Pat The Library's Databases: Migrating from VM to UNIX
3/3/1996

SLAC ARCHIVES COLL
SERIES 2 SUBSERIES 1
BOX 9 FOLDER 7

The Library's Databases: Migrating from VM to UNIX

presented to:

SLAC Database Forum

Pat Kreitz

February 23, 1996

Four issues define the problem:

- Size and complexity of databases
- Functional requirements
- World-wide user communities
- (Lack of) Resources and commercial products

The core databases are:

- ABSTRACTS 27389
  e-prints & abstracts, linked to HEP
- BINDING 134
  Journal binding, linked to SERIALS
- BOOKS 21754
  Book catalog for SLAC Library
- BOXES 1010
  Archive storage records
-

Core databases (cont'd):

- CODEN 1427
  Coden for journals, linked HEP
- CONF 6780
  Future/past conferences, linked HEP, BOOKS
- DISTLIST 2609 Distribution lists for SLAC Reports, preprints, and other documents

Core databases (cont'd):

- DOE 1043 Records mgt. info, archives
- EXPERIMENTS 1838 PDG (LBL91), active experiments. Includes URL's to their home pages.
- HEP 319659 Major database, eprints, preprints, theses, reports, journal articles in area of particle physics. Joint project of SLAC and DESY libraries. 26,000 links to full-text via W3.

Core databases (cont'd):

- HEPNAMES 25453 World-wide directory of e-mail addresses for particle physicists
- ILL 10195 Interlibrary Loan records
- INSTITUTIONS 5099 Master list of Particle Physics institutions, linked HEP/BINLIST /SSI & several other conference housekeeping databases. Includes URL's to their home pages

Core databases (cont'd):

- NEWPROTO 1332 Program library for various SLAC library SPIRES tasks
- OLDHEP 16590 Miscellaneous very old unpublished reports (pre-1975)
- PROCESS.CERN 198 Processing file for CERN contributions to HEP
- PROCESS.DESY 1017 Processing file for DESY contributions to HEP

Core databases (cont'd):

- CIRC 55017 Items circulated, linked HEP, BOOKS
- SERIALS 1265 SLAC Library journals
- SLACARC 182 Evaluated archival recs.
- SLACHIST 501 Uneval.archival records
- TIGERLIB 1127 Locator file for crucial ES&H documents examined by DOE Tiger Team. Now archival.

Functional requirements include:

- Maintain or improve response time
- Improve user/staff interface
- Maintain & improve development
- Preserve & enhance functionality
  - ➢ unique data/indexing requirements
  - ➢ distributed collaboration
  - ➢ integration ease

Migration affects three groups:

- World-wide user community of high-energy physicists
- Collaborating institutions and secondary data providers
- SLAC Library staff and SPIRES database owners

User requirements:

- Maintain connectivity, availability and response time
- Inexperienced users: better interface
- Experienced users:
  - minimize change
  - preserve expert search interface
  - enable better mousetraps!

HEP January use statistics:

Mo/Yr	All SPIRES-HEP Searches	QSPIRES	W3
Nov 1993	56,583	10,423	31,715
Nov 1994	97,113	7,377	81,104
Nov 1995	151,847	2,422	146,202
Jan 1996	170,628	2,721	166,075

These numbers refer to FIND or FIN commands in the HEP database only and don't include calls to postscript files or SHOW INDEX or BROWSE etc. The activities of the SLAC library staff are excluded.

Collaborations must be supported:

- Key data suppliers:
  - KEK, DESY, FermiLab
- Miniature "OCLC/RLIN"
- Secondary suppliers or connectors:
  - Raleigh/Durham
  - Serpukhov
  - LANL/eprint archives

Migration choices are constrained:

- SLAC environment:
  - lack of funds/FTE to support migration
  - cost of retraining
- Commercial environment:
  - costs (purchase & maintenance) of commercial systems
  - off-the-shelf doesn't fit
  - 2-3 year lag time for new developments

USPIRES provides a transitional solution:

- Low-cost migration of data to UNIX
- Minimizes retraining, retooling
- Builds on existing expertise: sub-programming languages of SPIRES
- Maximum flexibility to develop and expand system locally
- Permits further evolution incrementally

USPIRES Project:

- Evolved out of emulator
- Based on principle of targeted conversion
- Consists of:
  - SPIRES PL360 code
  - D. Guertin's Emulator code
  - Converted C language code for most heavily-used operations (~1M bytes)

Current status:

- Timing tests reaching acceptable levels
- Production testing stage:
  - DESY prototype installed, testing
  - Library migrating BOOKS
  - Some issues identified/solved
- User testing imminent:
  - convert necessary formats
  - test on SLAC UNIX platforms
  - open to external testing

USPIRES timing comparison:

- Typical times currently look like this for the standard 118 document search:

Version	search time	Default/118	Default/25	WWW/25
VM	0 sec	2 sec	0 sec	6 sec
95.09.25 (previous uspires version)	6 sec	22 sec	2 sec	18 sec
96.02.13 (newest uspires version)	2 sec	16 sec	2 sec	10 sec

USPIRES provides an alternative:

- Buys time to migrate for databases that must move quickly off VM
- 'Permanent' home for low-use/impact databases
- Sub-programming languages substantially unaffected
- User/inputter impact minimized

Challenges:

- Prism interface needed (?)
- Retains VM/interface weaknesses
- Security/access issues not fully resolved
- Migration support is developing
- HEP/SPIRES consortium issues
- Designing beyond current solution/needs--the real challenge!

06/14/1990

SLAC ARCHIVES COLL
SERIES 2 SUBSERIES 1
BOX 9 FOLDER 16

Update on the Web at SLAC

Presented to:
The Key Managers' Meeting
May 14, 1996

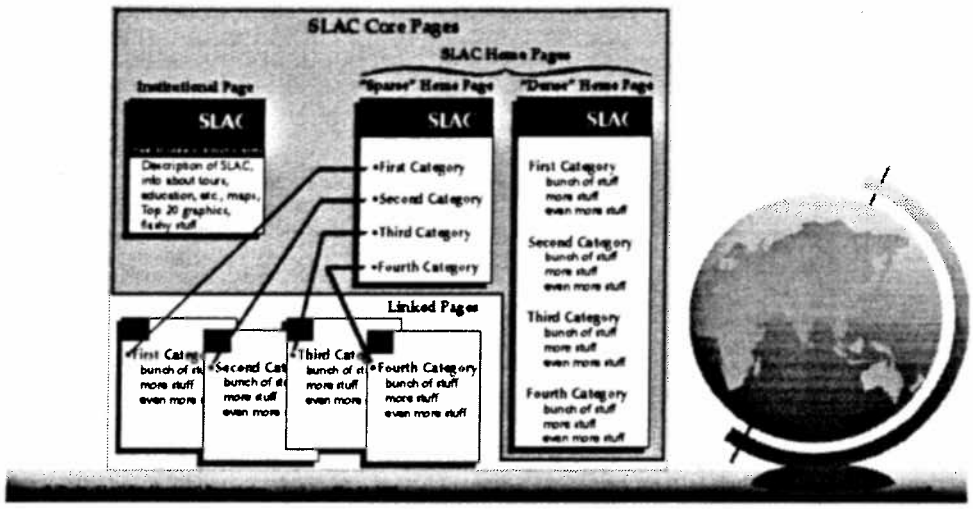
Patricia A. Kreitz
Technical Information Services

Update on the Web at SLAC...

- Status of Web Activity
- Guide to Web's Management
- Current Efforts and Future Developments

Snapshot of Current Web Activity....

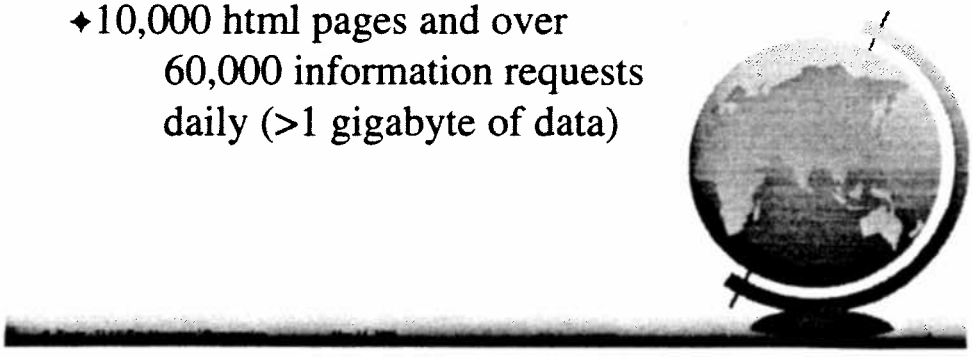
- One home page to three-page model



Snapshot of Current Web Activity....

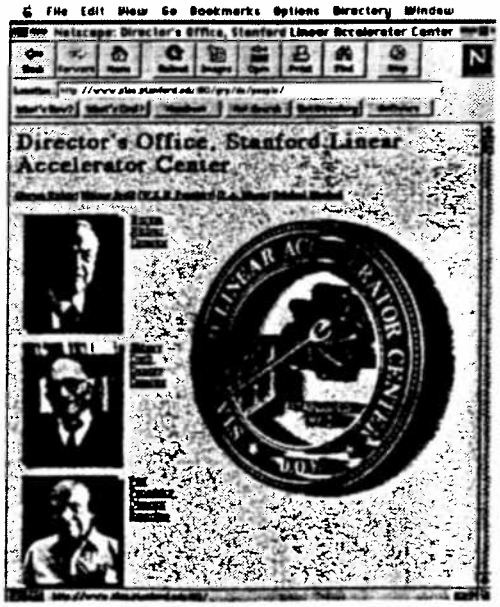
- From 10 Web "Wizards" to over 200 page authors

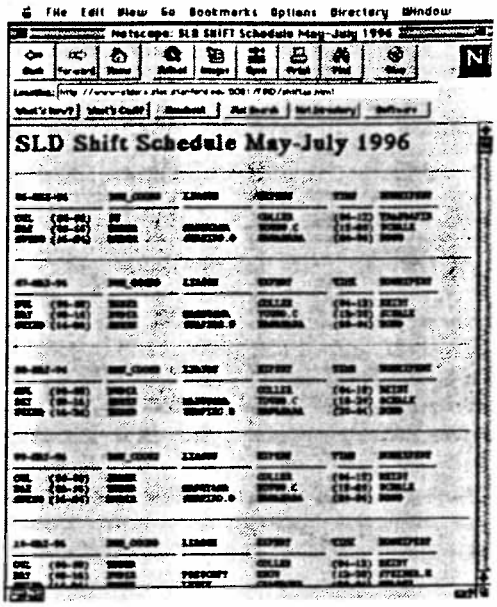
- 10,000 html pages and over 60,000 information requests daily (>1 gigabyte of data)

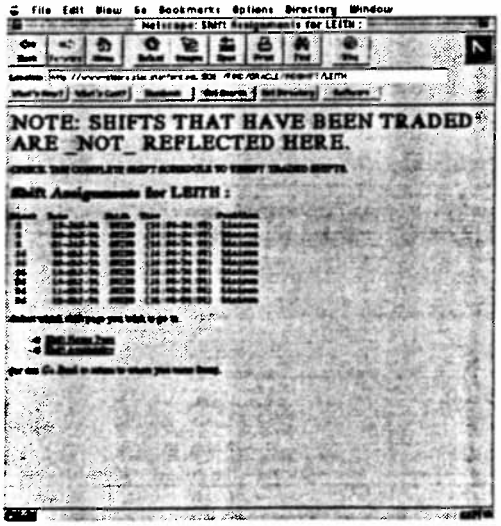


Snapshot of Current Web Activity (continued)....

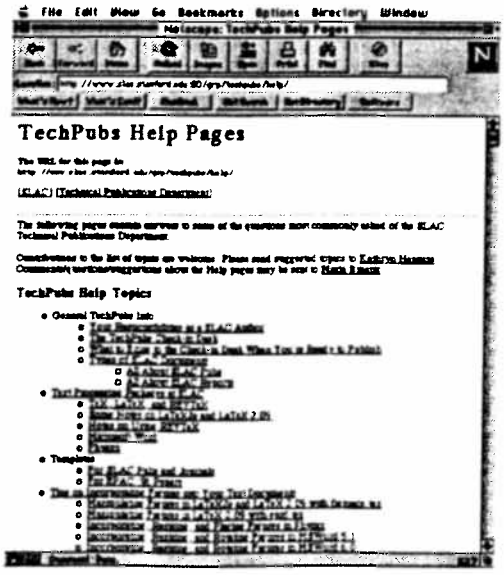
- VM Phaseout Committee's recommendations:
  - communications tool
  - information repository
  - save time and money
  - enable new collaborative models
    "build a shared community
    in a distributed environment"

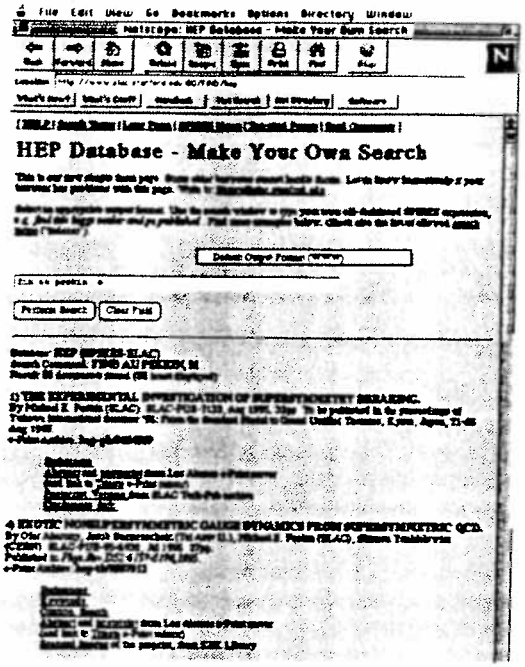


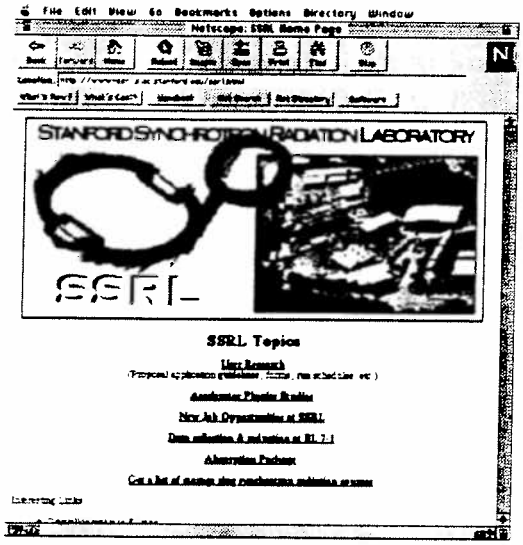

File Edit View Go Bookmarks Options Directory Window
SLD Shift Schedule May-July 1996

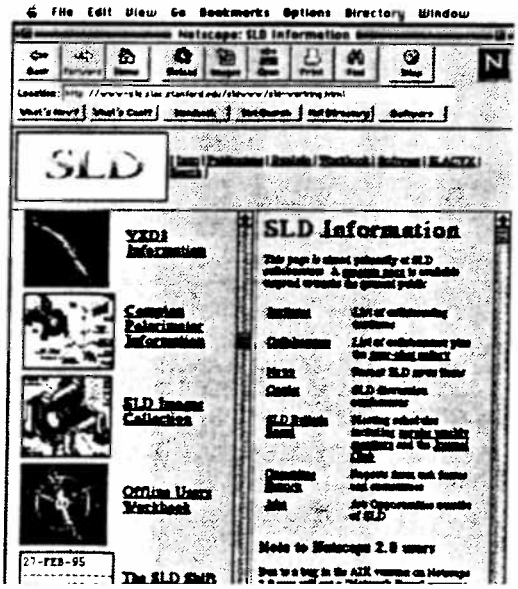

File Edit View Go Bookmarks Options Directory Window
NOTE: SHIFTS THAT HAVE BEEN TRADED ARE _NOT_ REFLECTED HERE.

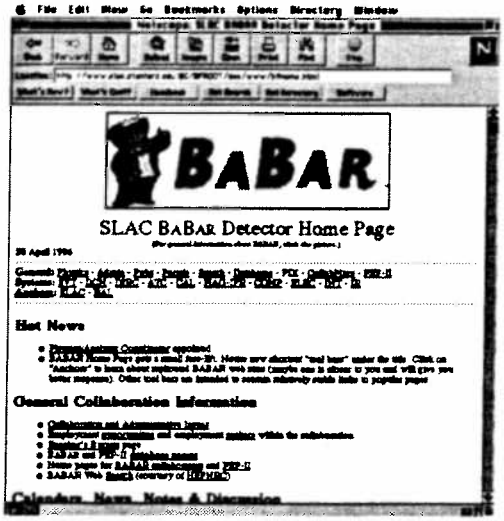
SLAC Administrative Guide on The Web

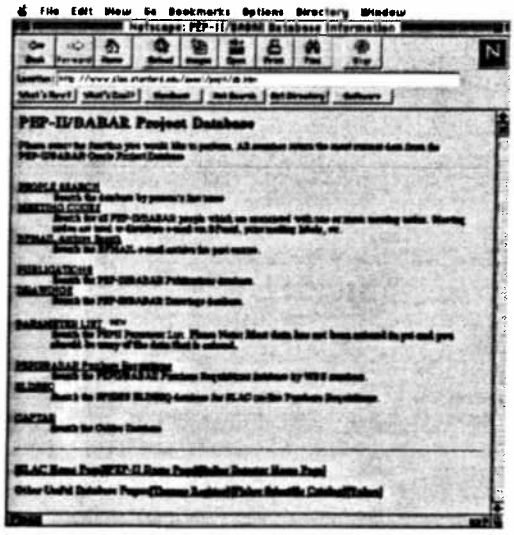



File Edit View Go Bookmarks Options Directory Window
Netscape: HEP Database - Make Your Own Search
HEP Database - Make Your Own Search

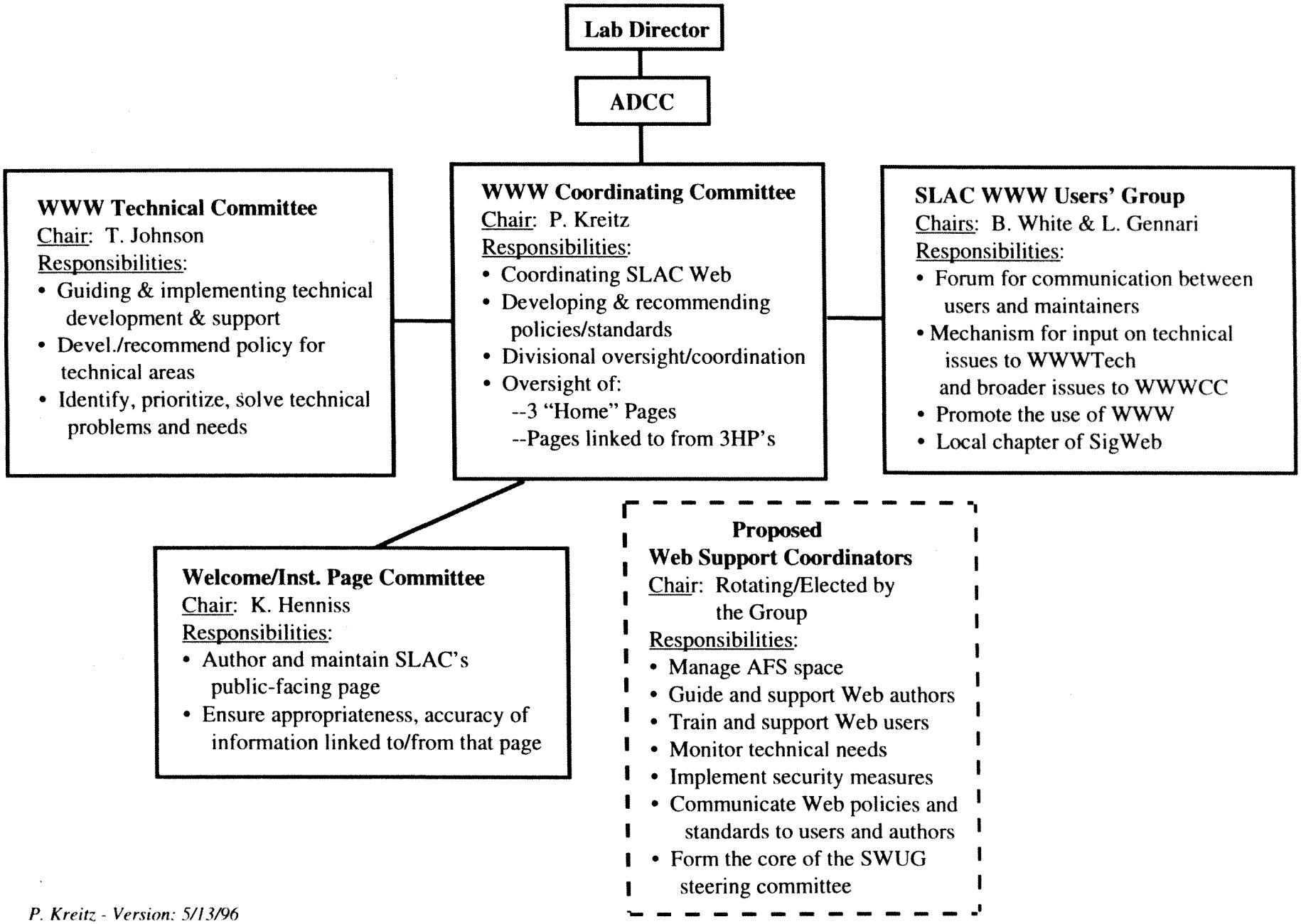

File Edit View Go Bookmarks Options Directory Window
Netscape: SSRL Home Page
STANFORD SYNCHROTRON RADIATION LABORATORY
SSRL
SSRL Topics


File Edit View Go Bookmarks Options Directory Window
Netscape: SLD Information
SLD
SLD Information
VXD3 Information
Compton Polarimeter Information
SLD Images Collection
Offline Users Workbook
27-FEB-95
Note to Netscape 2.0 users


BABAR
SLAC BABAR Detector Home Page
Hot News
General Collaboration Information


PEP-II/BABAR Project Database

Guide To Current Web Management (5/13/96)



P. Kreitz - Version: 5/13/96

Current Efforts: Out of the Terrible Teens....

- Policies and Standards
- Tools and Training
- Technical Support

Current Efforts: Out of the Terrible Teens....

- **Policies and Standards**
  - Security; privacy and confidentiality; individual home pages
  - Standard look-and-feel; links to three top-level pages
- Tools and Training
- Technical Support

Current Efforts: Out of the Terrible Teens....

- Policies and Standards
- **Tools and Training**
  - Organize currently-available resources
  - Identify, and fill in the gaps (e.g. Web Support Coordinators)
  - Advocate for users' and authors' needs
  - Provide training at needed levels for different groups
- Technical Support

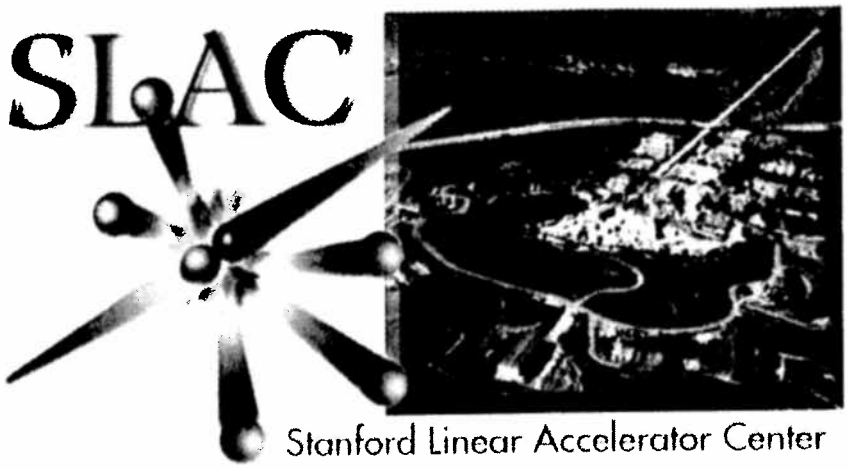
Current Efforts: Out of the Terrible Teens....

- Policies and Standards
- Tools and Training
- **Technical Support**
  - Standardize and support browsers; authoring tools, page/site management tools
  - Create Web interfaces to and among information resources
  - Establish maintainable systems for ensuring security and confidentiality
  - Provide tools for indexing, searching, change management, statistics

Future Developments....

- Organize the sites' information structure so it can evolve flexibly
- Watch trends and developments
- Experiment and change
- Continue to "build shared community"

SLAC ARCHIVES COLL
SERIES 1 SUBSERIES 1
BOX 9 FOLDER 6



A national laboratory funded by the US Department of Energy.

This page provides a public-oriented introduction to SLAC
(home of the first U.S. WWW site).

Media Information How to Reach Us Employment Opportunities

Welcome to SLAC

The Stanford Linear Accelerator Center (SLAC) is a national basic research laboratory, probing the structure of matter at the atomic scale with x rays and at much smaller scales with electron and positron beams. The laboratory is operated by Stanford University under a contract from the United States Department of Energy (DOE).

The combined staff is currently about 1300, 150 of whom are Ph.D. physicists. Typically 800 physicists from universities and laboratories around the world participate in the high energy physics program and 800 scientists from universities and industrial laboratories are active in the synchrotron radiation program.

Points of Interest

What We Do

- About SLAC's Research Programs
- About SLAC's Experimental Facilities
- A Bit of SLAC History

Learn About Science at SLAC

- Tours of the Site
- Education Programs
- The Virtual Visitor Center
- More on Science at SLAC

More on High-Energy Physics

- *Beam Line*, a quarterly journal of particle physics
- SPIRES-HEP Databases, including access to scientific papers by SLAC authors and others in the high-energy physics community

---

Disclaimers, Copyright, and Other Fine Print

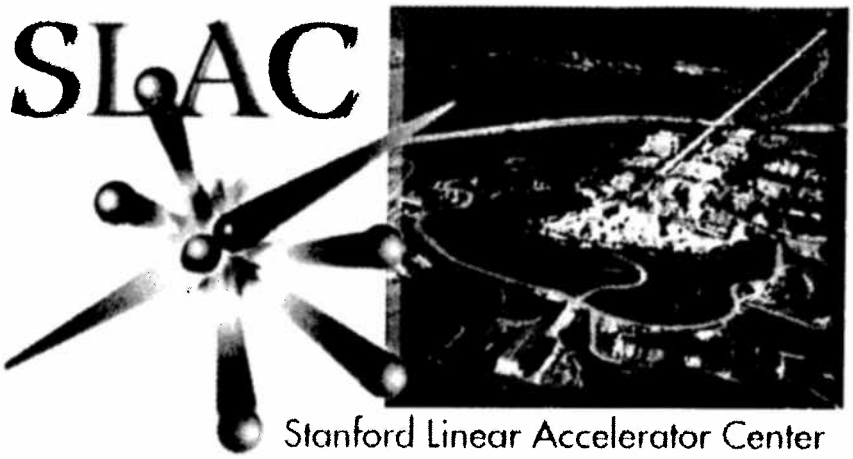
Last modified Tue May 27 18:47:08 PDT 1997
B. C. H. Quark

now at http://www.slac.stanford.edu/welcome/archive/980527/welc
htm

```
<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML//EN">

<html>
<head>
<title>Welcome to the Stanford Linear Accelerator Center</title>
</head>

<body bgcolor="FFFFFF">

<!-- <h1>Welcome to the Stanford Linear Accelerator Center</h1> -->
<p align=center><img alt="Blue colliding particles design superimposed over a
photo of the 2-mile linear accelerator at SLAC"
width=468 height=241
src="/icon/slac1.gif"></p>

<p align=center>
<large>
A national laboratory funded by the <a href="http://www.doe.gov/">US Department of Energy<
/a>.
</large>

<p align=center>
<!-- begin button bar -->

<a href="/"><img
alt ="SLAC Welcome"
border="0"
hspace="0"
width="115" height="18"
src="/icon/btn_welcome_down.gif"></a>

<a href="/highlighted.html"><img
alt ="Highlighted Home"
border="0"
hspace="0"
width="115" height="18"
src="/icon/btn_highlighted_up.gif"></a>

<a href="/detailed.html"><img
alt ="Detailed Home"
border="0"
hspace="0"
width="115" height="18"
src="/icon/btn_detailed_up.gif"></a>

<a href="/slac/www/how-to-use/whatsnew.html"><img
alt ="What's New"
hspace="0"
border="0"
width="115" height="18"
src="/icon/btn_whatsnew_up.gif"></a>

<a href="/slac/www/search/searchslac.html"><img
alt ="Search"
border="0"
hspace="0"
width="115" height="18"
src="/icon/btn_search_up.gif"></a>

<a href="/comp/telecom/phone-dir/search.html"><img
alt ="Phonebook"
border="0"
hspace="0"
width="115" height="18"
src="/icon/btn_phonebook_up.gif"></a>
```

```
</p>


<!-- end button bar -->

<p align=center>

This page provides a public-oriented introduction to SLAC
<br>
(home of the first U.S. WWW site).

<p>
<hr>
<center>
<table>
<tr>
<td align=center valign=center>
<small>
<a href="/slac/media-info/">Media Information</a>
      
<a href="/welcome/location.html">How to Reach Us</a>
      
<a href="emp/emp-opp/emp-opp.html">Employment Opportunities</a>
</small>
</td></tr>
</table>
</center>
<hr>

<h3>Welcome to SLAC</h3>

<blockquote>
The Stanford Linear Accelerator Center (SLAC) is a national basic
research laboratory, probing the structure of matter at the atomic
scale with x rays and at much smaller scales with electron and
positron beams. The laboratory is operated by <a
href="http://www.stanford.edu">Stanford University</a> under a
contract from the <a href="http://www.doe.gov/">United States
Department of Energy (DOE)</a>.

<p>

The combined staff is currently about 1300, 150 of whom are Ph.D.
physicists.  Typically 800 physicists from universities and laboratories
around the world participate in the high energy physics program and 800
scientists from universities and industrial laboratories are active
in the synchrotron radiation program.
</blockquote>

<h3>Points of Interest</h3>

<blockquote>

<h4><img alt="*" hspace=5 src="/icon/redpushpin.gif">What We Do</h4>

  <ul>
  <li><a href="/welcome/program.html">About SLAC's Research Programs</a>
  <li><a href="/welcome/facilities.html">About SLAC's Experimental Facilities</a>
  <li><a href="/welcome/history.html">A Bit of SLAC History</a>
  </ul>

<h4><img alt="*" hspace=5 src="/icon/redpushpin.gif">Learn About Science at SLAC</h4>
  <ul>
  <li><a href="/grp/pao/tour.html">Tours of the Site</a>
  <li><a href="/gen/edu/education.html">Education Programs</a>
  <li><a href="/welcome/vvc.html">The Virtual Visitor Center</a>
```

```
  <li><a href="/welcome/questions.html">More on Science at SLAC</a>
  </ul>

<h4><img alt="*" hspace=5 src="/icon/redpushpin.gif">More on High-Energy Physics</h4>


<ul>
 <li><a href="/pubs/beamline/"><i>Beam Line</i></a>,
a quarterly journal of particle physics
 <li><a href="http://www-spires.slac.stanford.edu/FIND/spires.html">SPIRES-HEP
Databases</a>, including access to scientific papers by SLAC authors and
others in the high-energy physics community
</ul>


</ul>

</blockquote>
<p>
<hr>
<small><font size=-1>
<a href="slac/disclaimer.html">Disclaimers, Copyright, and
Other Fine Print</a></font></small>

<p>

<small>Last modified
Tue May 27 18:47:08 PDT 1997
<br>
<a href="/welcome/bchq.html"><i>B. C. H. Quark</i></a></small>

</html>
```



A national laboratory funded by the US Department of Energy.

This page provides a public-oriented introduction to SLAC
(home of the first U.S. WWW site).

Media Information How to Reach Us Employment Opportunities

Welcome to SLAC

The Stanford Linear Accelerator Center (SLAC) is a national basic research laboratory, probing the structure of matter at the atomic scale with x rays and at much smaller scales with electron and positron beams. The laboratory is operated by Stanford University under a contract from the United States Department of Energy (DOE).

The combined staff is currently about 1300, 150 of whom are Ph.D. physicists. Typically 800 physicists from universities and laboratories around the world participate in the high energy physics program and 800 scientists from universities and industrial laboratories are active in the synchrotron radiation program.

Upcoming SLAC-sponsored Conferences, Workshops, and Meetings

- 1998 Klystron-Modulator Workshop: Development of ILC Modulator Components and Modulators for L, S, and X Band Klystrons, June 29-July 2, 1998
- Sid Drell Symposium, July 31, 1998
- XXVI SLAC Summer Institute on Particle Physics -- Gravity: From the Hubble Length to the Planck Length, August 3-14, 1998
- International Computational Accelerator Physics Conference 1998 (ICAP'98), September 14-18, 1998, Monterey Conference Center, Monterey, California, USA
- 4th RF Workshop (RF98), October 5-9, 1998

For information on other upcoming physics conferences, check the SPIRES Conferences database

Points of Interest

What We Do

- About SLAC's Research Programs
- About SLAC's Experimental Facilities
- A Bit of SLAC History

Learn About Science at SLAC

- Tours of the Site
- Education Programs
- The Virtual Visitor Center
- More on Science at SLAC

More on High-Energy Physics

- *Beam Line*, a quarterly journal of particle physics
- SPIRES-HEP Databases, including access to scientific papers by SLAC authors and others in the high-energy physics community

---

Disclaimers, Copyright, and Other Fine Print

Owner: *Henniss*
Last modified 8 Jun 1998 by Henniss.

(now at http://www.slac.stanford.edu/Welcome/archive/980619/welcome.shtml)

```
<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML//EN">

<html>
<head>
<title>Welcome to the Stanford Linear Accelerator Center</title>
</head>

<body bgcolor="FFFFFF">

<!-- <h1>Welcome to the Stanford Linear Accelerator Center</h1> -->
<p align=center><img alt="Blue colliding particles design superimposed
over a photo of the 2-mile linear accelerator at SLAC"
width=468 height=241
src="/icon/slac1.gif"></p>

<p align=center>
<font size="+1">
A national laboratory funded by the <a href="http://www.doe.gov/">US Department of Energy<
/a>.
</font>

<p align=center>
<!-- begin button bar -->

<a href="/"><img
alt ="SLAC Welcome"
border="0"
hspace="0"
width="115" height="18"
src="/icon/btn_welcome_down.gif"></a>

<a href="/highlighted.html"><img
alt ="Highlighted Home"
border="0"
hspace="0"
width="115" height="18"
src="/icon/btn_highlighted_up.gif"></a>

<a href="/detailed.html"><img
alt ="Detailed Home"
border="0"
hspace="0"
width="115" height="18"
src="/icon/btn_detailed_up.gif"></a>

<a href="/slac/www/how-to-use/whatsnew.html"><img
alt ="What's New"
hspace="0"
border="0"
width="115" height="18"
src="/icon/btn_whatsnew_up.gif"></a>

<a href="/slac/www/search/searchslac.html"><img
alt ="Search"
border="0"
hspace="0"
width="115" height="18"
src="/icon/btn_search_up.gif"></a>

<a href="/comp/telecom/phone-dir/search.html"><img
alt ="Phonebook"
border="0"
hspace="0"
width="115" height="18"
src="/icon/btn_phonebook_up.gif"></a>
```

```
</p>


<!-- end button bar -->

<p align=center>

This page provides a public-oriented introduction to SLAC
<br>
(home of the first U.S. WWW site).

<p>
<hr>
<center>
<table>
<tr>
<td align=center valign=center>
<small>
<a href="/slac/media-info/">Media Information</a>
      
<a href="/welcome/location.html">How to Reach Us</a>
      
<a href="emp/emp-opp/emp-opp.html">Employment Opportunities</a>
</small>
</td></tr>
</table>
</center>
<hr>

<h3>Welcome to SLAC</h3>

<table>
<tr>
<td valign=top>
The Stanford Linear Accelerator Center (SLAC) is a national basic
research laboratory, probing the structure of matter at the atomic
scale with x rays and at much smaller scales with electron and
positron beams. The laboratory is operated by <a
href="http://www.stanford.edu">Stanford University</a> under a
contract from the <a href="http://www.doe.gov/">United States
Department of Energy (DOE)</a>.

<p>

The combined staff is currently about 1300, 150 of whom are Ph.D.
physicists.  Typically 800 physicists from universities and laboratories
around the world participate in the high energy physics program and 800
scientists from universities and industrial laboratories are active
in the synchrotron radiation program.
<p>
<hr width=50%>
</td>

<td valign=top>
<table cellpadding=0 border=0>
<tr>
<td bgcolor="red">
<table cellpadding=5 border=0>
<tr>
<td bgcolor="white">
<font face="Ariel,Helvetica">
<h5 align=center>
Upcoming SLAC-sponsored Conferences,
Workshops, and Meetings</h5>

<small>
```

```
<!--#include virtual="/conf/upcoming.ssi"-->

 For information on other upcoming physics conferences, check the <a
href="http://www-spires.slac.stanford.edu/spires/form/confspif.html">SPIRES
Conferences database</a>

</small>
</font>
</td>
</tr>
</table>

</td>
</tr>
</table>

</td>
</tr>
</table>

<h3>Points of Interest</h3>

<blockquote>

<h4><img alt="*" hspace=5 src="/icon/redpushpin.gif">What We Do</h4>

  <ul>
  <li><a href="/welcome/program.html">About SLAC's Research Programs</a>
  <li><a href="/welcome/facilities.html">About SLAC's Experimental Facilities</a>
  <li><a href="/welcome/history.html">A Bit of SLAC History</a>
  </ul>

<h4><img alt="*" hspace=5 src="/icon/redpushpin.gif">Learn About Science at SLAC</h4>
  <ul>
  <li><a href="/grp/pao/tour.html">Tours of the Site</a>
  <li><a href="/gen/edu/education.html">Education Programs</a>
  <li><a href="http://www2.slac.stanford.edu/vvc/">The Virtual Visitor Center</a>
  <li><a href="/welcome/questions.html">More on Science at SLAC</a>
  </ul>

<h4><img alt="*" hspace=5 src="/icon/redpushpin.gif">More on High-Energy Physics</h4>


<ul>
 <li><a href="/pubs/beamline/"><i>Beam Line</i></a>,
a quarterly journal of particle physics
 <li><a href="http://www-spires.slac.stanford.edu/FIND/spires.html">SPIRES-HEP
Databases</a>, including access to scientific papers by SLAC authors and
others in the high-energy physics community
</ul>


</blockquote>
<p>
<hr>
<small>
<a href="slac/disclaimer.html">Disclaimers, Copyright, and
Other Fine Print</a>
<p>
Owner: <a href="/owner/henniss"><i>Henniss</i></a>
<br>
  <SCRIPT language="JavaScript">
    <!---//hide script from old browsers
    document.write( "Last Modified "+ document.lastModified +" by Henniss.");
         //end hiding contents --->
```

```
  </SCRIPT>

</small>
</html>
```

ARCHIVES COLLECTION
SERIES_____ SUBSERIES_____
BOX_____ FOLDER_____

SLAC–PUB–7636
April 1998

First in the Web, But Where Are the Pieces?*

Jean Marie Deken

Stanford Linear Accelerator Center
Stanford University
Stanford, California 94309 USA

Invited talk presented at the Society of American Archivists Annual Meeting,
8/27/97—8/31/97, Chicago, IL, USA

* Work supported by Department of Energy contract DE–AC03–76SF00515.

A BEGINNING

This is a difficult and intimidating topic to address, attempting to preserve even a part of the World Wide Web, because there is too much material being generated on the World Wide Web for any one person to adequately research even a major portion of it, and because new developments in this medium are occurring at breath-taking speed. But one must begin. The beginning, for me, was in 1996, when I came to the Stanford Linear Accelerator Center, a national basic research laboratory operated by Stanford University under a contract from the United States Department of Energy. SLAC conducts research that probes the structure of matter at the atomic scale with x-rays, and at much smaller scales with electron and positron beams. [1] When I first arrived at SLAC, I was introduced around to the staff as the new archivist, and as people cast about for something to SAY to an archivist, the fact that SLAC was the "first World Wide Web site in the United States" was frequently mentioned (along with mentions of SLAC's past Nobel-prize-winning physics research and current high-energy physics and synchrotron radiation experiments). As I became more familiar with the SLAC archives, I naturally sought out documentation of all of the "milestones" that had been mentioned to me, including SLAC's U.S. primacy in the World Wide Web. My initial forays into the SLAC archives yielded absolutely no documentation of the progress of the World-Wide Web at SLAC. This is not at all surprising, since the Web itself is only 6 years old. [2] There really had not yet been time for the Web to become old enough for anyone to begin to think about documenting the "early days." Also, the Web is a digital phenomenon, existing in an electronic environment. Although SLAC does have an "archive" of electronic data tapes, in addition to a collection of backup tapes, none of these is either in the physical custody nor under the intellectual control of the SLAC Archives and History Office

As 1996 progressed, however, and the Web became an even more widely-dispersed cultural phenomenon, its size and importance at SLAC and in the world at large continued to grow exponentially. There began to be more and more written and said about the World Wide Web in the technical and popular media. At this point, some of the people involved in the early days of the Web — at SLAC and elsewhere — began to get serious about gathering and preserving historical documentation. [3] Also, documenting the history of the Web at SLAC became an issue when there began to be discrepancies in press accounts of the history of the early Web that slighted or ignored SLAC's role. There was one story, in particular, published in a U. S. science laboratory's newsletter that appeared to state that the World Wide Web had made its U.S. debut there.... Reaction at SLAC was swift, and internal e-mails flew back and forth for several days afterward.[4] Why does it matter? It is too early to tell, in my opinion, whether either the Internet or the World Wide Web is the cultural watershed, (or the moral wasteland) that various pundits claim. Yet the World Wide Web does matter to the SLAC Archives and History Office for two very important, and related, reasons. The first reason is that the early Web at SLAC is historically significant: it was the first of its kind on this continent, and it achieved new and important things. The second reason is that the Web at SLAC --in its present and future forms — is a large and changing collection of official documents of the organization, many of which exist in no other form or environment. As of the first week of August, 1997, SLAC had 8,940 administratively-accounted-for web pages, and an estimated 2,000 to 4,000 additional pages that are hard to administratively track because they either reside on the main server in users directories several levels below their top-level pages, or they reside on one of the more than 60 non-main servers at the Center. A very small sampling of the information that SLAC WWW pages convey includes: information for the general public about programs and activities at SLAC; pages which allow physics experiment collaborators to monitor data, arrange work schedules and analyze results; pages that convey information to staff and visiting scientists about seminar and activity schedules, publication procedures, and ongoing experiments; and pages that allow staff and outside users to access databases maintained at SLAC.

So, when SLAC's Archives and History Office begins to approach collecting the documents of our World Wide Web presence, what are we collecting, and how are we to go about the process of collecting it? In this paper, I discuss the effort to archive SLAC's Web in two parts, concentrating on the first task that has been undertaken: the initial effort to identify and gather into the archives evidence and documentation of the early days of the SLAC Web. The second task, which is the effort to collect present and future web pages at SLAC, will also be covered, although in less detail, since it is an effort that is only now beginning to take shape.

THE EARLY WEB: ASSEMBLING THE FACTS

One of the first tasks in the enterprise of documenting the early WWW has been to gather together what is already recorded about the Web at SLAC, and to assemble it into some kind of order. Through this process, the role of individuals who were instrumental in the adoption and development of the Web at SLAC can be documented, and the basic facts, the "who, what, when, where and how" of the World Wide Web, duly recorded

WHO

As everyone undoubtedly knows by now, the World Wide Web was invented at CERN, the European Laboratory for Particle Physics, by Tim Berners-Lee [5] based, in part, on a notebook program he had written in 1990 called "Enquire-Within-Upon-Everything." [6] What is not widely known at all, however, is that SLAC Physicist Paul Kunz brought word of the World Wide Web's existence to SLAC in September, 1991, when he returned from a meeting at CERN. Kunz had immediately seen the possibilities of the Web for streamlining access to a very popular high-energy physics database maintained, in part, by SLAC's Library. SLAC staff quickly saw the value of Kunz' proposal, and work to bring this new phenomenon to SLAC began. Technical difficulties ground the work to a halt soon afterwards, however, and it wasn't until Thursday, December 12, 1991 that the first WWW server at SLAC was successfully installed.[7] Soon afterward, George Crane provided an interface between the SLAC WWW server and SPIRES-HEP, the high-energy physics database.[8] (See Figure 1.) After December of 1991, though, the cast of characters involved with the Web at SLAC widened and branched out. At first there was a very informal group that met under the sponsorship of the Library, and called themselves the WWW Wizards.[9] Participation in Web activity by other staff also flourished. Individuals, scientific collaborations and departments within SLAC, including the Stanford Synchrotron Radiation Laboratory, which became part of SLAC in 1992, mounted home pages on the Web in what quickly became a vast array of offerings covering the whole range of activities undertaken at the Center

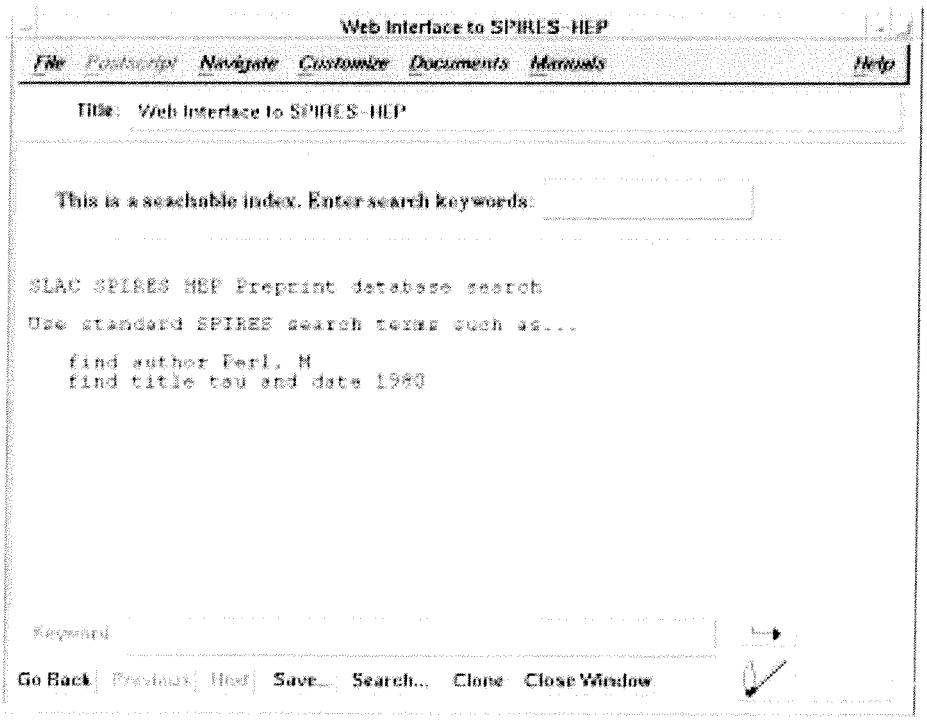


Figure 1

WHEN

The very early days of the World-Wide Web at SLAC date from 1991 to 1994.[10] This is the period that the SLAC Archives and History Office is attempting to reach back and capture, while a separate effort is being made to appraise and arrange for the orderly transfer of any permanently-valuable present and future Web documents. In the beginning, World Wide Web efforts at SLAC had no official status in the organization. This changed some time in 1994, when an ad hoc group called the SLAC WWW Technical Committee was formed to discuss and address technical, "as opposed to stylistic, aesthetic, content or policy" issues. [11] This was followed by the formation of a "SLAC WWW Users Group." Then, in 1995, a World Wide Web Coordinating Committee was established by the Associate Directors' Committee on Computing, "to provide SLAC web authors, groups and committees with policy, guidance, standards, and further support." [12]

WHAT

One of the first Web events at SLAC was a revolution in the delivery of bibliographic information. Since 1974, the SLAC Library had participated in providing SPIRES-HEP, a 300,000 record bibliographic database, to the world particle physics community via the Internet and through clone sites in Europe and Japan. The 1991 introduction of the e-print archives at Los Alamos National Lab, coupled with the World-Wide-Web from CERN, suddenly made it possible for the Library to expand its service and to provide easy linkage between bibliographic database records and the actual full-text of papers.[13] A further improvement was realized when SLAC physicist Tony Johnson introduced "MidasWWW," the first graphical Web browser that could handle compressed files in PostScript, a page description language favored by high-energy physicists because of its ability to describe images in a device-independent manner.[14]

According to Berners-Lee (the Web's creator), the mounting of SPIRES-HEP on the World-Wide Web was a vitally-important factor in the rapid acceptance and utilization of the Web in the international high energy physics community [15]. Driven by increasing pressure to disseminate results as quickly as possible and to as wide a group as possible, the high energy physics community has embraced the World Wide Web because it a significant improvement in the communication of time-sensitive, much-sought-after information. While some of the records mounted on the Web at SLAC are also generated in paper form, many — in fact, less and less as time goes by — exist only in their Web incarnation. The authors see no need to produce the document or information in two formats, and the interactive electronic format of the Web is much preferred

WHERE AND HOW

Because the WWW Wizards group met under the sponsorship of the SLAC Library, and because the maintenance and delivery of the Web HEP Pre-Prints interface was spearheaded by Louise Addis, SLAC Associate Head Librarian, the Library has been able to provide the Archives with documentation of the Wizards' activity and of the HEP Pre-Prints site. Some documentation has also been preserved by individual WWW Wizards, and all of the computer backup tapes from the founding era of 1991 - 1994 still exist (although not on the current main-frame platform). Individual Wizards have conveyed to the Archives paper documentation that they had retained for "historical" purposes in their own files, including paper copies of e-mail messages that were exchanged about setting up the first server, getting it to work properly, and negotiating changes and additions to the SLAC home page. Wizards and members of the Web Coordinating Committee have also provided hard copies of written papers and presentations on the WWW and SPIRES-HEP, and, those papers usually have included examples of then-current SLAC Web pages

APPRAISAL

One thing IS certain at this point, however, and that is that SLAC World Wide Web pages are official records. They fit the statutory definition of records [16] because they are documentary materials, in machine readable format, made or received in the transaction of public business and appropriate for preservation "as evidence of the organization, functions, policies, decisions, procedures, operations, or other activities of SLAC." [17] As official records, SLAC's early World-Wide Web pages, which were created by elements of the Research Division, need to be appraised and

scheduled. Because it is a U. S. Department of Energy contractor, SLAC's records are appraised and scheduled for either temporary or permanent retention based on records disposition schedules negotiated between SLAC, the Department of Energy, and the National Archives and Records Administration

The DOE disposition schedule for Research and Development (R&D) records [18] encompasses R&D records generated within all DOE contractors and national laboratories. The DOE R&D Schedule designates permanent research and development records as "Level I" records, and defines them as the records of :

"Projects which received national or international awards of distinction; active participation of nationally or internationally prominent investigators; research which resulted in a significant improvement in public health, safety, or other vital national interests; scientific endeavors which were the subject of widespread national or international media attention and/or extensive congressional, DOE or other government agency investigation; show the development of new and nationally or internationally significant techniques which are critical for future scientific endeavors, or made a significant impact on the development of national or international scientific, political, economic or social priorities."[19] The World Wide Web fits two of these criteria: it was and is the subject of widespread national and international media attention; and it is a new and internationally significant technique which is critical for present and future scientific endeavors. Although either one of these attributes would alone be sufficient to establish the permanent value of SLAC's early Web documentation, the latter attribute, of course, quite eclipses the former in importance

The identification of which past Web documents fall into the "permanent retention" category is the next step in this process. Retained backups from the early days are being moved into a single storage location in the SLAC Computing Services' vast array of storage sites. However, these backups can not be used on the current computing platform, since SLAC migrated this past spring from a VM main-frame platform to a distributed computing paradigm. [20] Besides the access problem migration has created, there are several other problems as well. The first has to do with the inadequacy of using computer backups as permanent records, particularly for Web documents; the second with the issue of whether it is appropriate to attempt to permanently preserve all of the dimensionality of a permanently-retained Web document

THE "BACKUP" PROBLEM

The first problem with backups is: everything is there. A "backup" contains all of the files that were on a computer system at the time the backup was made, and at any given point in time, much less than half of what is backed-up is of permanent value.[21] For the early web documentation at SLAC, this problem is being handled by knowledgeable staff, who — as previously mentioned — are patiently reviewing old storage disks, as time permits, and moving the World-Wide Web files to a separate storage location. Not every organization will be willing to support such a time-consuming process, however, and it is not a practical long-term solution. Although there is and have been "archive" capabilities on the current and previous computing platforms at SLAC, documents in this Computing Services "archive" are unsystematically selected for retention by their individual creators, who are saving or discarding web documents based on de-centralized, changing, and certainly non-uniform criteria. In addition, although the volume of "archived" documents is smaller than the volume of backups, both share a further attribute that makes them an archival problem. Over time, odd things happen to Web document names: the content, or document attached to a particular "name" changes at regular intervals, because the name is both an address and a name. The "Web document name" is what external and internal systems — the "Internet" and the "intranet" — use to locate a document. In Web lingo, the document name is a "URL," or "Uniform Resource Locator." The "Web document name"/URL is also what you and I type when we want to view a document on the Web. When a Web page owner updates a page, the content, the information — and therefore, the "document" — is changed, but the URL, necessarily, stays the same. Otherwise, system would not be able to find the page, either through a directly-typed request, or through a previously-established link on another page. (See Figure 2.) This ability to link and to renew links constitutes the communicative and functional beauty of the Web, but it makes the World Wide Web an artifact of the present with a disappearing past

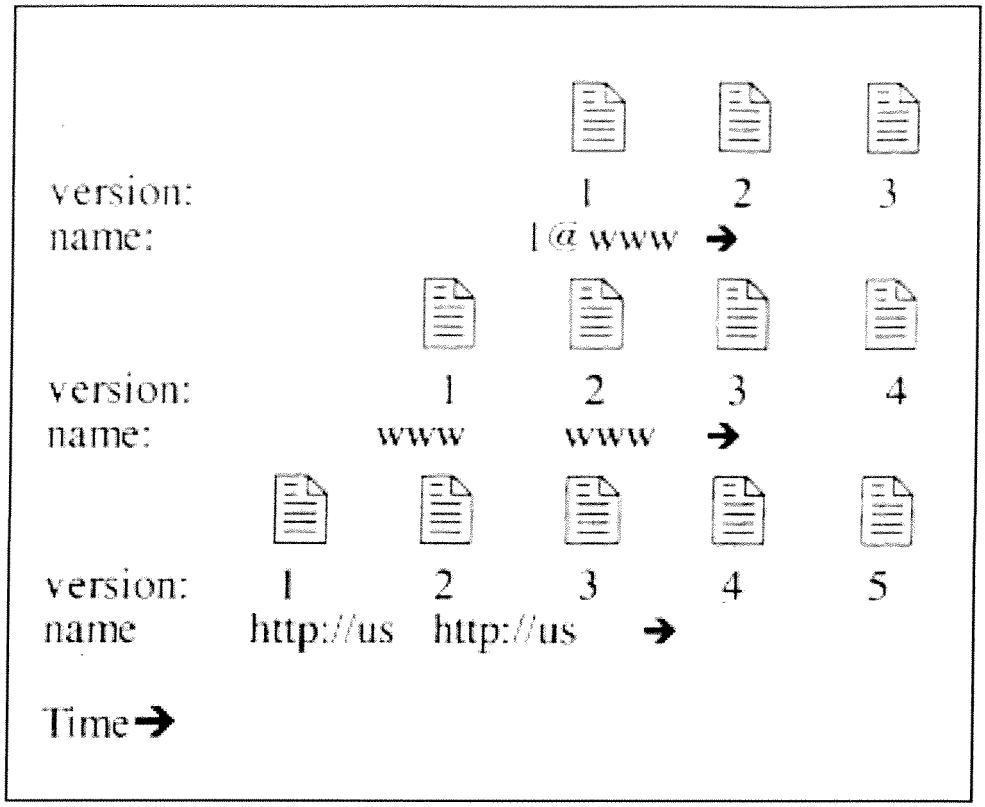


Figure 2

At SLAC, Web page changes have been automatically handled by the backup procedure on the HEP web site by means of a cascading name change that allows a limited number of iterations of a particular web page to be backed up and saved at any given time. In this process, the URL name travels to the newest iteration of the page; and backup name 1 is assigned to the immediately superseded iteration. At the time of the next backup, name 2 is assigned to the first iteration; name 1 is assigned to the second iteration, and the URL name is placed on the most current, or third, iteration. After the upper limit of backup documents of a single page have been stored (at SLAC, this number is six) the next subsequent backup cycle triggers the disposal of the oldest iteration in storage. Using file lists of the backup tapes, one can determine when a particular Web document was created ONLY IF a tape was taken out of the backup cycle, and therefore not overwritten, before an original Web document reached its seventh iteration. In fact, the earliest Web pages at SLAC were set aside through precisely such interventions. In the effort to locate SLAC's first web page on the "archived" backups, a combination of personal recollections of WWW Wizards and information from other electronic records (e-mail) was used to establish a rough estimation of the date of the page. "Archived" backups from that time period were then reviewed for the earliest-dated version of the document name. Using this approach, a digital copy of SLAC's first web page was located on the backups. This same approach will now have to be used to locate the other early web pages which will be appraised as permanent at SLAC, including the first web pages for each experimental group and, most likely, the first web pages for each department. Another historic page that will be sought in the early SLAC web backups is the code for the first World Wide Web page in China, which was set up with assistance from SLAC staff, in support of a collaboration with the Institute of High Energy Physics in Beijing. [22]

THE "DIMENSIONALITY" PROBLEM

Once Web documents have been appraised, and the electronic version of each permanent page has been located, the next issue becomes one of preservation. How should the Web documents be preserved, and who should preserve them? As a U. S. Department of Energy contractor, SLAC has an obligation to retire its permanent records to the

National Archives and Records Administration (NARA). NARA requirements for the retirement of electronic or digital records are quite clear: they must be retired "on one-half inch, seven or nine track reel-to-reel magnetic tape and 3480 class tape cartridges" in flat ASCII or EBCIDIC in a fielded format [23]; or on CD-ROM's which include fielded data files or text files and conform to the International Standards Organization (ISO) 9660 Standard and comply with the American Standard Code for Information Interchange (ASCII) [24] Any Web document transferred to NARA becomes "flattened," that is, loses its interoperability with other Web documents, both by its removal from active relation to the other Web pages to which it was designed to link, and by its conversion to one of the required transfer file formats. In a "flattened" Web document, the addresses of the links remain embedded, but the links themselves are broken, and can no longer be activated. However, Web documents on backups are already somewhat flattened, because they have been removed from the World Wide Web environment. In order to regain their dimensionality, such documents — AND all of the documents to which they are linked — would have to be once again mounted onto a Web server. As this issue arises for more and more organizations, one suspects that most will decide that they do not have the resources nor the institutional willingness both to set up and maintain permanent, fully-dimensional Web documents on an archival Web server on site AND to simultaneously fulfill their primary missions. It can be predicted that, unsurprisingly, most will opt, quite reasonably, to let their fully-dimensional Web documents expire

OTHER OPTIONS

There are some independent repositories, though — most notably The Archive of the Internet [25] and The World Wide Web History Project [26] — which are proposing to preserve fully-dimensional archives of active Web pages on servers dedicated solely to these archival documents. The mission of The Internet Archive is "collecting and storing public materials from the Internet...the Archive will provide historians, researchers, scholars, and others access to this vast collection of data (reaching ten terabytes), and ensure the longevity of this information." [27] The goals of the World Wide Web History Project are somewhat more focused: "The World Wide Web History Project is a collaborative effort to record and publish the history of the World Wide Web and its roots in hypermedia and networking...producing a definitive history and historical archive of the Web." [28] Although both of these organizations have high-powered staff and impressive goals, their ability to achieve those goals remains to be proven. It is unclear, for example, how they will handle web pages that are search forms for internally produced and maintained databases. Search form pages become quite meaningless without active links to the underlying databases, which are not on the web, and are not in markup language format. Nevertheless, I do plan to recommended that, in addition to the required retirement of SLAC Web documents to NARA, a parallel transfer to one or both these repositories be considered, because of the promised full dimensionality. In fact, it is likely that many of SLAC's public web pages have already been collected by the Internet Archive web crawlers, and in 1996, the co-founders of the World Wide Web History Project conducted extensive interviews and videotaping sessions with SLAC WWW Wizards, and these tapes and transcripts are now held by that repository.

CURRENT AND FUTURE WEB DOCUMENTS

The approach that has been taken with SLAC's prior World Wide Web documents has, of necessity, been adopted after the fact. We have an opportunity with the present and future web documents, though, to expedite and streamline the appraisal and transfer process. For organizations which are not as large as SLAC, or which are not as polymorphously web inter-active, the use of backup tapes as the primary instrument of archival appraisal may not present a significant challenge or hardship. For SLAC, however, using backups does not appear to be a practical way of conducting the ongoing business of appraisal and transfer to the archives of permanent Web documents.

In our situation, it will probably be both more efficient and effective to identify which elements of the organization are most likely to generate Web pages that need to be scheduled for permanent retention. Some form of "web crawler" technology, like that used by commercial Web search engines, could then be used to gather iterations of pages from the designated URL's at regular intervals and deposit them in a specified storage location on site. The Web documents in the designated storage site would then be converted to one of the required file transfer formats, and deposited with NARA.

The task for the coming months, therefore, will be to develop and gain approval for a disposition schedule for SLAC Web documents based on this automated gathering approach. We will also need to investigate the mechanics of retiring permanent Web documents to NARA, as well as to one of the previously mentioned independent repositories, if such a transfer is deemed to be worthwhile. As we go about these tasks, we will also maintain a keen interest in developments in thinking about World Wide Web archival and records management issues coming from other projects currently underway, particularly the Syracuse University project to investigate the status of material posted on state agency World Wide Web Sites [29].

NOTES

1. SLAC Welcome Page, http://www.slac.stanford.edu/

2. Hobbes Internet Timeline, http://info.isoc.org/guest/zakon/Internet/History/HIT.html; also The World Wide Web Consortium's "A Little History of the World Wide Web, http://www.w3.org/History.html.

3. WWW History Project, http://www.webhistory.org/project/book.html:

> "...Our goal is to establish a permanent process for recording and disseminating the ongoing history of the Web and networked information, in as close to real time as possible. Much of our work on the Project has been clearing up the six years worth of myths and half-truths which have accumulated around the Web's origins, precisely because there was no definitive history. We want to make it easier next time!"

4. Fermi-News August 16, 1996, p. 1 and ff. (It should be reiterated here that this Fermi-News article did not, in fact, claim that FermiLab was the first U.S. WWW site.)

5. http://www.w3.org/pub/WWW/People/Berners-Lee/.

6. http://www.w3.org/pub/WWW/History.html

7. ibid. Work on installing the server at SLAC was undertaken by both Terry Hung and Paul Kunz. (Paul F. Kunz, e-mail to Jean Deken 30 March 1998, Re: History of WWW at SLAC:

> "The time between September 1 1991 when I brought a Web browser to SLAC and source code for the server and December when [the] server came up is very long. The explanati[o]n is that Terry Hung .... did not know the SLACVM system very well so dropped the project...Finally in December, I put myself to finishing the server installation since I knew how to handle VM and the interface to SLAC Spires...")

8. Addis, Louise.

> "A Brief and Biased History of Preprint and Database Activities at the SLAC Library 1962-1994," p. [26] Appendix to: Addis, Galic, Kreitz and Johnson, "The Virtual Library in Action." Presented at the American Chemical Society (ACS) National Meeting "Chemical Information Symposium" (CINF) "The Library of the Future," Anaheim California, 4 April 1995 (unpublished manuscript, SLAC Archives, unprocessed papers of the WWW Wizards.)

9. 21 September 1994 Memorandum, To: Appendix to WWW Wizards Report to C. Dickens, From: WWW Wizards Committee, Subject: Brief background info on Web at SLAC (SLAC Archives, unprocessed papers of the WWW Wizards.)

10. 21 Sept 1994 Memorandum, To: Appendix to WWW Wizards Report to C. Dickens, From: WWW Wizards Committee, Subject: Brief Background info on Web at SLAC (SLAC Archives, unprocessed papers of the WWWizards; also Addis, Galic, Kreitz and Johnson, "The Virtual Library in Action" p. 5.)

11. http://www.slac.stanford.edu/slac/www/wwwtech/wwwtech.html

12. http://www.slac.stanford.edu/slac/www/wwwcc/charge.html

13. Addis, Galic, Kreitz and Johnson, "The Virtual Library in Action" p.1

14. http://www.cs.indiana.edu/docproject/programming/postscript/ what-is-it.html

15. Berners-Lee, Tim. Keynote Address, April 11, 1997, History/Developers' Day, Sixth International WWW Conference, Santa Clara, CA. (videotape) (Slides available at http://www.w3.org/Talks/9704WWW6-tbl/) Also: Johnson, Tony J. "Spinning the World-Wide Web" SLAC Beamline, vol. 24, no. 3/4, Fall/Winter 1994, pages 2-9.

16. 44 U.S.C. 3301; http://www.law.cornell.edu/uscode/44/3301.html

17. ibid.

18. National Archives Disposition Job N1-434-96-9 (pending)

19. ibid.

20. http://www.slac.stanford.edu/comp/vm/vmmigr.html

21. Estimates for the ratio of permanent to temporary records in an organization's paper-based system place the permanent records at less than 5% of the total volume. (See: Records Management Handbook, "Disposition of Federal Records" National Archives and Records Administration. 1989. p. 4)

22. Cottrell, R. L. A. et al. "Networking With China" SLAC-PUB-6478, April, 1994.

23. 36 CFR 1228.188 "Transfer of Machine Readable Records to the National Archives

24. NARA Bulletin 94-4, Use of Compact Disk-Read Only Memory (CD-ROM) medium to transfer records to the National Archives.

25. http://www.archive.org

26. http://www.webhistory.org/home.html

27. http://www.archive.org

28. http://www.webhistory.org/home.html

29. "Managing Web Sites is focus of Syracuse Study" NAGARA Clearinghouse v. 13, n. 3, Summer 1997, page 1; also http://istweb.syr.edu/~mcclure/

SLAC ARCHIVES COL
SERIES
SUBSERIES
FOLDER

Waterloo SCRIPT is going away February 1!

As has been announced in the VM Migration timetable (http://www.slac.stanford.edu/comp/vm/vmmigr.html), the license for Waterloo SCRIPT is not being renewed when it expires on January 31. We will remove SCRIPT from the system at that time. It will not be legal to run SCRIPT on the SLAC system after that date, because the license with Waterloo is a right-to-use license, which expires when we cancel the license.

There are alternate methods for formatting text on UNIX, Macs, and PC systems. You will have to port any remaining SCRIPT documents to one of these systems. Unfortunately, the complexity of the SCRIPT language precludes automatic conversion. There is not much that SCS can do to make the conversion of existing documents much easier; the best process is probably to begin editing the document in a word processor such as Microsoft Word, delete the script formatting controls and use the native word processing commands to duplicate the effects of the formatting.

From BOEHEIM 27Jan97 97 01 27 16:58:06

VM Archive Data - Last Call

Files which need to be preserved and which were stored on ("round") tape reels, including data archived on VM through April 10, 1989, must be retrieved before VM goes away (now scheduled for mid-June of this year). You are encouraged to read the details at: http://www.slac.stanford.edu/comp/vm/backarch-ann.html, but we are providing a summary below.

To obtain a list of files archived from VM onto tape:

- If your VM account is still active, use the "archive list" command to check the archive dates of the files;
- If you have a Unix account, the "vmarchcat vmuserid" command will build a file that contains archive dates and will include the tape volume-names (names starting with "RS" refer to cartridges, the rest are the older, round tape reels);
- Use the web form (http://www.slac.stanford.edu/comp/vm/slaconly/vmarch.html) to display the files in the archive (available to SLAC users only);
- If you don't have any of the forms of access listed above, contact the SCS Help Desk (xHELP) for assistance in obtaining a list of your archived files.

SCS will retrieve files from archive tapes for a $50 charge per **tape processed** to partially recover costs of retrieving the tapes and maintaining the drives. Due to the condition of the tape media and drives, there is no guarantee of successful recovery of the data, even with the best efforts of the staff. If you wish SCS to perform the file recovery, fill out the Archive Request (http://www.slac.stanford.edu/comp/vm/arch-request.html) form.

Cartridges

Files archived on (square) tape cartridges -- this started April 10, 1989 -- will remain accessible through Unix, but it is **FAR** easier to restore them now. The $50 retrieval charge also applies to cartridges. However, if you wish, you may restore your own cartridges. For further information please refer to: http://www.slac.stanford.edu/comp/vm/backarch-ann.html.

VM Backups and Archives

The phaseout of the VM system marks a significant transition point in the computing infrastructure for the Lab and serves as a time for examination of many policies developed for mainframe-style computing. The era of distributed computing has substantially altered the cost and resource tradeoffs; one such area affecting many computing users is archive and backup. While it is too early to announce a radical change in those policies (it will take some months to develop the full plan in conjunction with interested users), preliminary discussions within SCS underscored the need to clarify the current and future access to archive and backup data associated with the VM system. Since the VM system is currently scheduled to be shut down in May, 1998, this note serves as a warning concerning the schedule on which certain data will no longer be accessible.

Limitations

All media have limited lifespan and all data storage technologies have limited windows of cost effectiveness. The probability of successfully retrieving data from offline media diminishes over time and becomes fairly small at a time period of, say, ten years unless extraordinary measures are taken.

Additionally, hardware technology advances causing it to be impossible to recover data even from well-maintained media. Data stored in particular formats also may create problems as the hardware and software can no longer make sense of it. Examples of current data format problems include binary floating point data stored on the VM system, COPYFILE packed format for data used by various CMS commands, and backups performed in a device-dependent fashion.

These limitations have led us to review policies relating to backups and archives. What follows relates to data from the VM mainframe system, but these limitations will also be a strong factor in new archive and backup policies being developed for Unix and Windows NT systems.

VM Backups

Backups prior to April,1989 (tape reels)

Existing backup reels will be retained until the shutdown of the VM system, at which time the reels will be destroyed.

Backups since April,1989 (cartridges)

Existing backup cartridges will be retained until the shutdown of the VM system, at which time the cartridges will be recycled or destroyed, depending on their age and condition.

VM Archives

Archives made through April 10, 1989 (tape reels)

Our recent experience has been that the tape media have degraded, and that this data may not be recoverable from archive tapes. Hardware to read tape reels is limited and is fragile due to its age and limited availability of parts. Of course, the poor condition of the tapes creates further problems for the reliability of the hardware.

Until VM is shut down, we will make a best effort attempt to recover data from these tapes, within the constraints of our limited staff and hardware resources. Users should be aware the probability of successful recovery is low.

Archives since April 10, 1989 (cartridges)

SCS has some rudimentary Unix tools which can be used to recover data from VM archive tapes. These tools will produce the most satisfactory results on files which were originally in a standard CMS text file format (e.g. program source, listings, memos). Binary files, COPYFILE packed format files, and OS Simulation (CMS mode 4) files will be much more problematic to convert to a usable form in Unix. SCS may have very limited consulting available to provide advice in data recovery.

Prior to the VM Phaseout, requests for recovery of data from archive cartridges will be easier to satisfy but may well require intervention of the tape librarian to place the correct cartridge in the silo. There are no current plans for the removal of cartridge tape drives or the software allowing the VM archive tapes to be read from Unix. Current plans are to retain the VM archive tape cartridges as long as it is feasible to read them with functioning hardware supported by current software (operating system and utilities).

VM Minidisks

VM accounts which have been individually deleted (usually by a group czar) have had their minidisks archived at that time. (This procedure has been followed for many years, not just for accounts closed during the VM Phaseout.)

For accounts deleted from VM in large groups, binary images of the minidisks are accessible in Unix in /nfs/morgue/*group/userid* , where *group* is your two-letter group code, and *userid* is your VM userid. Each minidisk belonging to the account will be a subdirectory. Files from

VM are usually in EBCDIC code and need translation before being readable on UNIX. The 'vmfile' command will translate EBCDIC to ASCII and convert VM record conventions to UNIX conventions.

In general, if your VM access was removed in March, 1997, your VM accounts have been disabled for login, but their minidisks have not yet been moved to the morgue. Such minidisks can be NFS-mounted from UNIX just as before the account was disabled -- e.g., by `cd`-ing to a path of the form `/nfs/slacvm/`*`userid.vadr`* where *`vadr`* is the virtual address of the particular minidisk (e.g., "191" for your A-disk). Accessing a VM file via this path provides an automatic EBCDIC to ASCII conversion. If conversion is not desired, access the minidisk via a path of the form `/nfs/slacvm/binary/`*`userid.vadr`* .

To locate your minidisks, first check the `/nfs/slacvm...` path; if they are not there, then check the morgue. If not found in either place, ask SCS for an archive restore.

WDSF

The Workstation Data Save Facility (WDSF), which used the VM system for its data storage server, is being replaced by the Adstar Distributed Storage Manager (ADSM), which uses a Unix data storage server. More information about this replacement will be available soon. Files archived with WDSF are expected to be converted so they are recoverable using ADSM.

What Do You Do?

Data stored on ("round") tape reels, including data archived on VM through April 10, 1989, must be retrieved before VM goes away, now scheduled for late May of this year. Files archived on ("square") tape cartridges -- starting April 10, 1989 -- will remain accessible through Unix, but it is **FAR** easier to restore them now.

To obtain a list of files archived from VM onto tape:

- If your VM account is still active, use the "`archive list`" command to check the archive dates of the files;
- If you have a Unix account, the "`vmarchcat `*`vmuserid`*" command will build a file that contains archive dates and will include the tape volume-names (names starting with "RS" refer to cartridges, the rest are the older, round tape reels);
- Use the web form to display the files in the archive (available to SLAC users only);
- If you don't have any of the forms of access listed above, contact the SCS Help Desk (xHELP) for assistance in obtaining a list of your archived files.

SCS will retrieve critical files from archive tapes for a $50 charge per **tape processed** to partially recover costs of retrieving the tapes and maintaining the drives. Due to the condition of the tape media and drives, there is no guarantee of successful recovery of the data, even with the best efforts of the staff. If you wish SCS to perform the file recovery, fill out the Archive Request form.

If you have any questions or comments on this information, please send email to archive-admin@slac.stanford.edu

Page Owner: *Bob Cowles*

VM Migration

SLAC
18 October 1996 **Have you completed the VM Migration Questionnaire?**

Page contents

- Introduction
- What are we migrating to?
- How are major services migrating from VM?
- Progress
- What is the schedule for service migration?

Introduction

As described in the SLAC Strategic Plan for Computing, SLAC is moving from mainframe computing based on VM to a distributed computing paradigm. Current VM mainframe services will be made available through client-server technology to users on desktop PCs and Macintoshes, UNIX workstations, and X Terminals.

End of general service for SLACVM is **March 1, 1997.** After that date justification will be required for continued use. The final shutdown of SLACVM will be **December 31, 1997.**

Effective immediately:

- The VM Operating System and all compilers and products are frozen at their current levels. No fixes will be applied to any products on VM.
- CERNLIB is frozen at the current level.
- No new devices such as printers will be attached to VM. No new network printers will be made available to VM. An exception will be made only for new network printers that **replace** existing IMAGEN printers. For information on replacing IMAGEN printers, contact Teri Church.
- Any existing VM account that has been inactive for a period greater than six months will be automatically closed.

The following sections describe the schedule and strategy for transferring major services from VM. Announcements will be made via VM news, the slac.announce.slacvm netnews group, and on this WWW page prior to discontinuing services.

If you have concerns about the VM Migration plans, please contact Chuck Boeheim for assistance.

What are we migrating to?

SLAC has central support for Windows NT, Macintosh, and UNIX. Your choice from these platforms will be governed by the type of computing you do and possibly by the people with whom you

communicate.

- Windows NT and Macintosh have easy-to-use graphical interfaces, are relatively inexpensive, and have many commercial and free applications available. Word processing, spreadsheets, graphics programs, etc. are strengths of these platforms. The electronic mail programs are in general much easier to use than those on UNIX.
- UNIX is primarily command-oriented, although many applications do have a graphical interface, which is displayed with the X-Windows protocol. You may either purchase a UNIX workstation or use X-Windows from an X-Terminal, Macintosh, or PC to access the central UNIX servers. Program development and high performance computing are the strengths of UNIX. The SLAC Compute Farm is made up of UNIX systems. Some VM compatibility tools are available.

You may find you need a mixture of both styles of computing. Since it is relatively easy to access UNIX systems from Macintoshes or PCs, but not vice-versa, a good model for mixed computing is a Mac or PC on the desktop with X-Windows software to access the UNIX systems.

There are many classes and seminars available to help you learn about the new systems.

How are major services migrating from VM?

- **Electronic Mail**
  Electronic mail is readily available on all other supported platforms. SCS recommends and supports the *Eudora Pro* mail reader for PCs and Macs, which is freely available to the SLAC community through a site license. See the classes and seminars list for descriptions and times for Eudora Pro classes.

  There are many mail readers available on UNIX, and SCS supports and recommends the elm mail reader.
- **Netnews**
  There are many newsreaders available for all other platforms that are at least as satisfactory as the one on VM; however, SCS has not yet selected the ones that it will support. Netscape 3.0 is a promising candidate for support, however it is not compatible with the production news server. SCS is presently evaluating several news support models whose goal is to provide reliable news service using news readers on all supported platforms.
- **Batch Computing and Tape Access**
  The **SCS Compute Farm** provides batch computing and has strong similarities to the SLAC Batch system on SLACVM. It provides mechanisms for accessing tapes in the StorageTek Silo or in 8mm or 4mm tape stackers.
- **Databases**
  The following production database applications on SLACVM will continue to be supported on other platforms after SLACVM is shut down. In some cases the migration effort has already begun and in others a migration strategy needs to be addressed.

Application	Migration Strategy	Person Responsible
Binlist	much of functionality will be provided by PeopleSoft; other data provided in Oracle database PERSON;	Diana Gregory
BFMail	functionality assumed by MajorDomo mailing list facility;	George Crane and Bebo White
Depot	Spires application; migration to be determined;	
ELDREQ	will be available in PeopleSoft;	
Stores Catalog	will be available in PeopleSoft;	
Draw	Spires application; will be converted to Unix Oracle;	
Captar	moved to Unix Oracle;	
TMS		Sandra Crawford
Phone		Janet Dixon
Library	will be available via Unix Spires and WWW;	Pat Kreitz
Cables		
Ports		
Purch	will be available in PeopleSoft;	
Trainrpt	will be moved to Unix Oracle;	
ChartA	will be available in PeopleSoft;	
Phonlist	will be available in PeopleSoft;	Diana Gregory
TechDAB		
Stores	available via WWW;	
Pubstore	will be available in PeopleSoft;	
Commits	will be available in PeopleSoft;	
Labor	will be available in PeopleSoft;	
Property	will be available in PeopleSoft and accessible via WWW;	
Cost Reporting	will be available in PeopleSoft;	
SNAP Vendors	will be available in PeopleSoft;	

Owners of SPIRES databases may choose among three basic migration strategies:

1. Discontinue providing the database. This will be the choice for many of the SPIRES databases that are to be supplanted by the new Business Information System.
2. Convert the database to ORACLE, the supported database system on Unix;
3. Move the database to uSPIRES, an experimental portable version of SPIRES.

Statistics on the current usage of all VM SPIRES databases are available.

George Crane has developed guidelines for VM ORACLE table owners migrating to UNIX. These owners should consider doing the following:

1. Exporting and archiving existing tables;
2. Establishing an ORACLE userid in UNIX;
3. Moving the needed tables, forms, etc. to UNIX;

- **Departmental Information Sharing**

The **Web** is taking over the distribution of much of the information that is formally provided by departments and users to others. Classes on accessing the **Web** are given frequently, and classes on providing information via the **Web** are scheduled from time to time.

Casual file sharing, of the sort exemplified by `GIME <username>` commands, is more difficult to provide in a distributed computing environment. The AFS file system provides a common file

system to all supported UNIX systems at SLAC (and indeed to anyone running AFS anywhere on the Internet). However, there are not yet good AFS clients for PCs and Macs, though SCS is actively investigating those and expects to provide them in the future.

Progress

VM Maintenance Reduction	**10/1/96**	Maintenance on the ES9021-580 mainframe has been reduced from 24 hours per day, 7 days per week to a per-call basis.
3800 Printer Removed	**8/96**	The 3800 printer has become too expensive to operate and maintain, and the floor space is needed for an additional tape silo. Please re-direct small print jobs to your local departmental printers, using the PRT command. For larger jobs, SCS has acquiring a Xerox 4090 printer, which can print 40 sheets per minute. Instructions on the use of this new printer will be made available soon.
VM Capacity Reduced	**7/22/96**	The VM system reduced from a 3 CPU configuration to a single CPU. The number of 4490 (cartridge) tape drives has been reduced to 4, and most of the 3380 disk has been taken out of service. These changes are in response to the drastically reduced load on VM after the shutdown of the batch system.
Morgue for Dead Minidisks	**7/22/96**	A morgue is now available for minidisks removed from the VM system. On any UNIX system, cd to /nfs/morgue/*group*/*userid*, where *group* is your two-letter group code, and *userid* is your VM userid. Each minidisk belonging to that account will be a subdirectory. Files from VM are usually in EBCDIC code, and need translation before being readable on UNIX. The vmfile command will translate EBCDIC to ASCII and convert VM record conventions to UNIX conventions. The morgue will provide a snapshot of files on VM when any account is closed, and will have a complete set of all VM accounts when VM is shutdown. The purpose of the morgue is to ease the transition away from VM by providing continued access to VM files after the end of VM service.
VM Account and Minidisk closures	**4/2/96**	On July 1, 1996, unused VM minidisks were deleted and unused accounts closed. All minidisks* that have not been linked since November 10, 1995 are considered unused. All accounts* that have no remaining minidisks after these are removed are considered unused. The data from the minidisks will be recoverable from the VM backup system for the remainder of the lifetime of the VM system.
VM Batch and Staging systems closed	**4/2/96**	The VM Batch and Staging systems restricted to only authorized users who have submitted justification for continued use of the system.

* SLAC-only links

What is the schedule for service migration?

Action	Relevant Date
Budget application moves to PeopleSoft	2/1/97
Justification for VM required	3/1/97
HRMS and Payroll move to PeopleSoft	7/1/97
ELDREQ moves to PeopleSoft	7/1/97
SLAC Library data moves to Unix	7/1/97
Financials application moves to PeopleSoft	10/1/97
Purchasing moves to PeopleSoft	10/1/97
SLACVM shutdown	12/31/97

Future items to be scheduled:

- The 3420 tape drives (round tapes) will be phased out.
- The label printer will be phased out. Many departmental printers can print labels.

Owner: Bebo

VM Migration Questionnaire

Note: The function of this questionnaire is to gather information on the usage of services on SLACVM. Your responses will assist the VM Migration Committee in making the necessary decisions leading up to the shutdown of SLACVM. Your cooperation is appreciated.

Name:

SLACVM userid(s):

E-mail:

SLAC extension:

Which SLAC computing platforms do you use? (check all that apply)

VM ☐ **VMS** ☐ **Unix** ☐

Macintosh ☐ **NT** ☐ **Windows** ☐

Other

When was the last time you logged onto SLACVM?

Today ○ **Less than 1 month ago** ○

More than 1 month ago ○ **More than 6 months ago** ○

Do you feel as though you still need your SLACVM account?

Yes ○ **No** ○

If **No**, please have your group leader submit an account form to close your SLACVM account. You need not complete the remainder of the questions, but select the **Submit** button at the bottom of this form.

What services do you still use on SLACVM? (check all that apply)

Mail ☐ **News** ☐ **WWW** ☐ **Editing** ☐

ELDREQ ☐ **Phone ATOM Support** ☐ **Binlist** ☐

Interactive ☐ **Batch** ☐

Special services and applications (e.g., SPIRES, TeX) ☐ **Other**

What are your reasons for continuing the use of the applications listed in the previous question?

No adequate replacement ☐ **Haven't had time to pursue replacement** ☐

Need training ☐ **No training available** ☐

Other/Explain ☐

Do you still maintain files on SLACVM?

Yes ○ **No** ○

If **No**, you need not complete the remainder of the questions, but select the **Submit** button at the bottom of this form.

If you still maintain SLACVM files, can these files be transferred to another SLAC-supported file system?

Yes ○ **No** ○ **I don't know** ○

If you still maintain SLACVM files, do you expect to need these files after 3/97?

Yes ○ **No** ○ **I don't know** ○

Additional Comments

Thank you for your cooperation in completing this questionnaire. If you have any questions, please send e-mail to the SCS Help Desk or telephone them at x4357.

If you are satisfied with the answers which you have completed, please submit this form.

SUBMIT CLEAR

If you have completed this questionnaire in hardcopy, please fold, staple and send to:

SCS Help Desk
MS 97