrs./cu.ft.: the higher the IL/cu.ft., the higher the hrs./cu.ft. This supports the intuitive notion that when there are more folders in a box, it takes longer to process that box. The steady increase in IT and IL produced per hour reflects both the increasing experience of the processors and the streamlining of work procedures.

The average processing rate of 3.8 hrs./cu.ft. is substantially lower than the 20–25 hrs./cu.ft. average often expected for manuscript materials. There are three possible explanations. First, the University of Washington processing methods and/or personnel may be superior. This conclusion cannot be supported, given the slim statistical base on which these figures rest. The following two explanations seem more reasonable.

Both the Magnuson and Jackson papers more closely resembled corporate office records than classic personal papers. With the exception of several cartons of stray items, the records arrived at the university in clearly-labeled folders, many of which were still organized in their original series order. Skeletal container lists accompanied the boxes. Although most of the container lists proved inadequate or misleading, the inherent folder level organization in the records helped speed processing. For this reason it is most useful to compare the Magnuson/Jackson project figures only with those for corporate record accessions.

The third explanation for the lower processing rate is that the reporting of archival work rates is not yet standardized, leading to some confusion over how to interpret those rates. The figures reported here are roughly in line with those previously reported by the University of Illinois at Urbana-Champaign, but they are lower than those reported by Washington State University and are much lower than those reported to the NEH and the NHPRC on a number of grant proposals.[1] Yet Terry Abraham, *et al.*, have argued that such discrepancies between these processing times are mainly due to the habit of reporting the total

[1]William J. Maher, "Measurement and Analysis of Processing Costs in Academic Archives," *College and Research Libraries* 43 (January 1982): 59–67; "The Importance of Financial Analysis of Archival Programs," *Midwestern Archivist* 3 (1978): 3–23; Terry Abraham, Stephen E. Balzarini, and Ann Frantilla, "What is Backlog is Prologue: A Measurement of Archival Processing," *American Archivist* 48 (Winter 1985): 31–44; Karen Temple Lynch and Thomas E. Lynch, "Rates of Processing Manuscripts and Archives," *Midwestern Archivist* 7 (1982): 25–33.

time accessions are "in process" rather than the time actually spent on processing. Since the Magnuson/Jackson project figures only include actual processing time, one would expect them to be relatively low.

One must also ask whether the whole cross-repository comparative rate issue is even relevant. As mentioned above, it is not just the total processing time, but what is done in that time that counts. Since processing and inventory construction procedures often vary from repository to repository, cross-repository comparisons of work rates may not be very useful. What the University of Illinois calls series level control may be very different from what the University of Washington calls series level control. Each may be perfectly functional, but unless they are the same product and are arrived at in the same way, any comparison of processing rates between the two repositories may be meaningless. Yet, even if processing rates cannot easily be compared between repositories, they are useful for analyzing the work done within a given repository. Even within the limited scope of the Magnuson/Jackson project there was s sizable variation in processing rates, and the variations bear examination.

Tasks Other Than Refoldering

A detailed breakdown on the time spent on each processing task for each accession is provided in Tables 2-5. Figure 2 is a key to the task name abbreviations. The following is not a complete description of what should be done to process

Figure 2

Task Name Abbreviations

B	Legislation—Sponsored Bills
CPN	Campaign Materials
DP	Departmental Correspondence
DSC	Discarded
GC	General Correspondence
GE	Guide Enhancement (outside staff hired to research and write narrative description of accession materials; discontinued after WGM-3)
INV	Inventory typing and indexing
JB	Jobs (appointments and nominations for Senate confirmation)
LC	Legislative Correspondence
MV	Moving materials to and from remote storage
NB	Numbering folders
OUT	Outgoing letters
P/P	Personal/Political Papers
PC	Personal Correspondence
PF	Project Files
PL	Planning (initial series sorting and appraisal)
PR	Press Releases
S/W	Speeches and Writings
SC	Scrapbooks
SEG	Segregated (case files put aside for later processing)
SG	Subgroups (mainly committee files)
SS	Subject Series
WS	Washington State Matters (special legislative issues)

Table 2

WGM-3

Task	Cu.Ft.	Hours	IT	IL	Hrs./Ft.	IT/Ft.	IL/Ft.	IT/Hr.	IL/Hr.
B	6.0	58.0	224	275	9.67	37.33	45.83	3.86	4.74
CPN	16.0	228.5	219	326	14.28	13.69	20.38	0.96	1.43
DP	66.5	284.0	1715	2425	4.27	25.79	36.47	6.04	8.54
DSC	26.5	0.0	0	0					
GC	7.5	23.0	200	418	3.07	26.67	55.73	8.70	18.17
GE	0.0	160.0	0	0					
INV	0.0	135.0	0	0					
LC	16.0	164.0	276	320	10.25	17.25	20.00	1.68	1.95
MV	0.0	69.0	0	0					
NB	0.0	54.0	0	0					
OUT	40.5	61.5	1	59	1.52	0.02	1.46	0.02	0.96
P/P	3.0	122.5	15	68	40.83	5.00	22.67	0.12	0.56
PC	15.0	97.0	338	650	6.47	22.53	43.33	3.48	6.70
PF	6.5	67.0	101	183	10.31	15.54	28.15	1.51	2.73
PR	3.0	77.0	145	677	25.67	48.33	225.67	1.88	8.79
S/W	4.0	31.0	38	145	7.75	9.50	36.25	1.23	4.68
SC	10.0	2.5	2	39	0.25	0.20	3.90	0.80	15.60
SEG	48.0	0.0	0	0					
SG	14.5	100.0	180	308	6.90	12.41	21.24	1.80	3.08
SS	5.0	124.0	71	131	24.80	14.20	26.20	0.57	1.06
Overall	288.0	1858.0	3525	6024	6.45	12.24	20.92	1.90	3.24

congressional papers. This article simply describes a portion of what was done to several of the series in the Magnuson and Jackson collections. As will become evident, in a few cases mistakes were made and in several cases the initial experiences led to modifications in processing instructions. The reported figures reflect this learning process.

Refoldering and alphabetizing accounted for approximately 80 percent of the total time spent on each accession. Moving boxes to remote storage (MV), numbering folders (NB), and typing and indexing inventories (INV) accounted for another 15 percent of the time. Work planning (PL), including series level appraisal and initial arrangement, made up the final 5 percent of the total time. These relative proportions varied slightly from accession to accession.

The relatively low percentage of appraisal and arrangement time is no accident since most of that work was done at the box and folder levels. The reasonably coherent order of the records when received contributed to this low time, but the project supervisors still had to sift through all of the 2,000 plus cartons of materials and had to handle most of the individual folders in order to reconstruct the basic series arrangement. Actually physically doing this arranging, rather than just sketching it out on paper was not only a fast process, but also extremely useful. It focused the subsequent refoldering and alphabetizing, cleared up questions raised by the misleading container lists, and eliminated most of those unwelcome surprise discoveries of "lost" series segments that usually turn up during the course of large processing pro-

Table 3

WGM-4

Task	Cu.Ft.	Hours	IT	IL	Hrs./Ft.	IT/Ft.	IL/Ft.	IT/Hr.	IL/Hr.
B	2.0	26.0	75	102	13.00	37.50	51.00	2.88	3.92
CPN	5.0	35.0	75	186	7.00	15.00	37.20	2.14	5.31
DP	71.0	208.0	1230	1423	2.93	17.32	20.04	5.91	6.84
DSC	77.0	0.0	0	0					
GC	17.5	20.0	52	482	1.14	2.97	27.54	2.60	24.10
INV	0.0	120.0	0	0					
JB	2.0	10.0	52	62	5.00	26.00	31.00	5.20	6.20
LC	21.5	125.5	413	687	5.84	19.21	31.95	3.29	5.47
MV	0.0	36.5	0	0					
NB	0.0	84.0	0	0					
OUT	49.0	51.0	1	72	1.04	0.02	1.47	0.02	1.41
P/P	2.5	10.0	9	28	4.00	3.60	11.20	0.90	2.80
PC	23.0	126.0	581	857	5.48	25.26	37.26	4.61	6.80
PF	3.5	73.0	94	88	20.86	26.86	25.14	1.29	1.21
PL	0.0	30.0	0	0					
PR	24.5	51.0	239	282	2.08	9.76	11.51	4.69	5.53
S/W	2.5	33.5	140	231	13.40	56.00	92.40	4.18	6.90
SC	16.0	1.5	1	52	0.09	0.06	3.25	0.67	34.67
SEG	50.0	0.0	0	0					
SG	11.5	63.0	247	363	5.48	21.48	31.57	3.92	5.76
SS	14.0	31.0	257	368	2.21	18.36	26.29	8.29	11.87
WS	3.5	38.0	25	31	10.86	7.14	8.86	0.66	0.82
Overall:	396.0	1173.0	3491	5314	2.96	8.82	13.42	2.98	4.53

jects. In these three ways the planning time helped reduce subsequent processing times. It is beyond the scope of this paper to discuss the progression of the work on the Magnuson and Jackson papers in any detail. Interested readers will find Lydia Lucas's advice on dealing with large accessions most helpful.[2] It is important to note, however, that progressively refined control measures, such as initial subgrouping and series sorting, are especially effective with large accessions.

Two of the three tasks singled out above—moving and numbering—were a by-product of the physical location of the project. Moving boxes to and from remote storage was necessitated by the separation of the processing area from both the holding area and the final stacks. Although moving involved manually loading and unloading the library's small van on each trip, this procedure proved to be more of a physical nuisance than a major time waster. The numbering of the processed folders was left for last—just before the materials were moved out to the stacks—because the work area was not large enough to allow all of the materials for each accession to be completely assembled at any one time.

[2]Lydia Lucas, "Massive Collections: From Warehouse to Reading Room," *Georgia Archive* 4 (Winter 1976): 56-63; "Managing Congressional Papers: A Repository View," *American Archivist* 41 (July 1978): 275-80.

Table 4

HMJ-2

Task	Cu.Ft.	Hours	IT	IL	Hrs./Ft.	IT/Ft.	IL/Ft.	IT/Hr.	IL/Hr.
B	2.0	9.0	60	70	4.50	30.00	35.00	6.67	7.78
CPN	3.5	35.5	12	70	10.14	3.43	20.00	0.34	1.97
DP	19.0	33.5	267	304	1.76	14.05	16.00	7.97	9.07
DSC	1.0	0.0	0	0					
INV	0.0	34.0	0	0					
LC	15.0	39.5	143	202	2.63	9.53	13.47	3.62	5.11
MV	0.0	2.0	0	0					
NB	0.0	30.5	0	0					
PC	7.0	78.5	513	539	11.21	73.29	77.00	6.54	6.87
PL	0.0	6.0	0	0					
S/W	1.5	3.5	14	16	2.33	9.33	10.67	4.00	4.57
SEG	6.0	0.0	0	0					
SG	7.5	28.0	111	120	3.73	14.80	16.00	3.96	4.29
SS	9.5	19.0	177	207	2.00	18.63	21.79	9.32	10.89
Overall:	72.0	319.0	1297	1528	4.43	18.01	21.22	4.07	4.79

Table 5

HMJ-3

Task	Cu.Ft.	Hours	IT	IL	Hrs./Ft.	IT/Ft.	IL/Ft.	IT/Hr.	IL/Hr.
B	16.0	33.0	333	417	2.06	20.81	26.06	10.09	12.64
DP	88.5	118.0	1142	1427	1.33	12.90	16.12	9.68	12.09
DSC	23.0	0.0	0	0					
INV	0.0	55.0	0	0					
LC	111.0	161.0	973	1230	1.45	8.77	11.08	6.04	7.64
MV	0.0	28.0	0	0					
NB	0.0	52.0	0	0					
PC	8.5	42.5	273	256	5.00	32.12	30.12	6.42	6.02
PL	0.0	11.0	0	0					
PR	1.0	8.0	1	3	8.00	1.00	3.00	0.13	0.38
S/W	4.0	25.0	155	165	6.25	38.75	41.25	6.20	6.60
SC	9.0	5.5	103	102	0.61	11.44	11.33	18.73	18.55
SG	24.0	72.5	290	484	3.02	12.08	20.17	4.00	6.68
SS	4.0	4.5	35	57	1.13	8.75	14.25	7.78	12.67
WS	5.0	8.5	18	37	1.70	3.60	7.40	2.12	4.35
Overall:	294.0	624.5	3323	4178	2.12	11.30	14.21	5.32	6.69

The final accession inventories were relatively long: about 1 page per 2 cu.ft. Nevertheless, the time spent preparing the inventories was relatively short. One reason the inventory work moved quickly was that the work was done using a personal computer-based word processing package (Samna) that included a "go-word" indexing feature.

Refoldering Rates for Various Series

The decision to refolder every series in both the Magnuson and Jackson collections was reached for two reasons: first, the original file folders were in fair to poor condition; second, heavy use of these collections was expected. Since up to 80 percent of the processing time was spent on tasks related to refoldering, it is easy to see that this decision was a key determinant of the processing rates reported for each series.

Processing rates varied greatly, both between different series in a given accession and between similar series in different accessions. Most of these variations can be attributed to the differing characteristics of each series and/or to the processing instructions for each series.

Processing of the outgoing correspondence, departmental correspondence, and legislative correspondence series mainly involved refoldering of materials that on the surface appeared very similar, yet the processing rate for each of these series varied greatly from accession to accession. With the benefit of hindsight and the IL and IT data, it is possible to see why many of the processing rate variations occurred.

Magnuson's outgoing letters (OUT) were processed very quickly (average 1.3 hrs./cu.ft.), largely because the chronological sequence of the materials itself serves as the only "description." Index terms are virtually nonexistent (0.02 IT/cu.ft.) since the folder headings give only span dates. Inventory lines are also few (1.46 IL/cu.ft.) since the inventory description is box by box rather than folder by folder (e.g., "Box 1: Nov. 15, 1945–Jan. 30, 1946"). Most of the items in this series were duplicates of letters scattered throughout the other correspondence series. There were also some non-duplicate telegrams. The series was retained mainly as a convenience for researchers who might be interested in the overall activities of the Senator's office during a given time period. Under these circumstances the project staff felt this series warranted little attention, hence the "quick and dirty" processing.

The departmental correspondence (DP) was arranged by agency and occasionally subdivided by subject. Processing included refoldering, removing paperclips and rubber bands, and straightening or buffering selected materials. The original arrangement of these materials was scrambled and had to be reconstructed. Almost every folder had to be touched in the process. These additional steps—above and beyond the simple refoldering undertaken on the OUT—are reflected by the higher processing rates for the DP (1.33–4.27 hrs./cu.ft. vs. 1.04–1.51 hrs./cu.ft.). Also in contrast to the OUT, the DP description is much fuller, with 12–25 IT/cu.ft. and 16–36 IL/cu.ft. The refoldering rate (IL/hr.) tended to improve as the IT/cu.ft. and IL/cu.ft. decreased, but WGM-4 bucked this trend. The falling IT/cu.ft. ratios indicate that the senatorial staffs did not create as many special subject files as time went on. The falling IL/cu.ft. ratios reflect the increasing volume of mail to the Senators. This led the staff to utilize more multiple-folder files, in effect reducing the number of different folder headings per foot.

The legislative correspondence (LC) was arranged entirely by topic. Processing was the same as with the DP, though it involved much less rearrangement. Yet even with fewer IT and IL per foot, the LC took longer to process because some other tasks were added. WGM-3 took the longest because more work was done on these materials, including sampling of pressure mail and chronological sorting of folder contents. The content sorting was discontinued after WGM-3 because it was not worth the time: the materials were quite usable even when they were not in perfect order. Also, while the cost of sampling is justifiable in many other situations, it was not in this case. The pressure mail turned out to be less than expected: only about six inches of pressure mail postcards were found and weeded out from among the 21.5 cu. ft. of material. This minimal space savings was not worth the cost of the sampling. Eliminating the sorting and weeding greatly improved the refoldering rates of the last three accessions. Yet WGM-4 still took a long time to process because of the high number of folders (IL/cu.ft.) in that accession. It was not until HMJ-3 that LC processing rate approached that of the DP.

The personal correspondence (PC) involved more work because of its potential richness. To uncover all of the major correspondents within this series, the correspondence was sorted item by item and correspondents with three or more letters were placed in individual folders. This method provided a reasonable amount of name control while leaving most of the less interesting mail in the "Miscellaneous" folders. This item-level processing of the PC took 3–5 times longer than the folder-level processing of the OUT, DP, and LC. The HMJ-2 PC took twice as long as the other accessions because it was a very rich source of names (IT/cu.ft.) and because it included about twice as many IL/cu.ft. That the PC refoldering rates (IL/hr.) are in line with the LC and DP rates (5–9 IL/hr.) further indicates that the IL/cu.ft. ratio was a major factor in slowing the processing of the PC series.

Several other series show the effects of detailed processing, in some cases of overly detailed processing. The almost daily press releases (PR) were a major concern of Senator Magnuson's office staff. Hence each release received its own folder. This was useful as long as multiple copies of each release were kept; however, for archival purposes most extra copies could be discarded. It was nevertheless tempting to maintain the original folder headings. This was done with WGM-3. Unfortunately, this meant virtual item by item refoldering, a task that took almost 26 hrs./cu.ft. The resulting description is superb—48 IT/cu.ft. and 225 IL/cu.ft.—but one has to ask, "Who needs such a detailed description of press releases?" The item refoldering was stopped with WGM-4. For this accession the releases were kept in chronological order but were aggregated month by month. As a result the processing rate decreased to 2 hrs./cu.ft. The description was also less detailed—10 IT/cu.ft. and 11 IL/cu.ft. Individual releases were still listed in the inventory, but only on a key-word basis, not as before with their full original "title." Even this descriptive treatment may have been too detailed given the low potential use for this series, but at least this description was written fairly quickly.

The subject series (SS) presented a different problem. The WGM-3 files were processed in great detail, essentially at the item level, by an overly zealous volunteer. The result was a detailed description but one that in some instances represented the processor's notations—rather than Magnuson's—of subject files. This detailed work was again very

time-consuming, taking 25 hrs./cu.ft. About one-third of this time can be attributed to the added supervisory time needed to clean up this series once it had been "processed." Archival arrangement theory aside, this case is a good example of how expensive it can be to sort subject series when one fights the existing folder headings and of how costly it can be to lose control of volunteer labor. For the other three accessions the processing was confined to refoldering and to clarification of the original folder headings. The resulting processing rates (1.12–2.21 hrs./cu.ft.) are in line with the LC and DP processing rates. This increase in speed was achieved with no loss in descriptive detail: the IT/cu.ft. and IL/cu.ft. for WGM-4 and HMJ-2 remained about the same as for WGM-3, a further indication of the futility of the processing undertaken on WGM-3. The decrease in IT/cu.ft. and IL/cu.ft. in HMJ-3 appears to have been a function of increased file size, since the refoldering rate (IL/hr.) remained about the same as for WGM-3 and HMJ-2.

Summary

Many of the processing rates for the other series in the Magnuson and Jackson papers also exhibited considerable variation, but these will not be discussed here. In general those series that were processed in more detail or that were initially more disorganized usually took longer to process, just as would be expected. The causes of the processing rate variations found in the work on the Magnuson and Jackson papers are also not surprising. The examples presented above make it clear that the relative level of descriptive detail was a very important determinant of the processing rate. Just as important was the initial decision to refolder all of the series. These conclusions are not new. The reported figures merely give some quantitative support to well-established professional rules of thumb.

Detailed processing figures also have little predictive value, even if they do have some analytical value. That is, they cannot be used to suggest which series ought to be processed in detail. They can only indicate which series have been processed at a certain level of detail. Undoubtedly there are many series that, like the OUT, can be processed very quickly but will yield little description. There are many other series that, like the OUT, can be processed very quickly but will yield little description. There are many other series that, like the PC, can be processed in great detail, but only at a relatively slow rate. Most archivists already recognize such series by instinct or experience. The value in detailed processing figures lies in spotting series that, like the WGM-3 SS, are being processed too slowly or should be yielding more access points for the amount of time being invested. The decision to process in detail cannot be based only on such figures, however. That decision must trace back to the qualitative analysis of the series, to the subjective assessment of its potential value to users. Statistics on past use of similar series can help with this decision. Once the potential usefulness of a series has been established, the average processing rate figures can help determine how much staff time should be budgeted to process that series.

Filling the GAP: Planning on the Local and Individual Levels

GREGORY S. HUNTER

Planning has been the focus of much recent attention within the archival profession. Most of this attention, however, has been at the national level: for example, the final report of SAA's Task Force on Goals and Priorities (GAP) and articles by F. Gerald Ham, Edwin C. Bridges, and Richard J. Cox.[1] This article is an attempt to "fill the gap"[2] (pun intended) by considering planning at local and individual levels and exploring the implications and possible implementations of the GAP report for regional archival organizations and for individual archivists and institutions.

The Goals and Priorities Task Force

Archivists have had a love-hate relationship with planning; while there has been an appreciation for and fascination with planning, there has also been a systematic avoidance of its use. Most archivists understand neither the basic terminology nor the methods of planning. This lack of understanding leads to fear, which at its worst leads to avoidance of the whole topic.

The archival profession first became formally involved with planning in 1949, when SAA President Philip Brooks established a special committee to examine the nature, purpose, and composition of the archival profession. As with later efforts, this committee functioned for a short period of time and left no ongoing structure to continue its work. Since 1949 planning, like the locust, has returned periodically, approximately every ten years, to gnaw at the archival profession. The recommendation of the 1959 Advisory Committee on Long-Range Planning for a permanent, paid SAA staff was not implemented until 1974. The "Committee of the Seventies" followed and focused largely on the structure of SAA.[3]

The most recent visitation of planning, the Goals and Priorities Task Force, went beyond previous efforts in two key ways. First, it focused on planning for the entire profession and not just for SAA. This, in turn, has generated the discussion of planning for the local and individual levels. Secondly, the GAP Task Force established an ongoing mechanism —the Committee on Goals and Priorities —for the continuation of planning in the profession. Archivists should not have to

[1] *Planning for the Archival Profession: A Report of the SAA Task Force on Goals and Priorities* (Chicago: Society of American Archivists, 1986); F. Gerald Ham, "Planning for the Archival Profession," *American Archivist* 48 (Winter 1985): 26–30; Edwin C. Bridges, "Can State Archives Meet the Challenges of the Eighties? Four Recent Views on the Condition of American State Archives," *ARMA Records Management Quarterly* 20 (April 1986): 15–21, 52; Richard J. Cox, "Strategies for Archival Action in the 1980s and Beyond: Implementing the SAA Goals and Priorities Task Force Report," *Provenance* 3 (Fall 1985): 22–37. A general definition of planning is: "the management technique of systematically establishing program goals, and organizing staff and allocating resources to meet those goals, by established deadlines." See Bruce Dearstyne, "Planning for Archival Programs: An Introduction," *MARAC Technical Leaflet Number Three* (New York: Mid-Atlantic Regional Archives Conference, 1983). Dearstyne also provides a brief bibliography on planning.

[2] Ham urges archivists to do this, without the pun: "We need to identify gaps and holes" ("Planning," 30).

[3] For more on the history of planning, see *Planning for the Archival Profession*, 2–4.

The author is Manager, Corporate Records and Micrographics, ITT Corporation. *Earlier versions of this paper were delivered at the Long Island Archives Conference on 19 October 1985 and the Archivists Round Table of Metropolitan New York on 8 April 1986.*

begin the process from scratch in another ten years.

The GAP Task Force issued a preliminary report in August 1984 and a final report in March 1986.[4] At the heart of both reports are two items: a statement of the mission of the archivist and the development of three overarching goals for the profession.

On the mission of the archivist, the GAP Task Force was extremely terse and clear: the mission of the archivist is "to ensure the identification, preservation, and use of records of enduring value."[5] Despite its apparent simplicity, this mission statement embodies quite a few nuances. Note, for example, the avoidance of archival jargon that would be confusing or meaningless to the general public. There is no mention of appraisal, accessioning, provenance, or similar technical terms. Note also the definition of archival records as having enduring rather than permanent value. While the point may seem minor, its implications are quite important: "enduring" value allows for the periodic reappraisal of archival records as suggested by Leonard Rapport,[6] as well as the expansion of archival control over long-term (but non-permanent) records now in the custody of records managers.

The three basic goals of the profession, which flow naturally from the mission of the archivist, are (1) the identification and retention of records of enduring value, (2) the administration of archival programs to ensure the preservation of records of enduring value, and (3) the availability of records of enduring value.[7]

The rest of the GAP report develops these themes, using objectives and activities. While it is at this point that the report offers some of its greatest contributions, it is also here that most people close its covers. In terms of readability, archivists need less of a dirge and more of a rousing march, to use a musical analogy.[8]

Numerous questions come to mind when the three major archival goals presented in the GAP report are considered at the local and individual levels. For example, how can these goals make a difference in a regional organization? What are some concrete ways in which a local organization can implement the GAP's recommendations? How can individual archivists become a part of the national movement toward better planning?

Planning on the Regional Level

One logical way for a regional organization to implement the GAP report is by making its recommendations the focus of future meetings. For example, each of three consecutive meetings might focus on one of the three basic goals. Each meeting would explore the relevant section of the GAP report, assess its applicability for the region, and develop specific activities for implementation. The Long Island Archives Conference has already done this.

With or without such a focus for future meetings, the GAP report offers a number of opportunities for any regional organization, which can be seen in a more detailed look at each goal.

[4]See also the GAP Task Force's earlier document: "A Statement of the Mission and Goals for the Archival Profession," *SAA Newsletter*, March 1983, 6–7.

[5]*Planning for the Archival Profession*, vi.

[6]Leonard Rapport, "No Grandfather Clause: Reappraising Accessioned Records," *American Archivist* 44 (Spring 1981): 143–50.

[7]*Planning for the Archival Profession*, 8, 14, 22; Larry J. Hackman, "Historical Documentation in the United States: Archivists—And Historians?" *OAH Newsletter*, August 1985, 17–18.

[8]Even Ham, chairperson of the GAP Task Force, admits that the planning hierarchy "may seem dense and impenetrable." ("Planning," 27.)

The first goal, identification and retention of records of enduring value, has two aspects: archivists must educate themselves about the records of contemporary society and improve archival practices accordingly, and archivists must educate records creators and the general public about the importance of retaining records of enduring value.[9] The following activities suggest ways in which a regional organization can help archivists educate themselves, records creators, and the general public:

(1) Develop a region-wide strategy to document key aspects of the area's life and history.
(2) Develop cooperative collecting policies for regional repositories.
(3) Hold a joint meeting with the local chapters of the Association of Records Managers and Administrators (ARMA) and the Association for Information and Image Management (AIIM).
(4) Hold a workshop on developing records retention schedules, inviting archivists and non-archivists to attend.
(5) Work more closely with regional library consortia.
(6) Encourage archivists to speak about the importance of archival preservation at meetings of local civic groups, fraternal organizations, school districts, professional organizations, and political parties.
(7) Present an annual award to the community group or individual which did the most to advance historical preservation in the region during the previous year, such as the President's Award to be presented by the Midwest Archives Conference in the spring of 1988.
(8) Establish a detailed telephone tree in order to mobilize quickly all members of the regional organization when it is necessary to support legislation and other interests. The Coalition for New York Documentary Heritage has established such a telephone tree.

The second major goal in the GAP report deals with improving the administration of archival programs. After records of enduring value are identified, archivists must store, arrange, and describe the records; provide assistance to patrons; select and supervise staff; and generate and administer funds. The key to accomplishing the goal, however, is to remember that these aspects of administration are not ends in themselves. Rather, they are means toward the ultimate end: promoting the use of the archival records.[10] The following are ways a regional organization can advance this goal of improved archival administration:

(1) Establish a central research collection on archival science, comprised of finding aids, statistical reports, annual reports, policy statements, budgets, job descriptions, and relevant publications deposited by regional repositories. The items could circulate through interlibrary loan.
(2) Develop a regional "Adopt an Archives" program similar to the national one of the SAA College and University Archives Section. As in the SAA program, experienced archivists would give free, informal advice (by mail or telephone) to another archivist. A regional program would encourage the exchange of information, develop new friendships, involve many

[9] *Planning for the Archival Profession*, 8; Hackman, "Historical Documentation," 19.
[10] *Planning for the Archival Profession*, 14; Hackman, "Historical Documentation," 19.

more people in the workings of the organization—and cost next to nothing.

(3) Develop a comprehensive disaster plan for archives and manuscript repositories within the region, similar to the award-winning *Ounce of Prevention* developed by the Toronto Area Archivists Group.

(4) Preserve the records of the regional organization. If there is no central repository for organization records, designate one as soon as possible.

(5) Consider regional cooperative purchasing of archival supplies to achieve volume discounts.

(6) Share specialized facilities (on a charge-back basis), such as conservation labs, fumigation equipment, and microfilm facilities. Publish a directory listing these facilities and explaining their use.

The third goal recommended by the GAP Task Force—encouraging the use of archival records—is really the purpose for archivists' existence. Despite our best efforts to date, the many possible uses of archives are not widely recognized by the public; it does no good to identify and preserve records if they are never used. Archivists need to consciously reduce existing barriers and take positive steps to promote the use of archives.[11] The GAP task force asks archivists to change their thinking, focusing not on the order in which activities are performed but on the ultimate goal of these activities.[12]

Two recent complementary trends have brought use to the forefront. The first, which began with the renewed interest in family history that followed the publication of Alex Haley's *Roots*, has been an increasing use of records by the general public rather than just by scholars. The second, more fundamental trend has moved use to the center of the archival world and defined all other activities by this standard. Articles by Elsie Freeman, and Frank Boles and Julia Marks Young illustrate this trend.[13]

A regional organization can encourage the use of archival records in the following ways:

(1) Encourage the exchange of finding aids within the region.

(2) Hold a workshop on promoting the use of records and identifying barriers to use.

(3) Survey users of the regional archival repositories in order to learn how better to serve them.[14]

(4) Establish liaisons or joint committees with local library, history, genealogy, and similar groups. Invite them to meetings and offer to speak about archives at their meetings.

(5) Encourage schools to invite archivists as guest speakers.

(6) Capitalize upon anniversaries, centennials, and dedications as occasions for raising community consciousness about archival records.

(7) Write feature articles about archival preservation in the region for local newspapers.

(8) Seek underwriting for a slide-tape or videotape production about regional archives.

[11] *Planning for the Archival Profession*, 22; Hackman, "Historical Documentation," 19.

[12] *Planning for the Archival Profession*, 23.

[13] Elsie T. Freeman, "In the Eye of the Beholder: Archives Administration from the User's Point of View," *American Archivist* 47 (Spring 1984): 111–23; Frank Boles and Julia Marks Young, "Exploring the Black Box: The Appraisal of University Administrative Records," *American Archivist* 48 (Spring 1985): 121–40.

[14] See *Midwestern Archivist* 11, no. 1 (1986), for articles on user studies.

To implement all of these recommendations would tax the resources of a large organization but perhaps these suggestions will stir thought about concrete ways in which the GAP report can be implemented on the regional level.

Planning at the Individual Repository Level

It is also important to consider planning at the individual repository level because the planning process forces an institution to answer three fundamental questions:

1. *Why* does the archival program exist?
2. *What* are its goals?
3. *How* can it accomplish these goals?

Difficult though it may be, it is essential that archivists—individually and collectively—attempt to answer these questions.[15]

Every archivist need not develop a lengthy formal plan for his or her repository. Naturally there is a question of balance: a "lone arranger" may be too busy with daily activities to spend months developing and refining a voluminous planning document. No matter what the individual situation, however, archivists as professionals owe it to records creators and donors and to funding sources to provide answers to the three fundamental questions listed above.

The first question—why does the archival program exist?—can be answered through a mission statement or statement of purpose. Not only is the development of such a statement well within the means of every archival institution, it should be one of the priorities of each repository. Developing a mission statement encourages reflection on the ultimate purpose of the archives. Such a statement, once prepared, will be useful for informing administrators, potential donors, and the general public about why the archives exists. A mission statement is an outreach or public relations tool, as well as a planning document.

As mentioned earlier, a regional organization could help its members by holding a session on developing mission statements and encouraging the deposit and exchange of mission statements among repositories. Such a centralized collection of mission statements could be a key resource for demonstrating to the public the importance of archives.

Once the "why" question has been addressed, it is time to turn to "what"—what should the program accomplish? It is easiest to think in terms of two or three basic goals for the repository, such as documenting the experience of a particular ethnic or religious group, encouraging school groups to learn about historical records, or helping labor organizations to identify historically valuable records. Whatever the goals, make them explicit and put them in writing for future reference and revision.

In terms of "how"—the third basic question—planning for a small repository need not be elaborate. While there may be a hierarchy of goals, objectives, and activities, having a hierarchy is not a prerequisite for planning. If the structure becomes an end in itself, an all-consuming purpose, the repository may be better off without it. What is needed is a practical approach that reflects the economic and political realities on the local level. The focus should be on making planning a central part of archival work, not on whether planning follows one particular model.

Planning "how" to achieve repository goals calls for developing specific, unambiguous courses of action. If, for example, the goal is to increase archival aware-

[15]Dearstyne, "Planning for Archival Programs," 7.

ness in the schools, one specific course of action would be to personally contact twenty schools during the year and invite them to visit the repository. At the end of the year there would be no doubt about whether or not the target number of schools had been contacted. While the goal of increased awareness may not have been achieved, at least one can determine whether or not the more detailed plan was met.

A final caution about repository planning: the plans must be possible within available institutional resources. Do not, for example, plan a large direct-mail campaign if there are not sufficient funds for postage. Similarly, do not plan to invite the entire sixth grade from a local school, no matter how well the class visit would achieve the goal of increasing the awareness of historical records among students, if the repository can only accommodate ten visitors at a time. If the archivist does not know what institutional resources are available, it would be appropriate to first carry out a self-analysis or self-study.[16] As with other elements of the planning process, such self-examination need not be elaborate, especially in a small repository. At the same time, also remember that planning may actually increase archival resources. Clearly showing administrators all that needs to be done may over time lead to an increase in the resources allocated to an archival program.

Conclusions

Planning, at whatever level and in whatever form, offers a number of benefits. It forces archivists to define mission and goals. It provides a clear way to justify archival programs to administrators and the general public. It furnishes a framework of accountability for resources expended. And it ensures that archivists will do more than respond to day-to-day crises.[17] These benefits more than justify the time spent on the planning process.

To return to the musical analogy, archival planning is a composition with a great versatility. It can be played by the full symphony—SAA at the national level—or can be played by string quartets—the regional organizations across the country. And it can be played by the solo musician in individual settings.

[16]Ibid., 2. See also *Evaluation of Archival Institutions: Services, Principles, and Guide to Self Study* (Chicago: Society of American Archivists, 1982).

[17]Dearstyne, "Planning for Archival Programs," 2.