

January 1997

Taking a Byte Out of the Senate: Reconsidering the Research Use of Correspondence and Casework Files

Naomi L. Nelson
Emory University

Follow this and additional works at: <https://digitalcommons.kennesaw.edu/provenance>



Part of the [Archival Science Commons](#)

Recommended Citation

Nelson, Naomi L., "Taking a Byte Out of the Senate: Reconsidering the Research Use of Correspondence and Casework Files," *Provenance, Journal of the Society of Georgia Archivists* 15 no. 1 (1997) .
Available at: <https://digitalcommons.kennesaw.edu/provenance/vol15/iss1/4>

This Article is brought to you for free and open access by DigitalCommons@Kennesaw State University. It has been accepted for inclusion in Provenance, Journal of the Society of Georgia Archivists by an authorized editor of DigitalCommons@Kennesaw State University. For more information, please contact digitalcommons@kennesaw.edu.

Taking a Byte Out of the Senate: Reconsidering the Research Use of Correspondence and Casework Files

Naomi L. Nelson

In the mid-1970s, a sustained discussion about the management of modern congressional collections first emerged in archival literature.¹ Much of the debate over congressional collections during the intervening twenty years

¹ Looking back from the perspective of 1994, Senate Historian Richard A. Baker identified several factors leading to an increased public awareness of the disposition of the papers of public officials in the 1970s. The unexpected death of influential Senator Richard B. Russell in 1971 resulted in the very visible transfer of forty-five tons of records in three tractor-trailers to the University of Georgia. Richard Nixon's resignation after Watergate and the legal battle over the ownership of the secret recordings made in the Oval Office led to a debate over which papers created by elected officials should be considered private records and which should be considered public records. Finally, between 1976 and 1980, fifty-three senators left office (through resignation or election defeat), the greatest turnover in Senate history. When the dust cleared, congressional papers remained private records, and increasing numbers of repositories faced the challenge of accessioning the huge collections. See Richard A. Baker, "Congressional Papers: the Legacy of Richard Russell and Richard Nixon," in *Proceedings of the Congressional Papers Conference Held in Portland, Maine, 16-17 September 1994*, eds. Gregory P. Gallant and William E. Brown, Jr. (Waterville, ME: Atkins Printing Service, 1995), 15-21.

concerned the appropriate disposition of the voluminous constituent correspondence and casework files. Most archivists agreed that the casework and constituent correspondence records created and filed under the old paper-based system were bulky, hard to use, and of little research value.²

In the summer of 1976, James K. Benson presented two papers to the Minnesota Historical Society assessing the potential research uses for constituent mail.³ He identified three possible areas of focus: the content of the mail, the people who wrote, and the impact of the mail on the political decision making. He also identified several potential barriers to research use of these records. These barriers included the large volume of the records, the organization of the records, the inconsistency with which information about the

² Almost every speaker at the 1978 Conference on the Research Use and Disposition of Senators' Papers addressed the research value of constituent mail, with many concluding that such files were problematic at best and of little use to the social scientist or historian. Lydia Lucas, however, argued that "the way in which a member defines and expresses his relationship to his constituency, and the way his papers reflect this relationship, also shape their most unique and enduring values"; and Frank Mackaman pointed out that constituent correspondence and case work documented a kind of political participation by non-elite members of society. J. Stanley Kimmitt and Richard A. Baker, eds., *Proceedings of the Conference on the Use and Disposition of Senators' Papers, Washington, DC, September 14-15, 1978* (Washington: U.S. Government Printing Office, 1979); Lydia Lucas, "Managing Congressional Papers: A Repository View," *American Archivist* 41 (July 1978): 280; and Frank Mackaman, remarks during Archivists Panel in *Proceedings of the Conference on the Use and Disposition of Senators' Papers*, 68-9.

³ James K. Benson, "Political Research on Constituent Mail: A Report on Problems and Prospects" (paper prepared for the Minnesota Historical Society, summer 1976), and Idem, "Letters to Congressmen as Sources for Research: A Report on the Constituent Correspondence of Congressman Clark MacGregor" (paper prepared for the Minnesota Historical Society, summer 1976).

constituents appears in the letters, the difficulty of categorizing letter content, and the time needed to estimate total quantities of mail on a given topic.⁴ The congressmen Benson included in his study all used the paper-based filing systems in use in Congress prior to the introduction of automated correspondence management systems.

In 1978, the Senate began to automate the handling of constituent correspondence, and several archivists and records creators expressed hope that automating (or "computerizing") mail processing would solve many of the processing and access problems posed by the voluminous mail and case work files. F. Gerald Ham suggested that "[t]hese records possess great advantages for our users. The information they contain can be rearranged, aggregated, compared, and subjected to statistical tests without the laborious tasks of sample selection, data collection, coding, and data entry."⁵ Margery Sly sagely predicted that "some archivists will be lucky and will be able to use computerization to their advantage; others will be faced with an unholy mess."⁶

Repositories receiving senatorial papers must now evaluate whether the constituent correspondence and casework records created and organized through the use of these early correspondence management systems are easier to access than records created under the paper-based systems and whether automation might offer any benefits to the archivist and researcher. Senator Sam Nunn served from 1972 to 1996, and his papers, now at Emory University (Atlanta, Georgia), provide an example of the types of benefits and challenges offered by correspondence manage-

⁴ Benson, "Political Research," 7-8, 10-11, 15.

⁵ F. Gerald Ham, "Archival Choices: Managing the Historical Record in an Age of Abundance," *American Archivist* 47 (winter 1984): 19.

⁶ Margery Sly, "Access to Congressional Case Files: Survey of Practices, Implications for Use" (paper presented at the annual meeting of the Society of American Archivists, 30 August 1986), 20.

ment systems. The Senate has provided repositories with uniform electronic databases of coded information about constituents and their interests that should appeal to researchers interested in quantitative analysis. The systematization and standardization offered by these files, however, are a mirage. Senate staffers adapted the systems to individual office needs, and the data contain errors and irregularities. Constituent correspondence and casework files continue to be bulky and difficult to use.

Correspondence Management Systems Come to Capitol Hill

The handling of United States Senators' constituent correspondence⁷ did not change appreciably with the adoption of automation.⁸ Staff members answered letters using paragraphs pre-approved by the senator and filed the original letter and a copy of the response for later reference. Indexes provided access to the filed correspondence through key access points, usually including constituent name, subject of the letter, and date of the letter. Staff members also compiled lists of constituent names and addresses for follow-up letters, newsletters, or future mailings and generated

⁷ For the purposes of this article, *constituent mail* and *constituent correspondence* will be defined as including all kinds of correspondence between a member of congress and his constituents. These will include letters on legislative issues, requests for flags and other routine matters, letters requesting that the senator intercede on the constituent's behalf with another federal agency, thank you letters, and mass mailings. Letters on legislative issues will be referred to as *issue mail*, and letters requesting intervention on the constituent's behalf with a federal agency will be termed *casework*. In the Senate, the correspondence management system index provided to the repositories upon the senator's retirement includes all mail indexed on the system, regardless of type.

⁸ For an interesting assessment of the impact of computer applications on Congress itself, see Stephen E. Frantzich, "The Implications of Congressional Computerization," *Bulletin of the American Society for Information Science* 13 (February/March 1987): 13-14.

reports tracking hot topics, mail volume, and other useful derivative information.⁹

In the mid-1970s, Congress embraced automated, word-processing systems as the answer to the increasing volume of constituent inquiries. Senate facilities literally were unable to handle the mountains of constituent mail, and the floors of the Senate office building used to store the addressograph plates began to buckle under the weight of the plates.¹⁰ During a hearing before the Senate subcommittee that oversaw computer services in the Senate, Senator Alan Cranston estimated that in 1979 his office alone received from 10,000 to 15,000 letters per week.¹¹ Members sought a faster way to send high-quality responses to constituents and a more cost-effective way to keep constituents apprised of member activities. They also wanted to reduce staff time spent on producing, filing, and retrieving correspondence and to institute more managerial control over the mail process.

The constituent mail function was automated first by using word processing and then by using increasingly more complex correspondence management systems. Word processing combined technologically more advanced office equipment with a systematic approach to office workflow in order to increase both the quality and volume of correspondence

⁹ See Senate Committee on Rules and Administration, Subcommittee on Computer Services, *Report on Computer Services to the Committee on Rules and Administration*, 95th Cong., 1st sess., 1977, Committee Print, 9.

¹⁰ Stephen E. Frantzich, *Congressional Applications of Information Technology* ([Washington, D.C.]: Office of Technology Assessment, [1985]), 22.

¹¹ Senate Committee on Appropriations, Subcommittee on the Legislative Branch, *Oversight on Computer Services in the Legislative Branch: Hearing before a Subcommittee of the Committee on Appropriations, Special Oversight Hearing, Legislative Branch*, 96th Cong., 1st sess., 1979, Committee Print, 14.

produced.¹² Building on the systematization and standardization provided by word processing, correspondence management systems offered sophisticated word processing; the capability of inserting selected, approved paragraphs; personalized salutations and closings; personalized text; the ability to create targeted mailing lists; correspondence records; mail count on issues; automatic filing; and correspondence tracking.

Starting in the early 1970s, the Senate Computer Center developed the first database systems—the Automated Indexing System (AIS) and the Senate Mail File (SMF). They designed AIS to store the basic identification information about a document (name or subject, date, staffer, city, document number, and so forth) and then to provide lists of the correspondence sorted by any of those fields. The goal was to end the time-consuming practice of maintaining carbon copy cross-reference files and to facilitate faster filing and retrieval time.¹³ The correspondence was filed by a system-generated document number. Name and topic indexes (see figure 1, page 43.) to the senator's correspondence were generated periodically from the AIS so that the staff could locate a letter by name or topic.¹⁴ The SMF was a centralized database of correspondent names and addresses that could be used to create labels or for follow-up mailings. Initially, staff manually typed the information about constituents and correspondence into these databases, but

¹² General Accounting Office, *Federal Productivity Suffers Because Word Processing Not Well Managed: Report to the Congress*, report prepared by the Comptroller General of the United States ([Washington, D.C.]: U.S. General Accounting Office, 1979), 1.

¹³ *Report on Computer Services to the Committee on Rules and Administration* 1977, 11.

¹⁴ Karen Dawley Paul, *Records Management Handbook for United States Senators and their Repositories* (Washington, D.C.: U.S. Government Printing Office, 1992), 50.

Name index (generated by CMS)

09-25-81	INDEX FILE FOR SENATOR X	NAME REPORT	DOCUMENT #	MICROFILM #
	CRUMLEY, HARRY B., MR. 343 JENIFER STREET PORTLAND, OR 97201 MJM WHITE HOUSE TOUR	REQUEST TOURS	09-22-81 WASHINGTON DC	1264062007
	JOHNSTON, JENNY, MRS. 1802 NORTH 24TH STREET SPRINGFIELD, VA 22100 ADT ITEMS: 3, 11, 5	ISSUE ENVIRONMENT INTERIOR	09-21-81 WATER NATIONAL PARKS	1263101001
	MULLIGAN, BART, MR. AND MRS. 492 COPELY LANE BORDENTOWN, VA 22102 GPS ITEMS: 1, 47, 6	CASE CONST INITIAL	09-24-81 CONTACTED AGENCY	1266100002
	RUTHERFORD, EILIS, MS. APARTMENT 499 230 KEY ROAD ARLINGTON, VA 22101 ADA ITEMS: 1, 20, 0, 7	ISSUE INTERIOR	09-21-81 NATIONAL PARKS	1263101002

Figure 1

with the adoption of the Senate's Correspondence Mail System (CMS) in 1978, they could download information in batch files from the CMS system to the AIS and SMF.¹⁵

The centralized constituent mail system known as CMS was designed to "perform centralized indexing, filing, and retrieval functions and maintain central indexes and mailing lists in accordance with Senate rules."¹⁶ Like the AIS, it produced indexes. In addition, it included a topic listing that allowed for easier cross reference for letters with multiple topics. CMS could produce reports to help office managers summarize the opinions expressed in incoming mail and to assess the efficiency and effectiveness of office staff in responding to mail. It cost more per letter, but the additional capabilities it offered were supposed to compensate for this extra expense. By-products from the system included management reporting; casework management; high speed, production printing; mailing list maintenance; and indexing and filing of correspondence. These additional capabilities became a part of the offices' correspondence function.¹⁷ In the late 1980s, CMS was upgraded and renamed the Constituent Services System (CSS). In 1991 the Senate Mail

¹⁵ Ibid., 50.

¹⁶ General Accounting Office, *The Senate Should Explore Other Word Processing Alternatives to Improve Cost Effectiveness and Productivity: Report by the Comptroller General to the Chairman, Committee on Appropriations, United States Senate* (Washington, DC: U.S. General Accounting Office, 1980), 6.

¹⁷ General Accounting Office, *The Senate Should Explore Other Word Processing Alternatives* (1980), 6, 10-11. Eight-eight percent of the offices using CMS reported that they found the CMS management reports useful. Offices that did not use CMS generated the workload and hot topic reports manually. In contrast, only twenty-five offices used the casework subsystem, and some senators complained that the system included features that they did not want to use.

System (SMS) was developed to replace CSS, the SMF, and the AIS with a single database.¹⁸

In 1994, the Senate Computer Center decided to stop supporting SMS and began the process of moving all the Senate offices still using SMS to stand-alone correspondence management systems developed by outside vendors. These systems were designed for local area networks (LANs) and located in the senators' District of Columbia offices. The transition to the new systems was completed in 1996. Approved systems included InterAmerica's *CapitolCorrespond*, Intelligent Solutions, Inc.'s *Quorum*, and Electronic Data Systems' *Quick Response*. Because these new systems resided in the senators' offices, they gave both more control and more responsibility to senators and their staffs. Individual office staffs designed and generated their own reports, and those senators interested in having a mail file for mass mailings had to maintain it in-house.

When a senator left office, the Senate Computer Center sent a copy of selected data fields from the correspondence management systems to his or her designated repository. (See figures 2 and 3, pages 46 and 47.) Since the center created the files using proprietary software that the repositories could afford neither to purchase nor to maintain, they sent data in a flat ASCII format that could be accessed using other software. Prior to 1996, they transferred files using seven-inch magnetic reels, nine-inch magnetic reels, or data tape cartridges. In 1996, they sent the files on CD-ROMs.

Electronic files stored on seven- and nine-inch reels require the use of a mainframe, and even files stored on data cartridges and CD-ROMs require large amounts of storage space and specialized software. Understandably, repositories have not been anxious for researchers to use these files and have not worked to make them accessible by researchers. A

¹⁸ Paul, *Records Management Handbook*, 51.

Record layout for Correspondence Management System files
sent to repositories by Senate Computer Center in 1996

Field	Length	Position	Type
1. Name (last, first middle, prefix, suffix) (Ex. Public, John Q., Mr., Jr.)	39	1-39	char
2. Title	30	40-69	char
3. Organization	30	70-99	char
4. Address line 1	30	100-129	char
5. Address line 2	30	130-159	char
6. City	30	160-189	char
7. State code	2	190-191	char
8. Zip code	10	192-201	char
9. Correspondence type	50	202-251	char
10. Correspondence topic	50	252-301	char
11. Correspondence subtopic	50	302-351	char
12. Letter date	6	352-357	ymmdd
13. Staffer initials	4	358-361	char
14. Document number	10	362-371	char
15. Comments	100	372-471	char

Figure 2

Records from Senator Nunn's CapitolCorrespond files (without name and street address fields)

City	StateZipcode	Document Type	Topic	Subtopic	Date	Staff	Doc No.	Item Paragraphs
SILVER SPRING	MD 20906	ISSUE	LABOR	STRIKES	940602	MAT	4152296148	ITEMS: 709
LAWTON	OR 73505	MILITARY AFFAIRS	DEFENSE	MANPOWER AND PERSONNEL	940208	CHK	4038293164	ITEMS: 635
LAWTON	OR 73505	MILITARY AFFAIRS	TRADE	PERMANENT RESTRICTION	940208	CHK	4038293164	ITEMS: 635
MARIETTA	GA 30090	ISSUE	GENERIC	BILL STATUS	930708	MKS	4118672081	
ROSWELL	GA 30075	ISSUE	CONGRESS	OPERATIONS	930708	MMS	3188171149	ITEMS: 786
HARTWELL	GA 30643-0283	MILITARY AFFAIRS	DEFENSE	RETIRES	940609	CHK	4159300462	ITEMS: 209
WASHINGTON	DC 31642	ISSUE	JUDICIARY	CRIMINAL JUSTICE SYSTEM	920204	MMS	2034171012	ITEMS: 678
LAKE	WV 25121	MILITARY AFFAIRS	DEFENSE	RETIRES	930702	CCH	3182171028	ITEMS: 0,CONCUR.PRM
CLEVELAND	OH 44135	PRESS	REQUEST	PHOTO	931116	CES	3320092071	
DES PLAINES	IL 60018	ISSUE	LABOR	STRIKES	930810	MAT	3221293329	ITEMS: 709
ACWORTH	GA 30101	ISSUE	SCIENCE & TECHNOLOGY	SPACE PROGRAM	930114	NLU	3013293264	ITEMS: 0,SPACE.PRM
ALBANY	GA 31705	ISSUE	TAX	EXCISE	930623	NLU	3173293026	ITEMS: 154
ATLANTA	GA 30305	ISSUE	HEALTH	HOMEOPATHIC	940301	MAT	4059294392	ITEMS: 1255
NORCROSS	GA 30092	MILITARY AFFAIRS	DEFENSE	MANPOWER & PERSONNEL	940302	CHK	4060293023	ITEMS: 72
SAVANNAH	GA 31404	ISSUE	CIVIL RIGHTS	GAY RIGHTS	920901	MMS	2244293034	ITEMS: 172
CLERMONT	FL 34711	MILITARY AFFAIRS	FOREIGN POLICY	EASTERN EUROPE	930824	CCH	3235053019	ITEMS: 826
ATLANTA	GA 30327	MILITARY AFFAIRS	FOREIGN POLICY	MIDDLE EAST	920923	CCH	2266171113	ITEMS: 0,IRAQ.PRM
STOCKBRIDGE	GA 30281	MILITARY AFFAIRS	DEFENSE	RETIRES	940629	CHK	4193292046	
CAPISTRANO BEACH	CA 92624	ISSUE	GOVERNMENTAL AFFAIRS	GOVERNMENT OPERATIONS	940309	NDW	4102672009	
WASHINGTON	DC 20041	PRESS	THANKS	AWARD	940307	AN	4074672051	
WASHINGTON	DC 20005	ISSUE	LABOR	UNIONS	940407	WMBM	4123672072	NRN
LILBURN	GA 30247	ISSUE	REQUEST	GENERAL	940314	MS/	4181292041	
ATLANTA	GA 30301	INVITATIONS	MEETINGS	DECLINED	930929	MUR	3348672049	BILLY BRAHAM CRUSADE
CHANG SHA HUNAN	GA 30301	MILITARY AFFAIRS	FOREIGN POLICY	CHINA	930413	CCH	3308092001	
WASHINGTON	DC 20002	PRESS	PR	SURVEY	940407	SBB	4103672020	
CORTLANDT MANOR	NY 10566	MILITARY AFFAIRS	DEFENSE	HOMOSEXUALS	940309	CHK	4080672070	
HICKORY	NC 28603	MILITARY AFFAIRS	DEFENSE	PROCUREMENT	940302	CHK	4080672028	
WASHINGTON	DC 20004	ISSUE	FINANCE	ECONOMY	931009	MJW	3345672067	
DECATUR	GA 30031	ISSUE	BUDGET	BUDGET	930810	NLU	3321672063	
HOUSTON	TX 77002	ADMINISTRATION	PR	PERSONAL	940203	MAH	4075672044	
COLUMBUS	GA 31904	REQUEST	REQUEST	WHITE HOUSE TICKETS	940221	JRT	4081672044	
COLUMBUS	GA 31906	MILITARY AFFAIRS	DEFENSE	POW-MIA	940310	CHK	4080672045	
ATLANTA	GA 30339	INVITATIONS	MEETINGS	DECLINED	930915	RMJ	4082672071	9 93
ROSWELL	GA 30076	PRESS	CONGRATS	EAGLE SCOUT	940215	CES	4068672002	
CORDELE	GA 31015	MILITARY AFFAIRS	DEFENSE	HOMOSEXUALS	930309	CCH	3305672005	ITEM 72
LOUISVILLE	KY 40205	MILITARY AFFAIRS	DEFENSE	ALL VOLUNTEER FORCE	940602	CHK	4169672067	NRN
GALAX	VA 24333	MILITARY AFFAIRS	FOREIGN POLICY	KOREA	940410	CHK	4118672018	
MORROW	GA 30260	ISSUE	JUDICIARY	GUN CONTROL	940209	LAL	4195292055	
MORROW	GA 30260	ISSUE	TAX	TAX	940209	LAL	4195292055	

Figure 3

few repositories, like the Richard B. Russell Library at the University of Georgia and Special Collections at Emory University, have worked with information technology experts and political scientists at their institutions to examine some of the data they have received and to explore possible research applications.¹⁹ To date, however, no researcher has studied data from Senate correspondence management systems. In the case of the Nunn papers, use of the correspondence files has been limited to requests by Senator Nunn for information about particular correspondents.

Barriers to Research Revisited

Volume is the most cited barrier to research use of constituent correspondence.²⁰ The adoption of automated correspondence management systems by Congress, other federal agencies, and lobbying organizations made it easier to send mail and contributed to a further increase in the volume of mail handled by Senate offices, making this problem more acute.²¹ The amount of mail generated by Congress

¹⁹ For a summary of the work done at the University of Georgia and Emory University, see Todd Kosmerick, "Congressional Papers Roundtable Minutes, 1998 Annual Meeting, Orlando, September 4, 1998," *Congressional Papers Roundtable Newsletter* [distributed through e-mail, 2 November 1998].

²⁰ See Lucas, "Managing Congressional Papers," 280; Eleanor McKay, "Random Sampling Techniques: a Method of Reducing Large, Homogenous Series in Congressional Papers," *American Archivist* 41 (July 1978): 284; Ham, "Archival Choices," 18; and Patricia Aronsson, "Appraisal of Twentieth-Century Congressional Collections," in *Archival Choices: Managing the Historical Record in an Age of Abundance*, ed. Nancy E. Peace (Lexington, Ma. and Toronto: Lexington Books, 1984), 97. Frank Mackaman, on the other hand, argued that it is the nature of a collection and its arrangement and description, and not its volume, that discourages use. See Mackaman, Archivists Panel in *Proceedings*, 68–9.

²¹ See Paul Chesnut, "Appraising the Papers of State Legislators," *American Archivist* 48 (spring 1985): 165, for a discussion of rising mail volume at the state level.

increased significantly beginning in the mid-1960s. Volume peaked in the late 1980s, averaging 700 million pieces per year from 1984 to 1989. In 1990, Congress responded to pressure to curb the use of franked mail by imposing new restrictions that reduced the volume of mail sent.²² Constituent correspondence, however, continues to constitute up to one-third of the volume of members' papers.

The automated correspondence management systems did end the need for carbon copy cross-reference files. Unfortunately, the topically filed master file has been replaced by correspondence filed by system-generated document number. This number is virtually meaningless to the researcher.²³ In many cases, routine mail (namely, flag requests) and casework are interfiled with issue mail, making it difficult to weed the mail prior to accessioning.

Automated correspondence management systems, however, have allowed Senate staffers to avoid the problems of volume and file order by enabling them to retrieve information from the computer rather than from the correspondence itself. Nunn's staff usually wanted to find letters through personal name or subject and were therefore dependent on the computer system to match the information they had about a constituent or letter with the document number under which it was filed. When they located the on-line entry for the letter, however, they often found that the information they wanted was recorded in the computer file,

²² American Enterprise Institute, *Vital Statistics on Congress 1997-1998*, eds. Norman J. Ornstein, Thomas E. Mann and Michael J. Malbin (Washington, DC: Congressional Quarterly, Inc., 1998), 159. David Burnham, "Congress's Computer Subsidy: Federally Financed Computers, Franking Privileges and Public Funds for Direct-Mail Experts Have Given an Edge to Members of Congress Seeking Re-election," *New York Times Magazine*, 2 November 1980, 97.

²³ Document numbers are generally chronological by date and order of reply.

and that they therefore did not need to retrieve the actual letter.²⁴ (See figure 3.) When the letter was processed, the key information from the constituent letter and the senator's reply was captured in the on-line database. The correspondence itself was filed, more or less accurately, by document number and rarely referred to again.

Indeed, from Nunn's staff's point of view, the correspondence system records were the most important records concerning constituent correspondence. They demonstrated this by requesting that three years of data from the old Senate Mail System (or SMS) be migrated to the new CapitolCorrespond system when they converted in 1994, so that they would continue to have the previous three years' correspondence history on-line. The paper indexes to the correspondence were also available, but the speed of access and the clarity with which the system presented information about the correspondence could not be replicated using the paper records under the current filing system.

Paul Chesnut has argued that "most correspondence sent to state legislators is more useful in the aggregate than in its individual form," and Benson's studies demonstrate that the same is true for congressional collections.²⁵ If researchers are indeed more interested in quantitative studies of constituent mail, the correspondence data files sent to the repositories should encourage their research because much of the data collection has been done for them. Like the Senate staffers, these researchers will be able to bypass working with the actual correspondence. Researchers looking for particular letters or for anecdotes, however, may find these files more

²⁴ Staff members were typically searching for the date on which a constituent had previously written to the senator, which opening paragraph had been used in previous responses, the constituent's address, and the topics on which the constituent had previously written.

²⁵ Chesnut, "Papers of State Legislators," 164. See Benson, "Political Research" and "Letters to Congressmen."

frustrating.²⁶ It is often much easier to find a record for a specific piece of correspondence than to find the correspondence itself. A researcher looking for sample letters on a particular topic, for example, might have to request many boxes or reels of microfilm because the letters are filed according to a system-generated number rather than according to topic. Letters on the same topic often received identical replies, and these letters might be "grouped" or "batched" together when filed. Each group would then be filed under a system-generated number. In Senator Nunn's office, letters that were part of groups were filed separately from other constituent mail, and letters were arranged in no particular order within a given group. Some of the groups contain over ten thousand letters, and locating a particular letter in such a group takes time and luck.

In addition, data entry errors have resulted in numerous entries in the correspondence management systems with misspelled names, topics, and addresses.²⁷ File clerks filing the letters by name or topic might catch the error and file the letter under the correct name or topic. Computer-generated indexes, however, will sort the records as entered, leaving the researcher to scan through the entire index to be sure that the desired record was not accidentally entered with an "I" or a "Z" in front of the last name.²⁸ On the other hand, researchers can use software programs to search for "strings" or groups of characters, letting the computer do the work of scanning the index for the desired term. In addition, the on-line index can be sorted by address or subtopic rather than

²⁶ Patricia Aronsson has pointed out that many researchers appreciate the "anecdotal value" of casework. Aronsson, "Appraisal of Twentieth-Century Congressional Collections," 93.

²⁷ There are several examples of misspelled words in figure 3.

²⁸ For example, in a subset of Senator Nunn's 1990-1991 correspondence management system records, the document type "case" was misspelled in twenty-two different ways, including "CAS3E," "CO," and "DCAS."

name or topic, giving the researcher another way of narrowing the number entries to scan for the desired correspondence.²⁹

Many archivists recommend that constituent mail and casework be sampled, asserting that the volume of mail can be reduced without damaging whatever research value there may be in such files.³⁰ Other archivists warn, however, that sampling may "mislead a researcher by distorting the record of the interaction and priorities of legislative activities."³¹ Accessioning correspondence management systems files will allow repositories to retain a considerable amount of information about the constituent correspondence without retaining all of the actual letters. Researchers will be able to estimate the total volume of mail received and to compare the characteristics of the mail that was retained to the mail that was destroyed.

While researchers may be able to avoid the mountains of paper files by using the information contained in the correspondence management system files, however, the size of the electronic files themselves raise other problems. The size of the files received by a repository will vary, based on the congressman's length of service and his or her policies concerning constituent correspondence. Senator Nunn's file for the older CMS (1978-1994) contained 2,320,000 records and took up almost 1.1 GB. His largest files from the newer CapitolCorrespond system (1994-1996) took up a compara-

²⁹ Statistical software packages such as SPSS (Statistical Package for the Social Sciences) and SAS (Statistical Analysis System) can be used to sort and search large databases.

³⁰ McKay, "Random Sampling Techniques," 281. Aronsson, "Appraisal of Twentieth-Century Congressional Collections," 92-93. Mark Greene, "Appraisal of Congressional Records at the Minnesota Historical Society: a Case Study," *Archival Issues* 19, no. 1 (1994): 35-36.

³¹ Chesnut, "Papers of State Legislators," 166.

tively small 289 MB.³² Using these files requires a considerable commitment of file storage space and software designed to handle large databases.³³ Both repositories and researchers may be discouraged from working with these files because of their size. Repositories planning to offer access to correspondence management system files should break them down into small files that can be more easily accessed.³⁴ Doing so will require the use of servers or mainframes that can retrieve the data from its current storage format and then provide the space needed to manipulate it.³⁵ Researchers can combine these smaller files to make larger data sets if they so desire. The difficulties caused by the size of the files, however, may be short-lived as advances in technology promise more powerful computers that make processing large databases easier in the future.

Reports, indexes, and lists generated by the correspondence management systems serve as useful summaries of the constituent mail files. The reports helped the senator's staff to interpret constituent opinions expressed

³² Senator Nunn directed his staff to answer every letter, postcard, name on a petition, and most phone calls with a letter. He was in office for twenty-four years, the senior senator from Georgia from 1981 to 1996, and chairman of the Armed Services Committee from 1987 to 1994. Senator Nunn had a higher volume of mail answered and indexed than most other senators because of the leadership positions he held and his policies on answering constituent mail.

³³ Faye Phillips discusses these difficulties in *Congressional Papers Management: Collecting, Appraising, Arranging & Describing Documentation of United States Senators, Representatives, Related Individuals and Organizations* (Jefferson, N.C. and London: McFarland & Co., 1996), 178.

³⁴ For example, Emory University will breakdown Senator Nunn's correspondence management system files by year.

³⁵ For a brief summary of Beth Bensman's description of the Russell Library's attempts to work with such large files see Kosmerick, "Congressional Papers Roundtable Minutes," [distributed through e-mail 2 November 1998].

in the mail received by their office. When the Senate used the centrally controlled CMS, reports and indexes were generated automatically, and staffers had to make a special request to have a duplicate copy run later if the first report was mislaid. The reports and indexes that were important to the office therefore were filed fairly carefully. The systems implemented after 1993, however, resided on local area networks within Senate offices, and the staff maintained the system files directly. They generated reports as needed and may not have kept them as another could be generated on demand.³⁶ Accessioning the correspondence management system files would allow researchers to generate their own reports and to recreate reports the office may have lost or decided not to generate themselves. Researchers using reports and indexes generated by the systems, however, must be cautioned that the data on which the reports are based contains many irregularities. The reports and indexes do represent the information on the mail available to the senators and their staff, but this information may not reflect accurately the amount or content of the mail itself. Depending on a researcher's interests, what the senator knew about the mail he or she received may be more important than the actual content of the mail.

Those repositories choosing not to provide researchers access to the correspondence management system files should work with the senators' staff members before they leave office to determine which information was important to them and to make sure that reports have been generated to capture that information. For example, these could be reports listing mail volume per month or per year, lists of the most popular

³⁶ For example, the CMS automatically generates a weekly "hot topic" report listing the most frequently used item paragraphs. Senator Nunn's office maintained a file of these reports. The CapitolCorrespond system that they adopted in 1994 did not automatically generate this report, and the office staff only produced it sporadically.

topics per month, or indexes to correspondence on issues important to the senators. The repository might also want to contact potential researchers to determine what kinds of information they might be interested in seeing. Researchers using any of the reports generated or retained should be shown printouts of data from the system so that they can see the kinds of irregularities that exist in the data from which the reports are drawn.³⁷

The organization of the files forms a second barrier to research. When Benson took a representative sample from the paper-based Minnesota constituent mail, he discovered three problems. First, though the congressmen all seemed to have some rough, topical organization for their mail, their systems were different enough to make uniform sampling difficult.³⁸ Second, the topic categories used were too general to be useful for researchers. For example, a researcher looking for letters on open housing legislation would have to oversample the folders on civil rights in order to get a sufficient number of letters for her study.³⁹ Third, many constituents covered several topics in their letters. The letter most likely would be filed under only one of them. Benson's sample, therefore, would not be drawn from the total number of letters on that topic as some of those letters would be filed elsewhere under another topic discussed in the letter.⁴⁰ In addition, for quantitative analysis the topics

³⁷ The name and topic indexes generated by the CMS provide this kind of information. The systems implemented after 1994 may not automatically generate such indexes, and in such cases the repository should request that an index to a small portion of the correspondence be generated.

³⁸ Benson, "Political Research," 9, 10-11.

³⁹ *Ibid.*, 8.

⁴⁰ Historian Richard Lowitt also found this to be true in his research using Senator George W. Norris's papers. His research, however, was not quantitative in nature, and he felt that he found important information by browsing through the correspondence and reading documents not directly

covered by the constituent's letter must be put in rigorously defined categories. Given the wandering, unfocused nature of much of the correspondence, such categorization was time-consuming.⁴¹

Automation provides some solutions to these problems and presents other problems in a slightly different guise. The Senate Computer Center standardized the correspondence management computer files sent to repositories. The format changed slightly in 1996 after Senate Archivist Karen Paul solicited input from the repositories; however, in general, the same information has been transferred to the repositories over the years. (See figure 2, page 46.) Thus comparisons between the mail received by different senators should be possible. A uniform format, however, can mask differences in the way that the staff used the system. System documentation indicates only what the system was designed to do. It does not document the ways in which a senator's staffers worked within the system to record things not anticipated by the system designers. For example, Senator Nunn's Atlanta office overrode the system-assigned document number so that all mail related to a particular case would have the same document number. Lydia Lucas expressed a concern in 1978 that adopting standardized filing systems and means of "computerizing" congressional records would "submerge the individuality of the senator."⁴² The danger, however, seems to be not that individual senators will do things differently but that archivists and researchers will not recognize what they have done differently.⁴³

related to the topic that he was researching. Richard Lowitt, remarks during Historians Panel in *Proceedings of the Conference on the Use and Disposition of Senators' Papers*, 47.

⁴¹ Benson, "Political Research," 11.

⁴² Lucas, remarks during Archivists Panel in *Proceedings of the Conference on the Research Use and Disposition of Senator's Papers*, 73.

⁴³ For an expanded discussion of the need for archivists to work closely with congressional offices to document electronic records, see Phillips, *Congressional Papers Management*, 177-80.

Many of the difficulties that Benson had with topic categories were merely transferred to the new systems. Although the correspondence management systems did allow staffers to assign multiple topics and subtopics to correspondence records, there was no control on the terms entered. Topics remained broad and continued to reflect the interests and needs of the individual offices, making comparisons between different offices difficult. Perhaps more significantly, topics could be added at will or accidentally misspelled. Misspellings and unauthorized terms make it difficult to retrieve comprehensive listings of correspondence on a specific topic.

Automation does provide two possible solutions to these problems, however. First, a list of all topics can be generated and any misspellings or unauthorized terms corrected in a copy of the file. Second, researchers can take advantage of the information used to generate the reply letter to locate more accurately letters of interest and to categorize individual letters. In order to create a reply, the correspondence management system needed the codes for the item paragraphs that would make up the reply letter. These codes are listed in a field in the file sent to the repositories. An index for the item paragraphs can be generated from the correspondence management system (see figure 4, page 58), and the text of the approved paragraphs and their codes can be found in the library of approved items, often located in the Systems Administrator files. Many paragraphs were written to respond to particular kinds of letters; for example, supporting the Gulf War, opposing a milk tax, or opposing daylight savings time. The item paragraph codes were used to generate a report listing the most frequently cited constituent concerns or positions. Researchers can use these codes to design the rigorous content categories needed for

Index to Item Paragraphs (generated by CMS)

OFFICE: SENATOR SMYTHE		ABSTRACT REPORT		R02
Date of Listing: SEP-18-81				
Topic	Subtopic	Item #	Synopsis	Updated
Close		4	disagree on issue	JAN-07-80
Close		5	Thank you again: keep in touch	JAN-07-80
Close		6	Thank you again	JAN-07-80
Close		7	good to hear from you: keep informed of activities	JAN-07-80
Close		9	call on me, look forward to hearing from you	JAN-07-80
Intro		1	f. i. subj of ltr: Thanks for letter, happy to know views	JAN-07-80
Intro		2	f. i. date of ltr: Thanks for bring matter to attention	JAN-07-80
Intro		3	f. i. date of ltr: Thanks for ltr: Thanks for letter	JAN-07-80
Intro		8	f. i. publication name: thanks for publication	JAN-07-80
agriculture	dairy	11	dried milk to compete w/ fresh, anti	SEP-05-81
energy	public utilities	13	f. i. county - adverse impact elec. serv. in counties	SEP-17-80
environment	solid waste	16	aprec. learning concern about solid waste management	JAN-07-80
environment	solid waste	17	newspaper disposal	JAN-11-81
housing	rent control	22	support local rent control. I agree completely	JAN-11-81
interior	national parks	20	Green Valley strip mining	JAN-07-80
interior	national parks	47	ban strip mining and protect wildlife sanctuaries	SEP-17-80
public activities	congrats from senator	18	short - naturalization	SEP-16-80
public activities	congrats from senator	19	short, f. i. occasion - for example, 75th birthday	SEP-16-80
taxes	marriage penalty	17	encl - single and married tax differences	SEP-17-81
wildlife	animal abuse	14	soring of horses	JAN-11-81
wildlife	animal abuse	44	inhumane treatment of animals	MAR-14-81

Figure 4

quantitative analysis. In effect, the senator's staff has already coded each letter for content.⁴⁴

One of Benson's goals in sampling the Minnesota issue mail was to estimate the total quantity of mail received on a specific topic. The correspondence management files should allow a researcher to determine more easily the quantity of mail received on a given topic without having to sample. Researchers, however, will have to take time to examine carefully the data file for irregularities and will need to consult memos and other records concerning correspondence files to determine whether there has been any duplication or data loss. For example, some correspondence management systems allowed staffers to make copies of entries and then assign them different topics/subtopics. When Senator Nunn's office changed its subtopic for Desert Storm from "Middle East" to "Iraq-Kuwait," for example, the staff created a duplicate entry for all records related to the war and entered under "Middle East" to the new topic "Iraq-Kuwait." These records, therefore, appeared twice in the database. In another case, shortly after the change to the CapitolCorrespond system, several hundred new records were deleted when data entry operators accidentally pressed the wrong key. Information about these kinds of data irregularities can only be obtained from the staff members who worked with the correspondence management systems.

⁴⁴ Some letters, of course, were not answered using the pre-approved item paragraphs. In Nunn's office, these letters answered with customized text were known as "perms." In the correspondence management system file, instead of listing the item paragraph code, the staffer would enter the file name for the newly created language (that is, SPACE.PRM or IRAQ.PRM). "Perms" that were used to answer several letters were made into item paragraphs and assigned an item code. Letters that were not created using the correspondence management system, but were indexed in the system, were known as "handtypes" and might not have any item codes associated with them. Letters indexed but not answered were known as "no reply necessary" or "NRN" letters. Sometimes "NRN" was entered in the item code field. See figure 3 for examples.

The greatest amount of irregularity usually occurs during transitions from one system to another.⁴⁵

Benson also pointed to the need for research into who writes to their congressmen. The greatest problem he identified in this area, aside from the volume of the mail, was that vital information was frequently not present in the letters themselves, including age, race, and occupation.⁴⁶ This information is also unlikely to appear in the computer database. Benson suggested that researchers might be able to find additional information about constituents in local directories,⁴⁷ and the ability to create reports listing constituents by name or by address might make such work easier. Files that have the title data (Mr., Mrs., Dr., Ms., et cetera) separated into a separate field may allow researchers to categorize constituents further by sex. Data entry errors will make any study of constituents difficult, however. Senator Nunn's data files contain numerous examples of misspelled first and last names and incorrect zip codes and state designations in the address fields.

Both the 1978 Conference on the Research Use and Disposition of Senators' Papers and the 1986 *Congressional Papers Project Report* written after the conference on congressional papers sponsored by the Dirksen Congressional Center and the National Historical Publications and Records Commission emphasized that donor restrictions pose perhaps the greatest barrier to research use of congressional

⁴⁵ Many archivists have recommended that repositories work closely with congressional staff members to ensure that the transfer of records is complete and orderly and to allow the archivist to become familiar with the way that the office functioned. See Paul, *Records Management Handbook*, 129; Connell Gallagher, "A Repository Archivist on Capitol Hill," *The Midwestern Archivist* XVI, no. 1 (1991): 49-58; and Faye Phillips, "Harper's Ferry Revisited: The Role of Congressional Staff Archivists in Implementing the Congressional Papers Project Report," *Provenance* VI (spring 1988): 26-44.

⁴⁶ Benson, "Political Research," 11.

⁴⁷ *Ibid.*, 14.

collections.⁴⁸ Constituent correspondence is generally given long restriction periods to respect constituent privacy. It is a simple matter, however, to create a copy of the constituent management system files without the name and street address fields (see figure 3). The resulting file protects individual constituent confidentiality while allowing researchers access to aggregate data about the correspondence. Repositories that plan to offer access to the correspondence management system data files should try to open these files to researchers as early as possible. Computer files that are open and used are much more likely to be refreshed and migrated to new storage formats and are therefore more likely to be preserved in a usable format.

Conclusion

Automated constituent correspondence system records are well suited for aggregate, quantitative research. The correspondence management system records provided in electronic form by the Senate Computer Center are an important access tool, a source of significant information, and the only index to senatorial constituent correspondence. They can be used as a finding aid for the correspondence records and to sample or weed those files. Unlike the correspondence itself, they can be purged of confidential information easily and, therefore, more quickly opened for research. Perhaps most significantly, the Senate staffers have already coded demographic and topical information into the computer files, providing a database that can be adapted readily for use with statistical database software.

Correspondence management system records, however, promise more than they can deliver. Misspellings, missing data, missing records, and duplicate records combine to undermine the reliability of the data files as both indices and data sets. The repository must be familiar with how the

⁴⁸ Phillips, "Harper's Ferry Revisited," 34.

senator's staff used the system in order to help the researcher correctly interpret the files. The size of the computer files themselves make them difficult to manipulate and search, and using the data may require skills that most archivists currently do not possess. The correspondence management files that serve as an index are separate from the correspondence, and the correspondence is extremely difficult to access without that index. The key to the item paragraph codes that provide more precise subject access are also in a separate file.

Repositories planning to provide access to correspondence management files must commit time and resources to working with the Senate staff to document the systems and how they were used, to reformatting the data into smaller files, and to migrating and refreshing the data to keep it accessible as technology changes. These are significant commitments considering the problems posed by the data and the lack of interest researchers have shown in constituent correspondence, in general. Unfortunately, although correspondence management systems provide some advantages to users interested in data manipulation and quantitative analysis, data contained in them is, as archivist Margery Sly feared, "an unholy mess."

Naomi L. Nelson is the Modern Political Collections Archivist in the Special Collections Department, Robert W. Woodruff Library, Emory University. This article combines several papers given at the annual meetings of the Society of American Archivists and the Congressional Papers Roundtable between 1994 and 1998.