

SLAC-PUB-13214
April 2008

**HYBRID PAPER/ELECTRONIC ARCHIVAL COLLECTING, PROCESSING,
AND REFERENCE:
A View from SLAC**

Jean Marie Deken

*Stanford Linear Accelerator Center
Stanford University
2575 Sand Hill Road MS82
Menlo Park CA 94025*

Invited talk presented at
Future Proof IV
4/23/2008-4/25/2008,
Stockholm, Sweden

Abstract

Real-time archiving of mixed paper and digital collections presents challenges not encountered in the primarily paper environment. A few recent examples from the archives of the Stanford Linear Accelerator Center highlight obstacles encountered, and attempted and contemplated solutions.

Work supported by Department of Energy contract DE-AC02-76SF00515

Introduction

Lucinda Glenn, the current President of the Society of California Archivists, recently exhorted a group of student archivists with the following analogy:

Science Fiction is replete with issues concerning the space-time continuum. Space ships are caught in an anomaly and flung around the sun. The characters find themselves in the past trying to right a wrong by finding the information needed to save the galaxy from sure annihilation. That's pretty much our profession.

In our own quiet, persistent ways we fling ourselves into the past, and if not directly right wrongs, we provide the files, the photographs, the correspondence, the videos, the reel-to-reel tapes, the handwritten or mimeographed or typed papers, the obsolete computer discs by which wrongs may be righted. We go boldly where no one has gone before – the school basement, the shed in the back yard, the church steeple, the garage, the back room off the kitchen, the storage room in the boiler room under the pipes. And we fight through spider webs, dogs with cold noses, and whatever that is in the dark corner skittering around making noises. Our identity is seldom known and we often wear the cloak of invisibility, but that cloak is fastened by a badge of nobility...

Maybe we don't save the galaxy, but we do save the community, the organization, the school, the city, the county, or the country from the annihilation of memory and identity.¹

I find Ms. Glenn's analogy both charming and inspiring, and I must confess to you that lately I am an archivist in need of inspiration. Perhaps I have spent too much time in basements and attics, but it seems to me that my "badge of nobility" is getting a bit frayed around the edges. What is fraying it most are newer hazards to the archival mission of saving memory and identity, hazards that seem to threaten to prevent us from being able to "fling ourselves into the past" because our resources are diminishing, and because important parts of our past are fast becoming what seems to be a disappearing digital mist that we can locate only with difficulty, collect with extreme effort, and preserve only momentarily.

In my day-to-day attempts to navigate these new hazards, I have developed some hybrid procedures, kept an eye on emerging regional and international trends, and come to some tentative conclusions about best practices for my organization. I share these with you today as one archival "time traveler" to others, in the hopes that our separate experiences are enough alike that – our collective wisdom being equal to more than the sum of its parts – we will perhaps inspire each other to discover ingenious solutions for the changing archival journey that we share.

Collecting and Processing: New Difficulties and Efforts

What we in the Stanford Linear Accelerator Center (SLAC) Archives and History Office are experiencing lately is, in part, a **hybrid documentation environment**, and collecting records in this hybrid paper / electronic environment has made our work a bit more complicated. The dictionary definition of *hybrid* is: the offspring produced by crossing two individuals of unlike genetic constitution; specifically, the offspring of two animals or plants of different races, varieties, species, etc.²

As the SLAC Archives and History Office staff go about our duties, we are encountering individuals who are creating and saving all of their documentation electronically; some who are creating and saving some documentation only electronically and some only in paper format; and others who are retaining duplicate paper and electronic versions of their documentation.

We are also encountering senior scientists who at the same time as they are writing their own programs for their research, make minimal or no use of the commercial-off-the-shelf software (COTS) that is otherwise in general use at the laboratory. Their reasons may be as strong as their having moral objections to using a particular software, or simple as their never having taken the time to learn it. We have scientists and administrative staff at the laboratory using a variety of computer platforms (Microsoft, Apple, Unix, etc.). We have staff using many flavors of Open Source and commercial software, some supported by the laboratory supported, and some not.

As archivists all know, the procedures and practices for collecting paper and other analog types of records are well established and widely practiced. Archival procedures for dealing with electronic records, however, are only in the early stages of development. From 2003 to 2007, the SLAC Archives and History Office participated in the Persistent Archives Testbed (PAT) research project, about which I have previously reported to this group and elsewhere.³ The PAT project has given us a sense of the methodology we need to use to archive electronic records, as well as a template for the workflow of electronic records from the individual creators desktop into the holdings of the archives. We have developed a set of metadata elements for local use, and we have tested a software tool called PAWN (Producer-Archive Workflow Network) for the transfer of electronic records from creators and users to the archives.⁴

What was outside of the scope of the PAT project, however, and what is proving to be difficult to assimilate on a day-to-day basis, is the integration of electronic records archiving procedures and practices into our repository's established routines. As I reflect on this difficulty, it seems to me that it occurs for two reasons: that electronic records archiving requires a new economic model, and that it also requires a new collection model. The best way to frame the required new economic and collection models is to begin by describing the existing models for both.

The traditional economic model for archives, whether we like to face its reality or not, is that an archival operation is a low-overhead, low-cost affair. Archivists are

modestly compensated, generally make no large expenditures, and normally go about their business, as Ms. Glenn has stated, wearing a “cloak of invisibility.” We are only visible at those moments when we make a dramatic rescue of the organizational memory: for the most part we are barely noticed and, more significantly to our continued survival and well-being, we barely register on our organizations’ bottom lines.

Our traditional, paper-based, or – if you will, analog – collections are fixed entities that are relatively stable if given a proper storage environment away from extremes of temperature and humidity, and out of reach of the pests and predators that prey upon paper, photographs and the like. They are minimally affected by periods of neglect, and for this reason archives are able to function even though they possess large processing backlogs. Further, the use of analog / paper-based archives can be a relatively low-technology operation requiring, at a minimum, only sturdy furniture and good lighting on the part of the archival repository, and, on the part of the researcher, a reasonably well-developed ability to read and write.

Archiving of electronic records, however, completely changes **both** the economic and the collection models for archives. To begin with, electronic records archives are high-overhead affairs, requiring significantly higher expenditures for electricity and hardware than the older, paper-based collections. They also require, because of market realities, more highly paid staff. Computer programmers are a class of workers who are in great demand, and unless they live and work in a less-expensive labor market offshore, they are more generously compensated than most archivists I know.

A second factor affecting the economic model for electronic archival collections is the fact that they are inherently unstable entities, requiring regular and particular attention. Neglect of electronic archives – in the form of staff inattention, or of loss of electrical power, or of hardware malfunction or failure – **can be fatal** in a way that seldom occurs with paper-based / analog collections. Even though computer storage costs continue to decrease, the resource costs for maintaining expanding electronic archival collections will not. This is because the collections will have to be methodically checked for viability, and will also need to be migrated or converted to new hardware and storage media at regular intervals. As more electronic records are archived, the maintenance and migration workloads of the archives will increase exponentially. Even though many of the maintenance and migration processes may be able to be automated, those processes themselves will require regular, skilled attention to ensure that they are functioning properly.

A third factor complicating the economic model for electronic archiving, the electronic environment continues to evolve in ways that make electronic entities more closely resemble non-record cultural artifacts. This is commonly referred to as “Web 2.0.”⁵ Think about wikis, blogs, instant messages, YouTube, MySpace, RSS (Really Simple Syndication) feeds, and any number of web sites that now create web pages “on-the-fly” by executing a small program and pulling information into a “seen once, saved never” web page. Or consider the socially maintained sites that can be updated by users.

Do any of these web entities create records? Is it the business of archives to collect and preserve whatever it is that these entities create?

All of the aforementioned factors lead me to the conclusion that the collection of digital archives is a complicated process that must begin with an appraisal of what is to be preserved before it moves on to the question of how to best preserve it. In order to fully understand the nature and attributes of digital entities, it is necessary to place them in the context of some of the other types of cultural artifacts that have been preserved over time, and to look at how the issues of fixity (stability) and durability (longevity) of these other artifacts have been addressed. Cultural artifacts are products of the “processing,” if you will, of “parts” into “products” of varying fixity and durability. Efforts to define digital records and electronic archives are at the root of current research into the best ways of preserving both, and have been deeply affected by the blurred boundaries between parts, processes and products; by the continually changing nature of digital entities; and by the entities’ low fixity and low durability. Recent research projects into the preservation of digital entities have concluded that a necessary first step is the appraisal of the “significant properties” of digital objects. Now is the time for the creation of a range of digital-derivative products of high fixity and high durability that each effectively captures some significant property of the original “digital performance” or “digital organism.”⁶

The new hybrid digital archives model: multi-tracking

Recent research projects into the archiving and preservation of digital entities have concluded that a necessary first step is the appraisal of the “significant properties” of digital objects⁷. One working definition of the distinctive qualities of digital entities⁸ lists them as:

- ease of replication,
- ease of transmission and multiple use,
- plasticity/fluidity,
- equivalence of works,
- compactness, and
- non-linearity

According to a recent research project, the first task of those undertaking to preserve digital entities is to determine which of the distinctive qualities of the digital entity is significant enough to warrant its long-term maintenance in viable form⁹. Digital preservation research projects are also investigating and developing procedures for maintaining the viability of the distinctive qualities of digital entities over time¹⁰. While the results of these projects are encouraging, the solutions toward which they currently point are neither easy nor inexpensive.

For these reasons, such solutions should be undertaken only when the digital material is truly valuable, when the long-term resources of the creating or custodial organization are sufficient, and when there is no other acceptable preservation alternative. Even when future long-term digital preservation solutions become less labor and resource intensive, it may still be worthwhile for organizations to take a much more traditional approach to preservation of digital materials.

The “traditional” form of digital preservation that should be seriously explored is “multi-tracking.” Using multi-tracking, an organization will exploit digital entities’ most distinctive characteristics (ease of replication, ease of transmission and multiple use, and plasticity/fluidity) to periodically, and at regular intervals, intentionally create fixed and durable products that capture those attributes of the digital material that can be successfully fixed and extended.

This process would be similar to the process of creating periodic system “backups,” but unlike backups, the products of multi-tracking would be self-contained, complete, and system-independent entities. One need only look to the analog realm to locate many precedents for this type of preservation. Musical works are performed by living musicians, they are recorded (and played back) in a variety of media, and they are documented with ink on paper using musical notation systems. In the realm of science, students of natural history have traditionally maintained small, living collections (aquaria, zoos, greenhouses), as well as collections of non-living specimens (skeletal remains, dried plants, preserved tissue collections, etc.) and collections of detailed illustrations.

No one of these preservation “tracks” preserves every attribute of the original entity, but used together they provide good-enough information at a reasonable cost and in a sustainable way. Multi-tracking for preservation tacitly acknowledges that different media have different strengths and weaknesses, and it allows preservation of various attributes of cultural artifacts to be accomplished across a range of media. Preservation studies have repeatedly pointed out that digital preservation efforts must realistically account for institutional priorities, available resources, and the limits of technical feasibility¹¹. Now is the time for the creation of a range of digital-derivative products of high fixity and high durability that each effectively captures some significant property of the original “digital performance” or “digital organism.” Some of these fixed and durable products may be digital, and some may be analog. Some may be high-maintenance, but most should probably be relatively low-maintenance.

Defining and Discovering Best Practices

In the hybridized, multi-tracked electronic/analog archive that we are trying to develop at SLAC, we are embracing the fact that materials are coming to us in a variety of formats. When records are transferred to the Archives and History Office, our standard operating procedure is to process them according to a “triage” approach. Basic processing is accomplished as quickly as possible on all receipts, and a skeletal database record with basic information about the holding is created for each accession. As

accessioned records are consulted in response to reference queries, they are processed further, and their electronic database entries are improved and expanded. This second-level processing is sometimes followed – when resources permit and the importance of the records warrants – by the more traditional folder-level archival processing and collection guide preparation.

This procedure is followed for both paper and electronic records, although the amount of electronic records accessioned into our holdings to date is miniscule, amounting, to date, to 80 separate compact disks. These disks are the archives of SLAC technical publications, which are also available from the SLAC website, on pages administered by SLAC InfoMedia Solutions, our publications department. At present that Department is committed to keeping the publications online and available, but that commitment is subject to the vagaries of management, budget and politics. As funds have been available to us, the SLAC Archives has been creating microfilm copies of the publications, so that the corpus of work is moving forward through time on a multi-track of web pages (InfoMedia Solutions Department), and of offline electronic storage on compact disks and on microfilm in the archives. Should there be a management decision made to “save space” on the SLAC web server and jettison the older publications, the archives will have them stored in both digital (less robust, more accessible) and microfilm (more robust, less accessible) formats, so that they will not be lost.

Images transferred to the Archives and History Office, including photographs and drawings, are handled individually, and are indexed in a separate database (PHOTOINDEX). Since 1999 this database has been web-accessible¹², and in 2002 thumbnail images were added to the database records for a subset of the “most-requested” images. We have begun indexing SLAC’s digital photographs, which are stored in various virtual locations at SLAC, depending upon the divisional and departmental status of the photographer.

In order to provide enhanced access to analog photographs in our collection, we have instituted the practice of scanning photos on request, and creating thumbnails of the scanned photos for inclusion in our database index. Conversely, we plan to assess the newly received and processed collections of digital photographs, and to initiate projects to create prints of those images which provide especially useful documentation of the events they portray. The born-digital photographs are relatively easy to make accessible to our users: we will convert a judicious sample of them to analog format in order to ensure their continued survival in a long-term, economically sustainable and usable form.

Just recently, the importance and value of our multi-track approach to creating hybrid archives has been brought home to us in a very real way. Drastic and unexpected funding cuts in US federal science programs in late 2007 have resulted, at SLAC, in the elimination of over 200 jobs, and the unanticipated early termination of several large research efforts. As long-time staff leave, and large, important projects are ending, the bulk of records – particularly recently created electronic records – being retired to the archives has significantly increased. At the very same time, the prospects for increasing our resources to deal with these growing records retirements have become quite bleak.

Now, more than ever, we need to be judicious in our determinations of what to save, and how to save it.

Conclusion

The most threatened records in the current age are the newest¹³, and it is incumbent on us to not wait for some future “perfect” preservation solution to save our present-day digital entities. Daniel Boorstin, historian and former Librarian of Congress, has written: “The limits of historical discovery come from the physical qualities of objects as much as from the human activities which they suggest.¹⁴” He has suggested that what survives to inform and delight future generations is “collected and protected” information.

As the International Council on Archives (ICA) has stated: “Archives constitute the memory of nations and of societies, shape their identity, and are a cornerstone of the information society.”¹⁵

It is the task of those who would preserve today’s digital entities for tomorrow’s generations to accurately and realistically determine how best to collect and protect what we wish to endure. This means that we must take pains to ensure the future usability – within the limits of available resources – of the works of the digital realm that have been entrusted to our care. Although as archivists we may be cloaked in invisibility, if we want to continue to deserve to wear our “badges of nobility” we must take realistic and measured steps to ensure that our collections remain visible – and usable – for the foreseeable future.

Acknowledgements:

I wish to thank Peter Harper (National Cataloging Unit of the Archives of Contemporary Science) and Karl Grandin (Center for the History of Science, the Royal Swedish Academy of Sciences, Sweden) for organizing Future Proof IV, the conference at which this paper is being presented. Thanks are also extended to Laura O’Hara, Pennington Ahlstrand and Kathy Restaino (SLAC), Sarah Maxfield (DanceSpace), Steven Grote, and James Reed (History San Jose) for their comments on and contributions to the work in progress.

This work was supported by the US Department of Energy Contract No. DE-AC02-76F00515.

BIBLIOGRAPHY and REFERENCES:

Boorstin, Daniel J. 1987. *A Wrestler with the Angel*. In *Hidden History*. New York: Harper & Row.

Brown, John Seely and Paul Duguid. 2000. *The Social Life of Information*. Boston: Harvard Business School Press.

The Cedars Project. 2001. "CURL Exemplars in Digital Archives." *Digital Preservation and Further Information*. Consortium of University Research Libraries: United Kingdom, <http://www.leeds.ac.uk/cedars/DigPres.htm> (accessed September 15, 2003)

Chen, Su-Shing. 2001. "The Paradox of Digital Preservation," *IEEE Computer*: 34(3): 24-28.

Commission on Preservation and Access and The Research Libraries Group. 1996, May. *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*. <http://www.rlg.org/ArchTF/> (Accessed 9/15/2003).

Conrad, Mark. et al. "PAT Project Lessons Learned: Archivists' Perspectives." *Archival Outlook*, November/December 2005. Society of American Archivists: Chicago IL. p. 10-23.

Council on Library and Information Resources. 2001. *The Evidence in Hand: The Report of the Task Force on the Artifact in Library Collections*, <http://www.clir.org/activities/details/artifact-docs.html>.

Deken, Jean Marie. *Archiving SLD Records in SRB: The Persistent Archives Test-Bed (PAT) Project at SLAC in 2004*. SLAC, December 2004. Invited talk presented at Future Proof II, Munich, DE, April 20-22,2005. (SLAC-PUB-10857).

Deken, Jean Marie. "Preserving Digital Libraries: Determining 'What?' Before Deciding 'How?'" Co-Published simultaneously in *Science & Technology Libraries* V.25 No. 1/2, 2004, pp. 227-241 and: *Emerging Issues in the Electronic Environment: Challenges for Librarians and Researchers in the Sciences* (ed: Jeannie P. Miller) Haworth: Binghamton NY. p. 227-241

Gilliland-Swetland, Ann (1995). "Digital Communications: Documentary Opportunities Not to Be Missed," *Archival Issues* 20(1): 39-50.

Glen, Lucinda. "President's Message" *Society of California Archivists Newsletter*. Number 132, Winter 2008

Hedstrom, Margaret. 1991. "Understanding Electronic Incunabula: A Framework for Research on Electronic Records," *The American Archivist*, 54(3): 334-354.

Hedstrom, Margaret. 1997. "Digital Preservation: A time bomb for digital libraries," <http://www.uky.edu/~kiernan/DL/hedstrom.html>. (Accessed 5/20/2008)

Holdsworth, David and Seargeant, Derek M. 2000. A Blueprint for Representation Information in the OAIS Model .
<http://gps0.leeds.ac.uk/~ecl/dh/cedars/nasa2000/nasa2000.html> (Accessed 9/15/2003)

Levy, David M. 2001. Scrolling Forward: Making Sense of Documents in the Digital Age. New York, Arcade.

MacLean, Margaret & Davis, Ben H., eds. 1998. Time and Bits: Managing Digital Continuity. Los Angeles: Getty Conservation Institute.

Moore, Reagan, Chaitan Baru, Amarnath Gupta, Bertram Ludaescher, Richard Marciano, and Arcot Rajasekar. 1999. Collection-Based Long-Term Preservation. San Diego Supercomputer Center. San Diego, California. Submitted to National Archives and Records Administration, <http://www.sdsc.edu/NARA/Publications/nara.pdf>. (Accessed 5/20/2008)

Samuelson, Pamela. 1991. "Digital Media and the Law," Communications of the ACM, 34(10): 23-28.

United States National Archives and Records Administration (NARA). NARA Basic Laws & Authorities. General Counsel and Policy and Communications Staff, National Archives and Records Administration. 2000 Edition.
http://www.archives.gov/about_us/basic_laws_and_authorities/basic_laws_and_authorities.html, (Accessed 3/21/2003)

Warnow, Joan et al. A Study of Preservation of Documents at Department of Energy Laboratories. New York: American Institute of Physics, 1982.

Wolff, Jane. Files Maintenance and Records Disposition: A Handbook for Secretaries at Department of Energy Contract Laboratories. (DOE Report No. C00-5075.A000-16) New York: American Institute of Physics, 1982, Revised 1985

NOTES:

¹ Glen (2007)

² <http://www.yourdictionary.com/hybrid> (retrieved 4/17/2008)

³ Deken (SLAC-PUB-10857, 2004). And Conrad (2005).

⁴ Documentation on this project and the PAWN software is available at <http://www.slac.stanford.edu/history/projects.shtml>

⁵ From Wikipedia.com (retrieved 5/20/2008): a term describing the trend in the use of [World Wide Web](#) technology and [web design](#) that aims to enhance [creativity](#), information sharing, and, most notably, collaboration among users. These concepts have led to the development and evolution of web-based communities and [hosted services](#), such as [social-networking sites](#), [wikis](#), [blogs](#), and [folksonomies](#). The term became notable after the first [O'Reilly Media](#) Web 2.0 conference in [2004](#).

⁶ For further discussion, see Deken (2004)

⁷ Holdsworth & Seargeant, 2000

⁸ Samuelson, 1991.

⁹ Cedars Project, 2001

¹⁰ Moore, et al., 1999; Cedars, 2001

¹¹ CLIR 2001; Commission on Preservation 1996, Chen 2001; Hedstrom, 1991, 1997; MacLean & Davis, 1998

¹² <http://www.slac.stanford.edu/history/photos.shtml>

¹³ Brown & Duguid, 2000

¹⁴ Boorstin, 1987

¹⁵ <http://www.ica.org> (retrieved 4/19/2008)