

SLAC-PUB-10135
August 2003
Revised
September 2004

Whither Processing? An Incremental Approach to Archival Processing

Jean Marie Deken

*Stanford Linear Accelerator Center
Stanford University
2575 Sand Hill Road MS 82
Menlo Park CA 94022 USA*

Invited talk presented at the Society of American Archivists
Annual Meeting, 8/18/2003—8/24/2003, Los Angeles, CA, USA

This work was supported by the US Department of Energy Contract No. DE-AC02-76F00515, and by a grant from the American Institute of Physics.

The struggle to accomplish processing is shared, to a greater or lesser extent, by most archival repositories regardless of their size or subject matter. Many archives also, I suspect, employ similar coping strategies in their efforts to preserve the identity and integrity of the records in their care, while at the same time arranging and describing them in order to facilitate research. Making progress on processing is one of the most difficult tasks we face in the Archives and History Office of the Stanford Linear Accelerator Center, and it is my hope that a discussion of our efforts in this area will prove useful to others both for the general issues it can illuminate and for the specific strategies it may suggest.

SLAC's Archives and History Office

The Stanford Linear Accelerator Center is owned by the United States government, and is operated for the US Department of Energy by Stanford University. Its present-day mission is to design, construct and operate state-of-the-art electron accelerators and related experimental facilities for use in high-energy physics, particle astrophysics, and synchrotron radiation research. The idea for the construction of a two-mile linear accelerator at Stanford University was conceived in 1956 and authorized by the United States (US) Congress in 1961. Construction of SLAC began in July of 1962 and was completed February 10, 1966.

The SLAC Archives and History Office began its life in February 1986 as the “SLAC History Project,” with a records survey in all administrative groups. Identification of important records was followed by creation and population of an inventory database for 500 separate records collections, and by the establishment of a physical archive of important records no longer needed for current business. The records survey was then followed up with an oral history program to gather information not fully documented in the available records.¹ SLAC’s History Project officially became the “SLAC Archives and History Office” (AHO) in 1989. The period 1993 to 1995 saw some growth in the program, but progress was hampered by loss of staff. After a year and a half hiatus, the Archives and History Office began a second growth spurt in mid 1996, when I was hired as permanent full-time archivist and, later that same year I hired a permanent halftime archives assistant, Laura O’Hara. We have a dedicated 2400-cubic-foot capacity temperature and humidity controlled archival storage area (completed in 1996), and an Archives Program Review Committee (established in 1999) comprised of internal and external stakeholders. The Review Committee advises SLAC management on the goals, policies, and activities of the AHO program.² By the end of calendar year 2002, the AHO had collected and at least partially processed over 1600 cubic feet of SLAC

¹ Detailed history of the SLAC Archives program can be found in: Deken, Jean Marie. *Future Proof for Physics: Preserving the Record of SLAC*. SLAC-PUB-9686. April 2003.

² The program review committee meets biennially, and its reports are available online at <http://www.slac.stanford.edu/history/progrev/charge.html>.

historical records, and had accumulated a processing backlog of roughly equivalent proportions.

General Issues

This very brief sketch of the history of SLAC and its archival program's beginnings hints at one of the major problems we have encountered with processing, which is, that even at the beginning of our existence, we were already behind on our work. The organization had a thirty-year head start on us, so the Archives and History Office was born with a significant processing backlog. (It should be noted here that, because this is a mid- to late-twentieth century backlog, it is primarily in analog formats with paper and photographic prints and film predominating.)

Another major problem has been that the work of processing has tended to be completely invisible to the management of SLAC, or, when visible, has tended to be regarded as trivial. At times the prevailing opinion has seemed to be that the real work of the archives is accumulating the records and providing reference on them. In this view, once records are in custody they are, *de facto*, "reference-ready," requiring either no effort at all to arrange and describe them, or so little effort that it is hardly worth mentioning.

A third issue has been another prevailing opinion that can be summarized as "everything is electronic now, and that has solved all of the Archives problems." This opinion appears to have been widely held in the organization and was often expressed to me when I first came to SLAC in 1996. However, the intervening 7 years have seen the bursting of the Internet bubble, as well as some maturing of the no longer entirely novel digital age. These developments have been accompanied by the growing realization that all formats – even digital ones – have their strengths and weaknesses, and that every record, regardless of format, has to be stored in such a place and such a manner as to be preserved and retrievable for as long as necessary. The many major and minor repercussions of a 1997-1998 SLAC-wide platform migration from VM to UNIX³ have also played their part in the recent local decline of this particular point of view.

To deal with our large backlog as expeditiously as possible in a climate characterized by the issues I have just described, we have developed what I like to call an "incremental approach" to processing that is founded on two methods: "triage processing" and the "virtual sort."

Strategies of the Incremental Approach: Triage Processing

The triage, or ranked, approach to processing acknowledges that not all records in the backlog are of equal importance. Triage helps us to spend most of our processing time on the most important records.

³ VM Migration 9 April 1998. At: <http://www.slac.stanford.edu/comp/vm/vmmigr.html> (8/3/2003)

The first level of triage, or basic processing, is applied to all records. At this stage records are accessioned, and a skeletal entry is created on the archives' collections database, SLACARC. The skeletal entry consists of an accession number, a "short-title," a one-to-two sentence description, and an indication of whether the accession came with a folder list provided by the creating office, or needs to have a folder list generated by the archives. The "short title" of the accession follows the format: Division, Office, and Personal Name. Very basic holdings maintenance is also completed for this level of processing. Any 3-ring binders encountered are removed,⁴ highly acid folders are replaced and obviously non-archival materials are discarded. An accession that has been processed to the first level goes on a shelf in storage, and it may remain at the first processing level indefinitely. All fields of the SLACARC database entry for a first-level accession can be searched; so limited reference service can be immediately provided for these accessions.

The second level of triage, or secondary processing, is applied to records pulled in response to reference requests. As the accession is handled and reviewed in the course of researching and responding to a reference request, the database entry for it is opened and improved. Information added to an entry during secondary processing can include types of material contained in the accession, names of individuals who created or are the subject of the records, descriptions of projects or experiments documented, and key dates. At first blush this approach might seem to significantly lengthen our reference response time, but we have found that the time spent describing an accession as it is being reviewed for reference reasons—even when the information being collected and entered on the database has little or no bearing on the immediate reference request—ultimately saves time. This is because approximately 80% of our reference requests seem to involve only around 20% of our holdings, and, over time, the better we process the most-used 20% of our records, the more easily we are able to assist 80% of our reference clientele.

We also complete the secondary level of processing on collections for which I anticipate high reference interest in the near future. As significant anniversaries loom on the horizon – the 25th anniversary of colliding beams at SPEAR in 1997, the 10th anniversary of the SLAC World Wide Web site in 2001, and the 40th anniversary of the entire lab in 2002, for example – I have identified accessions in our holdings that are likely to document the history of these activities or facilities. We then attempt to complete secondary processing on the identified accessions so that, by the time interest generated by the approaching anniversary is beginning to rise, we have the secondary processing completed and are well-prepared to handle the increased incoming requests concerning those topics.

Our third level of triage, or high-level processing, is what is traditionally thought of as "archival processing." At this level, records are arranged and described in great detail and a guide to the collection containing the customary elements—a biographical sketch of the records creator, a scope and content note, series descriptions, and a container list—is prepared. In my seven years at SLAC, we have processed only two collections to this

⁴ Three-ring binders use 1/3 more space than documents removed from binders, and most binders are too tall to fit properly into the Archives' storage containers.

level: a three cubic foot collection of documents relating to our web site⁵, and the 141 cubic foot first installment of the papers of our second Director, Burton Richter. The processing of both of these collections required extraordinary staffing and funding. We were able to secure SLAC Research Division funding for the high-level processing of our three cubic foot web site collection in conjunction with the 10th anniversary of our web site, which was the first site established outside of Europe.⁶ And, in 2000, we applied for and were awarded a processing grant of \$10,000 from the American Institute of Physics for work on the papers of Burton Richter, SLAC Director and Nobel Laureate.

Strategies of the Incremental Approach: The Virtual Sort

Dr. Burton Richter began his association with Stanford University in 1956, as a Research Associate in the High Energy Physics Laboratory on campus. Throughout his long—and still continuing!—career at Stanford, he has held the positions of Research Associate, Assistant Professor, Associate Professor, Professor, Technical Director, and Director. Along the way he has been awarded the Nobel Prize in Physics and the U.S. Department of Energy's E.O. Lawrence Award (both in 1976), has been elected to the National Academy of Sciences, a Fellow and later President of the American Physical Society, and President of the International Union of Pure and Applied Physics. In addition, he has served on the Stanford University Faculty Senate, and on various University committees. In the succession of positions that Dr. Richter has held on campus and at SLAC, he has had the support of various administrative personnel—numbering at times from less than one to two or more full-time staff—whose major interests and tasks it can be safely said almost never included maintaining the files.

From its very beginnings in 1986, the SLAC History Project, and later Archives and History Office, has provided records management services to the Directorate of SLAC, including Dr. Richter, the Laboratory Director, and the Associate Directors for SLAC's five Divisions. In the case of the Richter papers, this has meant that they have been retired to the archives when they were no longer needed for current business and as file storage capacity in the Director's office dictated. Accessioned in 68 separate increments over a period of 11 years, they document the entire range of his various activities and affiliations for the past 43 years. The combination of the low priority given to the systematic organization of Dr. Richter's files while they were active, their handling over long periods of time by many different office staff in various parts of the organization, and their retirement to the archives in what to us appeared to be random batches, has made processing the collection particularly challenging.

At the time we initiated the Richter Papers Processing Project in 1999, all of his accessions in the archives had been processed to the second triage level. This meant we

⁵ SLAC Accession 00-072

⁶ Deken, Jean Marie. *First in the Web, but Where are the Pieces?* SLAC-PUB-7636, April 1998.

See also: The Once and Future Web Symposium A symposium to be held on the occasion of the 10th anniversary of the first U.S. Web site at the Stanford Linear Accelerator Center December 3-4, 2001. (<http://www-conf.slac.stanford.edu/webanniv/> 8/12/03)

had a folder list for each accession, and, because of the cooperation of his office staff, many of the folder lists were available to us as electronic documents. Knowing that we had to maintain quality reference service on the papers even as they were being processed, I cast about for a method to work with them that would take as little space as possible, would involve physically relocating the records only once or twice, would maintain them in searchable (and find-able) order throughout the length of the project, and could be undertaken with the assistance of a rotating series of assistants of various skill levels.

The technique developed to accomplish these goals is one I have dubbed the “virtual sort.” We began by relocating all 68 Richter accessions to contiguous storage space, and making note of the new locations. I then had our Project archives assistant Bill O’Hanlon (hired with the AIP grant money) copy all of the individual Richter accession inventories onto one spreadsheet database with separate fields for the current accession’s date range, accession number, box number, folder number and file folder label. He saved a copy of this spreadsheet as a backup, and then added to the working copy fields for the new series number, new series name, new series folder number and folder date range that would be assigned to each existing folder during the virtual sort (Figure 1). He then printed out the entire database and arranged the pages in strips (like hanging wallpaper) along one whole range of shelving at the end of our stack area. This allowed those of us working on the project to review the list in its entirety, and to begin to formulate a plan for organizing the records into series and sub-series (Figure 2). Initially I had hoped that the series lists developed during an early 1990’s project to process the papers of SLAC’s first Director, W. K. H. Panofsky would be a useful guide for us, but Richter and Panofsky’s career trajectories have been so different that the Panofsky guide was ultimately of little help. After I had developed a tentative series list, Bill applied that list to the database, and sorted all of the files—moving the entries on the database, and not the actual physical folders. He then printed the second sort according to the proposed series designations, and we analyzed whether the sort was sufficiently logical and complete. Of course, being a first effort, it was less than adequate, so we tweaked the series designations and tried again.

After several iterations of the virtual sort, we were quite satisfied that some of the series designations were correct, although we had specific questions or issues with others. At this point Bill was set to work converting the virtual sort into a physical sort for those series we had finalized, and Laura O’Hara and I continued our analysis of the other series. As we completed each series to our satisfaction, Bill printed out the lists in turn, pulled the folders from their original locations, and arranged them in new boxes in the new series order. Laura checked the pulls for accuracy and logical sequencing, and I provided a final review of the completed series. Questions and problems were flagged and resolved in consultation. In this fashion we worked our way through the entire collection over the course of 24 months. As we conducted the virtual sort, we did encounter evidence that at several times short-lived efforts had been made to systematically organize parts of Dr. Richter’s papers. We used the forensic evidence of these various systems to assist us in the sort.

On a project of this size and complexity, it is easy to lose track of progress, and for this reason I kept a very detailed log of what had been accomplished, what remained to be done, and who was assigned to what tasks. I posted the log in our work area, where we all consulted it frequently. In the course of the project the information logged was refined and updated. The first version of the log (Figure 3) was fairly simple—perhaps demonstrating our innocent confidence in the relative manageability of the task at hand. By the end of the project, however, the log had grown in size and detail (Figure 4). Although somewhat intimidating when taken as a whole, the log helped us effectively segment this huge project into manageable increments, and then to track those increments from initiation to completion. As the months passed, the log helped us to move forward, and served as a visual reminder that progress was indeed being made.

Lessons of the Incremental Approach

Although the official project ended almost exactly one year ago, you may notice that our project log is not entirely complete. This is one of the realities of life in a small repository: projects are picked up and put down based on many factors, some internal and some external. I hope we will get back to the Richter project soon—and when we do, because of the log, I will know exactly what work remains to be done, and exactly which task should be next undertaken.

The AIP grant we received covered 6.5% of the costs of the Richter Papers Project, so far, but it was the most critical portion of our funding, because the grant was a catalyst for acquiring SLAC support and funding for the entire project. Outside recognition of the importance of the Richter papers, and outside support of and validation for their processing have been critical factors in raising lab support for the project and support for archival processing in general as a necessary and worthwhile activity. The advocacy of the Archives Program Review Committee has helped enormously as well. The Committee lends authority to the Archives goals, and its reports educate the laboratory community about our activities.

Planning and documentation have been critical to the success of our incremental approach to processing. Detailed planning has allowed us to segment processing into discrete tasks, for which we have subsequently requested funds and staff. Smaller projects are often easier to sell to management, and smaller budget increases—particularly if they are for one-time projects with well-stated goals and well-defined limits—are easier for our managers to approve. Detailed documentation has been created for every step of our incremental processing. We have needed to be very clear on procedures, and have taken pains to think through our methods and routines thoroughly, because the incremental tasks have been picked up and put down in mid-stride many times by many people. The down-side to this approach is that it requires repeated training of new staff: the up-sides are that throughout the life of a project we benefit from the vision of fresh eyes looking at familiar problems, and the documentation also assists the continuing staff in picking up projects and moving them forward after a hiatus.

We could not have contemplated our incremental approach to processing without the use of computers. Computer technology has been a key factor in the success of triage processing and of virtual sorting, and we anticipate adapting the “virtual sort” for use with strictly, or mostly electronic records collections. Technology has introduced multiple duplications into the virtual and physical environments: every desktop computer has a separate hard-drive for storage, many users have dedicated storage space on mainframe systems, and almost every desktop computer is connected to a printer. With all of the varied records creators, records formats and records storage locations, accompanied by an increasingly flattened administrative structure that has resulted in the handling of record-keeping by records creators themselves, we anticipate receiving and processing additional collections that will have come to us in a condition similar to—or even more fragmented than—that of the collection of papers I have just described. It is our hope that in the not-too-distant future, the virtual sort approach can be effectively utilized for the processing of these types of collections as well.

Acknowledgements

I wish to thank Kristine Kaske, Smithsonian National Air and Space Museum, for chairing the session, “Approaches to Processing: Finding Time and Funding,” at the Society of American Archivists 2003 Annual Meeting. Thanks are also extended to Peter Harper, National Cataloging Unit for the Archives of Contemporary Science (UK) and R. Joseph Anderson, American Institute of Physics, for agreeing to present papers at that session. Bill O’Hanlon, Laura O’Hara, Pennington Ahlstrand, Kelly Chatwell, Brian Johnson, and Irene Sanchez have been, at various times, the Richter Papers Project staff, and I thank them for their creativity, diligence, and persistence. This work was supported by the US Department of Energy Contract No. DE-AC03-76F00515, and by a grant from the American Institute of Physics.

Virtual Sort								I
A	B	C	D	E	F	G	H	
1 Series 2: Stanford University 1960-Present								
Series	Series Box #	Series Folder	Folder Date Range	Acc Date Range	Acc #	Acc Box	Acc Folder	Folder Title
2	2			1960-1975	92-064	4		1 Ombudsman - H & S 1972
4	2			1960-1975	92-064	4		2 Subcommittee Research and Instruction 1975
5	2			1960-1975	92-064	4		3 Subcommittee Research and Instruction 1975
6	2			1960-1975	92-064	4		4 RDNA Earthquake Stds
7	2			1960-1975	92-064	4		5 Recomb DNA Panel - Minutes Panel on Recombinant DNA Research
8	2			1960-1975	92-064	4		6 Richter RDNA
9	2			1960-1975	92-064	4		7 DNA
10	2			1960-1975	92-064	4		8 Fac. Disip.
11	2			1960-1975	92-064	4		9 Academic Council 1970-71
12	2			1960-1975	92-064	4		10 Academic Council 1970-71
13	2			1960-1975	92-064	4		11 1974-75 Senate (cont)
14	2			1960-1975	92-064	4		12 Senate Documents, Reports, Misc., Memos 1971-72
15	2			1960-1975	92-064	4		13 Senate 1974-75 VII
16	2			1960-1975	92-064	4		14 Vth Senate 1973-74
17	2			1960-1975	92-064	4		15 Vth Senate 1972-73
18	2			1960-1975	92-064	4		16 CEC Subcom. on Res. and Instruction
19	2			1960-1975	92-064	4		17 Academic Council 1966-67
20	2			1960-1975	92-064	4		18 Academic Council 1967-68
21	2			1960-1975	92-064	4		19 Academic Council Before 1966-67
22	2			1960-1975	92-064	4		20 Academic Council 1968-69

Figure 1: Richter Project Virtual Sort Spreadsheet



Figure 2: Richter Project Sorting “Wallpaper”

JULY 10

SLAC ARCHIVES AND HISTORY OFFICE Richter Papers Project Status Log							
Series	Status	Virtual Sort		Sort Review		List & Pull	
		To Series Level	In Series Order	Review Virtual Sort	Assign Sub-Series	Physical Sort by Subseries	Review
		Bill O'Hanlon	BO	Laura O'Hara	Laura O'Hara	LO & or BO	Jean Deken
1A	completed	completed	✓	completed	completed	completed	completed
2	completed	completed	✓	completed	✓	✓	ready
3	completed	in process					
4	3	completed	✓	ready			
5	4	completed	✓	ready			
6	5	completed	✓	ready			
7	6	completed	✓	✓ READY	✓	✓	ready
8	7	completed	✓	✓ READY	✓	✓	✓ ready
9	8	completed	✓	✓ READY	✓	✓	
10	9	completed	✓	READY			
11	10	completed	✓	✓ READY	✓	✓	ready
12	11	completed	completed	completed	completed	✓ READY	✓ ready

Figure 3: Richter Project Work Log
(Version1)

SLAC ARCHIVES AND HISTORY OFFICE										
RICHTER PAPERS PROJECT										
STATUS LOG										
SERIES	STATUS	PRINT LIST	SORT REVIEW	REVIEW	LIST & PULL	MAKE	REFOLDER,	COMMENTS	QUALITY INV. LIST	FOLDER BOXES
#	#	VIRTUAL	& PULL	assign	physical	REBOX & SORT	REBOX & SORT	CHECK	UPDATES	SHELF
New	Old	SORT	IN SERIES	review	sort by	flag	add in or delete	SHLF	LIST	UPDATE
		to Series	virtual	sub-series	subseries	flaggedolders	SUBSERIES	UPDATES	SHLF	PRINT
		level	ORDER	sort	subseries	corrections,		COMP.	UPDATES	FINDING
						flaggedolders			BOXES	AID
						sortfrom	sort lists			
							all staff			
							BO			
								JD, LO	JD, LO	STAFF
								JD, LO	JD, LO	STAFF
								JD	JD	JD
1A	1A1	Uni. Gov.	✓	✓	✓	✓	✓	✓	✓	✓
1B	1A2	Acad. Council	✓	✓	✓	✓	✓	✓	✓	✓
1C	1A3	Stanford Phys.	✓	✓	✓	✓	✓	✓	✓	✓
2A1	1B1.1	Dir. Offc.	✓	✓	✓	✓	✓	✓	✓	✓
2A2	1B1.2	DOE	✓	✓	✓	✓	✓	✓	✓	✓
2A3	1B1.3	Education	✓	✓	✓	✓	✓	✓	✓	✓
2A4	1B1.4	SLAC Faculty	✓	✓	✓	✓	✓	✓	✓	✓
2A5	1B1.5	Legal Office	✓	✓	✓	✓	✓	✓	✓	✓
2A6	1B1.6	Personnel	✓	✓	✓	✓	✓	✓	✓	✓
2B1	1B2.1	Office Clerks	✓	✓	✓	✓	✓	✓	✓	✓
2B2	1B2.2	Comp. Pol. Ctr	✓	✓	✓	✓	✓	✓	✓	✓
2B3	1B2.3	EPAC	✓	✓	✓	✓	✓	✓	✓	✓
2B4	1B2.4	etc.	✓	✓	✓	✓	✓	✓	✓	✓

Figure 4: Richter Papers Project Log (Final Version)