Serena Jardine Fenton. A Case Study In The Organizational Development Of A Digital Library: SunSITE - MetaLab - Ibiblio. A Master's paper for the M.S. in I.S. degree. April 2002. 119 pages. Advisor: Gary J. Marchionini

July 2002, SunSITE-MetaLab-ibiblio will celebrate ten years of existence. It has been an exciting decade, full of innovation and experimentation. During this time, the pendulum of information organization has swung from being driven by the technology to savoring the liberation of the new medium – back and forth, never lingering long. Usability, visual interest and technology have led the dance of the development of the digital archive.

This paper is a case study of the SunSITE-MetaLab-ibiblio archive and the vision that allowed the archive to exist, as well as a record of the development of the SunSITE-MetaLab-ibiblio archive in relation to the trends of information organization of that time period. It is a collection of snapshots in the development of one of the world's first and largest digital archives.

Headings:

Digital libraries -- United States – Planning

Digital libraries -- Collection and preservation

Digital libraries -- Design

User interfaces (Computer systems) -- Case studies

To my family, Joe, Charliss, Marieke and Ty, without whose support and good humor all of this study would not have been possible.

ACKNOWLEDGEMENTS

Table of Contents

Chapter 1 – Tracking the Digital Library

Digital libraries are a new phenomenon, invented in the last quarter of the twentieth century and popularized in the last decade of the century. Explorations are ongoing in areas in such topics as what compromises a digital library, how digital libraries should conduct business, what protocols and technologies will prove to be enduring, and how scholars, as well as the general public, will utilize these digital archives.

At the time that this paper is being written, the digital library that began in 1992 as sunsite.unc.edu, is about to turn ten years old. One goal of this paper is to record some of the beginnings of the SunSITE-MetaLab-ibiblio archive and to capture a record of the initial proposals for funding, which allowed the archive to exist. Much of the thought that went into the creation of the digital archive and of its initial grant proposal occurred in informal sessions and email exchanges. Both types of communication have become obscured with the passage of time.

A second objective of this paper is to chart the SunSITE-MetaLab-ibiblio archive's development, in respect to the trends of information organization of the time in which they occurred. It seems obvious now to that optimal usability dictates a index based on a relational database with metadata linkages and xml-Dublin Core compatibilities, but ten years ago, the height of technology was a list of what was in the archive, with a 70-character descriptor limit for each object. Categories of metadata were rough proposals and the word itself (metadata) had not yet been coined. Ten years from now, digital libraries will have made new and significant advances.

This paper is almost a series of snapshots of the archive as it was founded, developed and grew. It is an anecdotal, not a comprehensive view of the growth. This is not an attempt to create a in-depth history of the people who founded and built the archive. Nor is it an attempt to follow the myriad chains of technologies and innovations that were spawned by the archive. It is a recording of the unfolding of the UNC-CH SUN archive. It is a capturing of the digital documents and ephemeral thoughts that went into the archive's creation. I have attempted to place these occurrences within a context of the time and the technologies of the time.

Dance of technology and usability in a digital library

The history of SunSITE-MetaLab-ibiblio (SMI) began in a dark windowless basement room below the old library at the University of North Carolina – Chapel Hill (UNC-CH). It began with a few computer geeks, an offer of money and equipment from a major computer manufacturer, and a chance to create a playground in the arena of the newest, most experimental technology. But in reality, the history of SMI began decades earlier with the dream of creating an information wilderness for an intelligent urbanite that could lead to adventures beyond anything else available in the late 1980s. The creators of SMI grew up with Star Trek and the promise of *space, the final frontier*. The attainment of outer space was not to be viable for their youth, but the creation of an elaborate inner or cyberspace was at hand.

To understand the history of SMI, it is necessary to briefly look back into the history of technology and the Internet and understand the momentum that propelled the gold-rush for information connectedness that drove the computer culture in the last quarter of the 20th Century.

Long before there was a SunSITE or MetaLab or ibiblio, the University of North Carolina at Chapel Hill had the sort of computer geeks that inhabit almost every institution of higher education. As with many, the administration had seen fit to place them in a dark, windowless basement beneath the old main library, with nothing to occupy their imaginations but the luminous glow of green text and black background coming from the terminals before them.   UNC-CH offered a distinct advantage for these computer devotees - it was one of the early sites on the globally connected UseNet computer network

UNC-CH was Internet connected - with fast cable connections. They could share emails with fellow techies across the country, challenge one another in the virtual MUD communities of Dragons and Dungeons and carry on lively conversations in the hundreds of bulletin board systems (BBS) that existed within the UseNet connected world. They stood at the crossroads of imploding and exploding technologies. One path was driven by programming and the love of software, hardware and the accoutrements that could make the vision happen. The other path was the quest to apply this technology to humanity and how it could be implemented. These two paths would begin

a strange seesawing dance that would continue well into the next millennium as one group fed off the other as both rushed forward in a giddy dance of exhilaration and anticipation.

Chapter 2 – UNC's place in the early Internet

UseNet – a presence from the outset

An early map of UseNet groups from 1981 shows UNC as a server node. Throughout the 1980's, UNC was active in the application of networked technologies and experimental networked communications.

```
Newsgroups: NET.general

After welcoming several new sites to Usenet, I'm enclosing the current map.
Any sites which are missing or wrong please let me know.

                USENET Logical Map
                June 1, 1981
                !-      Uucp links
                :       Berknet links
                @       Arpanet links


                                 pdp
                 (Misc)           ! (NC)        (Misc)
             decvax sii  reed  phs--unc--grumpy  duke34  utzoo  cincy teklabs
               ! !    !     !       !    !               !        !      !         !
               ! +--+----+-----+-+--+------------+-------+------+         !
               !        !                 !                                      !
               !                         duke                                    !
               !                          !                                      !
               !        +------+--+--------------------+--------+ !
               !        !         !                            !      ! !
          ucbopt  !  hocsr--mhtsa----research  allegra      harpo-----chico
            :     !         ! !                   !
          ucbcory !         ! eagle    ihnss   vax135  (Bell Labs)
        (UCB) :   !         ! !           !        !
          ucbvax--++----------+--+--+-----+--+------+--------+
            :    @            !         !                 ! (Silicon Valley)
          ucbarpa @     (UCSD) sdcsvax  !            menlo70--hao
            :    @      sdcattb-----+   !              !    !
          ucbonyx @              +-----ucsfcgl      sytek sri-unix
                 @      phonlab-----+
             cca-unix          sdcarl
```

figure 1- 1   UseNet groups, 1981 ¹

Campus-wide Information Service – an Early Internet

By the early 1990s, UNC had established an early campus-wide Information Service (CWIS) that became a model for others. A CWIS was a wide-area network that provided campus information of general interest, online, accessible from any workstation on campus that has communications capability.   The general configuration of a CWIS consists of an information server, terminals which provide access to the server, dial-in access from off-campus, and a link to the Internet.

The CWIS would typically provide such information as event schedules, directories, guides, policies, and other documents. The main function of the server was to provide public interest information. Therefore, this system had to be separate from administrative information systems, such as the academic information database or the financial affairs network.

Additionally, the CWIS had to be accessible throughout the campus by means of networked terminals or workstation computers. The information had to be up to date, which required maintaining an accessible, lively interchange. The UNC CWIS, which was considered an outstanding example of what CWIS could be, by February of 1992 had nine networked stations for information access for use of the general campus population. Though any campus-networked computer could access the system, they were not as readily available as one might hope.

> You can use Info from an Info Center (terminals set up to use only Info), a VAX workstation, or a microcomputer (IBM PC, compatible, or Macintosh) with a modem and communications software with VT100 terminal emulation. ACS has installed Info Centers at the following campus locations:
>
> Student Union (next to the information desk)
> Undergraduate Library
> Davis Library
> South Building lobby
> Manning Hall lobby
> Morehead Visitors' Center
> Personnel Office, Pettigrew Hall
> Phillips Hall (next to ACS's dispatch window)
> Rosenau Hall student lounge [2]

Bulletin Board Beginnings

The bulletin board system that allowed freely contributed and exchanged information was a significant part of the CWIS.   In March of 1992, when setting up UNC's CWIS network, Jim Fullton consulted Usenet groups to gather information for the best practices of constructing the network:

> I am in the process of setting up a bulletin board system for a large university campus. Our (rudimentary) board is running on a DECStation 3100 with 1.2 Gbytes of disk space and a capacity of 24 simulatneous network and dialup users. Perhaps some experienced sysops on the net can lend me some advice and suggestions about what is useful and desirable in a good bbs. Equally important, perhaps you can give advice about things to *avoid*.
>
> Our operating system is Ultrix and we are running a very heavily modified version of XBBS 1.3.
>
> Our main priorities are:
>
>> 1. Communications (Conferencing, mail, and messaging)
>> 2. Information Access (Bulletins, on-line library info)
>> 3. File access (uploads, downloads, interest groups, etc).
>
> Our (very unfinished) board currently provides local messaging, Internet mail access, access to Internet news, conferencing, and several other things.
>
> ------ material deleted sjf ----- [3]

There were no public UseNet responses to this query. A month later, Fulton sent out another inquiry that focused on the implementation of a BBS that could use client-server architecture. The client-server usage was a new concept and just beginning to go into production on an no-experimental basis. A post back to the newsgroup espoused the advantages of the system: "The "client" program would connect to the bbs "server" to access the bbs services. This method would allow the fast use of fancy graphics, local message reading and storage, and lots of other useful functions...TCP/IP applications that work over Internet already do similar things, and there is no reason it couldn't work over the phone system." [4]

The technology existed. UNC-CH had the interested community and the support. There was an established online community who used and was comfortable with the digitally transferred and archived information. Given the funding, the time was right for the archive to blossom.

Chapter 2 – Beginnings of the Digital Archive

Founding members

The Office of Information Technology (OIT) consisted of a number of smaller subdivisions; one of these was the Workstation Development Group (WDG) which was managed by Judd Knott and staffed by a collection of young and eccentric computer programmers.

Before joining OIT, Judd Knott spent eight years in the U.S. Navy teaching and performing computer systems maintenance. He then went on to earn the award of 'Computer Science Student of the Year' and receive his bachelors degree in computer science at Old Dominion University, Summa cum laude in 1980. After a three year hiatus from academia spent building and programming a 36-node parallel computer at NASA Langley, Knott resumed his studies in the Doctoral program at Duke where he completed all of the coursework for a Ph.D. in Computer Science but left before receiving his degree. In 1988 he came to work at UNC's Office of Information Technology, focusing his skills on providing UNIX support and in establishing a variety of computing services in support of Campus Wide Information Service.

Paul Jones had bachelor's degree in computer science from North Carolina State University, but his passion in the early 1990s was for poetry. In 1993 he would receive his MA from Warren Wilson College in poetry.   Robin 'roblimo' Miller of Slashdot described Jones in 2000, "He is a computer scientist, a poet, and a professor of both journalism and library science. He has eclectic tastes in music, a high forehead, hair that ripples over his shoulders, and speaks in an accent you could call Mayberry PhD; imagine a good ol' boy-talking leftover hippie who co-wrote The Web Server Book (which later morphed into The Unix Web Server Book, Second Edition, and you have Paul pegged". [5] At the time of the founding of SunSITE, Jones was providing tech support to UNC users as a systems manager of NeXT, SGI, DEC, IBM, and SUN equipment, Macintosh networking and IBM Mainframe programmer. He also served as Campus network services coordinator.

In addition to Jones and Knott, there was one additional major participant at the founding of SUNSITE, Steve Christensen, an adjunct professor of physics and astronomy at UNC-CH. How the idea for SunSITE actually began is a bit fuzzy, but "Christensen says he pitched the idea as a way to help universities get access to free software and to help scientists avoid the cumbersome and slow process of writing research for publication in scientific journals." [6]

A Proposal to SUN Microsystems

Steve Christensen was involved with SUN Microsystems computers and their newsletter, SunFlash. He simultaneously approached SUN and UNC's WDG with the idea for an Internet archive. On his personal website, Christensen recalls the beginning of SunSITE, "Several years ago, I presented the concept of what now is called the SunSITE project to Sun Microsystem's Education and Research Group. With funding from them, I produced a project design and solicited proposals from major university Sun sites. I recommended the University of North Carolina Office of Information Technology as the first place to run such a project." [7] SUN put Christensen in charge of gathering proposals from academic institutions.

SUN Project Guidelines

Working with SUN, Christensen established a set of primary guidelines for an unnamed archive project. In January of 1992, he emailed three universities, UNC-CH, State University of New York at Buffalo and Harvard, with a 'request for proposal' to be a possible site for an anonymous ftp server. This server would be devoted to Sun and Sun-related public domain, shareware, or free commercial demo and support software and archives. The request for proposal read:

Project Description

The Sun Microsystems Education and Research Group believes
that there is a need for a central location for the collection of
public domain Sun software. There are many hundreds of programs
already located in the over 800 anonymous ftp sites on the
worldwide networks. The anonymous ftp server in this project
will provide, over the next year, a new place for developers in universities
to place their software. This will help these developers obtain
comments about, and an audience for, their work and give Sun users a library
of free software that is organized and focused. This is be a major
service to the academic and research Sun user community.

It is estimated that there are several gigabytes of potential software and other archival information such as the SunSpots and SunFlash files. It is likely that several thousand Sun users will make use of the new server in the first year. Software will need to be organized in a logical fashion with continuing updates of already installed software and the addition of new programs.

Sun is soliciting proposals from a number of universities on how they might set up this project. The following guidelines and questions will help Sun and its representatives to determine which site will best serve the goals of this project.

Guidelines

1) The host institution will need to provide a network link to a
   Sun Workstation. The link should be maintained at all times
   and be accessible to the Internet. An appropriate Sun workstation
   will be provided by Sun. The kind of workstation is yet to be
   determined. So long as the workstation's primary ftp
   use is not affected, this workstation may be used for other purposes.

2) System administration and the setup of the anonymous ftp software
   on the workstation will be provided by the host institution.

3) The minimum duration of this program will be two years. The
   host institution will need to make a commitment for at least two
   years.

4) The host institution will provide logins to the ftp server machine
   so that Sun representatives can add, modify or remove public domain
   programs.

5) The host institution must allow Sun to publicize the availability of
   the ftp server. This will be done through newsletters distributed
   electronically and in print to Sun users.

6) The host institution may add its own Sun related files to the ftp
   server for local use.

7) Optionally, a BITNET access can be provided so that such users can
   receive certain portions of the archives via email requests.

8) The archive should be registered with the "archie" database of
   servers in Canada.

Proposal Requirements

Each proposal should address first how the Guidelines above could most efficiently and economically be satisfied by the proposers institution. Secondly, an estimate of the hardware, software, and personnel requirements to make the project work should be made. Your realistic comments on any of the elements of this project are welcome.

The following questions should be answered:
1) Expertise in setting up and running such a project

2) Costs of the project
3) Viability of the worldwide network connection
4) Reputation in computing of the institution
5) Potential to support Sun at the institution and in the area
6) Future potentials [8]

On January 17, 1992, Paul Jones, with Judd Knott and Jim Fullton responded to the request via email, providing answers to the series of questions that Christensen and SUN had initiated. (See appendix 2-A) Some of the key issues addressed in UNC-CH's response were:

General experience:

"We've been running an internationally accessed anonymous ftp site for over 5 years. Many days we have experienced over 100 ftp sessions in a 24 hour period. We use anonymous ftp as a way to distribute documents and software and fixes." [8]

Experience in Internet archive

"We have quite a bit of experience with the WAIS protocol and with publicly accessible systems. Our extended bulletin board (bbs.oit.unc.edu) averaged over 1200 logins per 24 period this week. We have had logins from six of the seven continents in the past month. We're still waiting for a connection from Antarctica." [8]

Experience in WAIS – proposing this as a way to organize the archive:

"anonymous ftp is a fairly standard set up. The interesting problem is finding what you want once you're there. Besides the inclusion of accurate descriptive README or INDEX files at each directory level, we would use the WAIS protocol to make the paths to the programs easily tracable and the contents known. By setting up a WAIS server on the ftp machine and making available both a simple WAIS client and an X-WAIS client, actually finding the documentation or software you desire should be quite simple. Further, much of the documentation could be WAISed making WAIS an attractive way to scan documents and download them as an alternative or adjunct to ftp." [8]

Staffing requirements:

"1 FTE who would also become the campus SUN contact. I would see the single FTE aided by 20 hours of student support." [8]

Inclusion of software in the archive beyond SUN public domain software:

"My experience tells me that demo software does very well in such an environment. SUN will have to remember that demo software often escapes. We have (for other vendors) set up demo accounts that are "captive" that is usable by a special login that requires no

password. This is a better alternative than actually distributing the software by ftp or other means." [9]

Jones and Knott were intrigued by the archive possibilities and UNC-CH, though them, was the first institution to respond to Christensen's request. However, a digital archive was a new concept and the responsibilities of what this might entail for the sponsoring institution were not fully understood. Jones expressed a few of these issues along with his answers to Christensen's questions.

> I'd like a chance to talk to the Associate Provost before speculating on the long term or committing UNC for that matter. There a few questions that we'll have to pass on to the University lawyers.
> Problem and how to deal with them is another paper in itself. Just the security angle, takes quite a bit of discussion. Of course, we would not allow uploads or downloads of untested programs because of liability problems. [9]

UNC Recommended as Best Candidate for SUN Archive

On February 4, Steve Christensen wrote to SUN to say that he had received the information from the three universities, had evaluated it and was recommending UNC as the best candidate. This email included a formal proposal for the project to SUN for UNC-CH to be an archive of the specified digital materials. (see appendix 2-B) His reasons for recommending the UNC site were largely based on three factors: cost, experience, and a regional interest in information technology. The recommendation section of the proposal reads:

> 1) The UNC Office of Information Technology contains a number of highly expert communications programmers and system administrators who already provide a large and advanced network service to worldwide users. They provide support both for the university campus but also for the off campus Institute for Academic Technology and others. Their office as already done significant evaluation and testing of every major workstation vendors products. They have access to a large force of student assistance. The campus is famous for its computer science department and research into computer graphics and visualization. It has provided numerous workers for the Sun facility in the Research Triangle Park. They an also be expect to be able to provide BITNET access.

> UNC is strongly supporting the WAIS research efforts and it is likely that developers from OIT will take over much of that development. They expect UNC to become "WAIS Central". This is allow the Sun archive to become a central part of any such project. Other vendors are highly supportive of this effort and have provide equipment to OIT. Sun has not.

The other institutions do not have the same connection and in one case even expressed a concern over potential network overload. They each lack the long term large project experience that UNC has.

2) Costs for personnel in New York and Massachusetts will be higher.

3) UNC-OIT has already been running a worldwide network access program with thousands of user connections. The other institutions have not. OIT was able to provide a more detailed plan based on the guidelines two weeks before the other schools. It is a very professionally run organization.

4) UNC is a well-known research university in a high tech area not strongly affected by the bad economic times. The Research Triangle area has a very high presence of research and academic works with the highest density of Ph.D.'s in the country - even higher than in the Boston area. UNC hosted the first NSF conference on workstation use in the physical sciences in 1984.

5) Sun has come under heavy competition for IBM, Data General and DEC in this area with HP and NeXT also present and growing. A Sun project here will help to fend off these other vendors.

6) The three major universities (UNC, Duke, NC State) and many smaller colleges in the area, plus over a hundred high tech companies and government research facilities (EPA, US Forest Service, Glaxo, MCNC, NCSC, National Statistics Research Center, ...) can provide a large resource for Sun development and customers, influencing sites nation wide. A very large number of students in the computational sciences can provide a willing force of programmers for new public domain software. [8]

UNC Awarded Grant to Create First SunSITE

Approval for the proposal from UNC-CH to create a digital archive was given by SUN Microsystems in the spring of 1992. It was seceded that the archive would be called SunSITE. There were one and a half funded positions. It was decided that Paul Jones would function as the system administrator. Initially, the half time position was unfilled. SunSITE officially went on-line in May 1992, using a Sun SPARCstation 2 with 32 megabytes of memory and approximately four gigabytes of disk storage. Paul Jones was director with Judd Knott as the university representative overseeing him. On July 22, 1992, they sent out an announcement for the halftime student position:

> As you may have read in the OIT NewsBrief, OIT has been selected to put
> together and manage SUN's International ftp repository and to develop a
> WAIS-FTP client for OpenLook. This grant includes funding for one part
> time (student) position. The job requires familiarity with the Internet
> and its resources especially FTP and archie. C (or C++) is a must.
> Graphical Interface Design experience and Fortran are helpful.
> [10]

In August of 1992, Jonathan E. Magid (jem) became SunSITE's first employee.

Chapter 3 – SunSITE.unc.edu – Pioneering the Digital Library

Announcement of SunSITE (partial – from the draft version)

SMCC AND THE UNIVERSITY OF NORTH CAROINA ANNOUNCE AVAILABILITY OF FREE OR LOW-COST SOFTWARE FOR EDUCATION AND RESEARCH

CHAPEL HILL, North Carolina - October XX, 1992 - Sun Microsystems Computer Corporation (SMCC) and the Office of Information Technology at the University of North Carolina have joined together to provide a dedicated anonymous-ftp server, "SunSite." SunSite will act as a central respository for a collection of public-domain software, shareware and other electronic material such as research articles and electronic images.

Academic software developers who create software destined for the public domain will be encouraged to submit their creations to the program at **waiting to get e-mail address from Judy Gallegos; she needs to do a test first**. Those who chose to do so will be taking advantage of a free, easy distribution channel as well as increased exposure to every user on the Internet. In addition, much like having research documents published, posting information on SunSite may also lead to additional prestige and tenure credit.

The 200-year-old University of North Carolina at Chapel Hill was the nation's first state university. Since the nineteenth century, it has offered distinguished graduate and professional programs. In February of this year, SMCC donated a SPARCstation 2 through the company's Academic Grant Program to the university. The SunSite project leader at the University of North Carolina is Judson Knott. "Due to the nature of the education and research markets, accessibility to free or low-cost software is extremely important." Regarding this joint project, Knott added, "The SunSite project at the University of North Carolina will allow academics to access public domain software easily and in one central source and will also provide them the vehicle to post research papers or software packages of their own."

Key to the success of any software in the public domain is its quality. Significant quality assurance will be built directly into the SunSite program to ensure that its software is as bug-free as possible. Data and software packages must be submitted to SunSite. After a prompt evaluation, the packages will be accessible to SunSite users.

The Technology

The SunSite server will use an advanced ftp interface and WAIS protocol, which will make it easy to use. Anonymous ftp is a process that permits retrieval of files from a properly configured machine, typically over the Internet, a global wide-area network. An example of the types of files that can be retrieved via anonymous ftp include archived public-domain software, shareware, and technical documentation, as well as other types of electronic information.

Worldwide, more than 1000 anonymous ftp servers provide a wealth of information across the Internet.

Although anonymous ftp servers have been available for some time, the original anonymous ftp interface is difficult to use and has changed very little since the mid 1970s.**NB: Knott thinks anonymous ftp may have been around only since 1984--Judy attempting to verify** Because this interface has been a barrier for average users, anonymous ftp has been almost exclusively the domain of advanced users. SunSite has attempted to overcome this barrier by incorporating Wide Area Information Server (WAIS) a new public technology, developed by Brewster Kahle of Cambridge, Massachusetts-based Thinking Machines. WAIS is an emerging open-systems standard protocol for querying and retrieval of information. It provides an effortless interface by accepting queries in plain English and returning programs that match the descriptions in the queries. WAIS will make anonymous ftp servers much more accessible to less experienced users.

Knott and his colleagues at UNC expect to see an explosion of electronic information in the coming decade. "We must act now to develop the tools to manage this expected information avalanche," says Knott."The development of standard wide-area network protocols such as WAIS and the integration of these standard protocols into new and existing services, such as ftp, are absolutely critical to our success in accessing, publishing and managing electronic information. The SunSite project's integration of WAIS into the standard ftp environment will allow people with little or no previous ftp experience to quickly and easily locate and retrieve relevant electronic information, in a wide variety of formats, from SunSite." [11]

Evangelizing SunSITE

To demonstrate the new Internet archive to the educational community, Magid, Jones and other OIT staff went to Baltimore for the EduCom conference. The goal was to get an Internet connection at the conference. Then they could connect to the archive and digital information access and retrieval.   Unfortunately, things did not go smoothly. First there were problems with getting an Internet connection.   Then once there was an Internet connection, there were still other problems.   After several unsuccessful attempts to connect to SunSITE via the internet, "We called home to be told that an 8 inch water main in the room that contains the head of UNC's network; UNC's bridge to the rest of the world had broken, dumping an entire water tower into the room. But nothing to worry about, because the rushing water broke the superheated steam pipes, which sent 250+ degree Celsius steam onto the Ciscos, melting them," Magid later reported. [12] Staff at UNC then loaded the SunSITE servers onto a truck, moved them to

another location and linked to the Internet from there. The rest of the conference was a success, according to Magid.

Expanding on the Campus-wide Information Server Library

In the time prior to receiving the SUN grant, UNC had created medium-sized, solid collection of information materials on their Campus-wide Information Server Library and were making some available to the networked world at large via the Internet. Their announcement to the bulleting board Usenet group in March of 1992 read:

> UNC's Internet Extended Bulletin Board Service has moved to a bigger and
> faster host. While samba.oit.unc.edu still exists it isn't the bbs host
> any longer. Please connect to our bbs via telnet to bbs.oit.unc.edu. If
> you must use IP numbers, the new address is 152.2.22.80. Folks who have
> already been using the bbs on samba will find their files on the new
> machine. New users will simply create their own ids just as they did
> before. Remember we require a signed release form before you are allowed
> to send mail or post news from the bbs.
>
> What does bbs.oit.unc.edu have to offer?
>
> Access to all the Usenet news groups via nn and/or rn.
> Access to WAIS databases (about 70 at the moment) via SWAIS.
> Access to documents about the Internet.
> Access to about 200 libraries and other bbses via libtel.
> Access to UNC's Campus-Wide Information System (INFO).

Paul Jones and Judd Knott at UNC's Office for Information Technology (OIT) were intrigued by the possibilities of creating an online free information resource. They were proceeding to archive any digitally available information that they could gain access to. By April 21, 1992, the additional materials archived included:

> National Institute of Health documents
> Recipe archive from UseNet
> Reports on a new phenomenon: internet viruses (internet worm)
> Unix tutorials
> Software and directions for joining the internet
> Joke archives
> Full-text electronic books, including:
> the bible
> the Koran
> all of the plays of Shakespeare
> many of Shakespeare's poems
> Project Gutenberg archive of hundreds of classic book texts
> Free software archives for SUN computers
> On Line IETF (Internet Engineering Task Force) Information

SunSITE began with the SUN software and the few, eclectic digital archives that OIT had been maintaining. To expand the content, the SunSITE staff sent out invitations for contribution and collaboration to other institutions and individuals who might have the interest to create a collection, but not the means.

Russian Archives – First Collection

One of the earliest collections made available on SunSITE was an archive of Russian documents from the Library of Congress. Jones moved them over to the server, made them available and sent an announcement on August 3 to the bit.listserv.pacs-l listserv:

> We have the documents all on line on our anonymous ftp site--sunsite.unc.edu in a directory called pub/russia. Each file is available in Russian (in gif format) and in the English translation with background documents associated with each as well. We picked them up from Library of Congress when their availability was announced. The archive was originally on seq1.loc.gov in pub/soviet.archive. We've found them very interesting.   [13] (See appendix 3-A)

Various Collections – September 1992

In September of 1992, Jones wrote an email appraising the state of the SunSITE collections:

> "Our student Jonathan Magid has done a bang up job of gathering "good stuff" at SunSITE and we're getting over 200 ftp sessions per day during the week and about 120-150 per day on the weekends. Thanks to Jon we have an incredible collection of sounds and we should have the Smithsonian photo archives in gif format by Educom (or shortly after). We've also gotten a couple of folks to moderate their special interest areas in neural nets, sustainable agriculture, and model rockets--not the hottest topics but a good way for us to learn how to distribute authority.
>
> We also continue to WAIS index and archive all the Sun realted news groups as well as the Clinton and Bush speeches. We are getting about 300-400 WAIS searches per week day so far. We've also played a bit with the WAIS-FTP indexing and it works just fine. We need more and better readmes in place for that to work just right and I've hired another student to help out in that area." [14]

Linux Archives

A major resource to arrive at SunSITE was the Linux archives. Linux was a new technology. Most of the software materials that were available were on a server called banjo, which was ending the hosting arrangement. Magid contacted the maintainer of the Linux software collection and made arrangements for the entire archive to be moved to SunSITE. This was the beginning of the Linux Archives, project that would put SunSITE into a position of prominence in the Open Source software world and establish the atmosphere of SunSITE as an experimental archive devoted to programming and open source.

> hi,
> I've been talking to Alan, and I have agreed to take the task of replacing
> banjo. I am in the process of setting things up and transfering all
> the linux tree, so there may well be a period (hopefully no more than
> a week) between the demise of banjo and the presence of the linux stuff
> at the ftp site sunsite.unc.edu. I'll post an annoncement to the group
> when things are going again.   [15]

Presidential and Clinton Whitehouse Archives

 The expansion of the collections archive was relatively swift. SunSITE staff contacted the campaign committees of all of the presidential candidates and offered to coordinate, organize, and electronically distribute any information that the committees were willing to make available to SunSITE. A major collection added early in the history of SunSITE was a collection of materials relating to the 1992 Presidential Election campaigns.   Clinton, in particular took advantage of the offer to archive all of his materials in an Internet archive (this was before there was a whitehouse.gov.) "The Clinton people kept sending us more and more good stuff," Jones said. "The Bush people didn't send very much after the convention, and when Jim Baker joined the campaign we didn't get any more from them." [16]

Once Clinton was elected, the White House press office continued to share information with SunSITE for public distribution, including full text of presidential press conferences, statements, town meeting transcripts, and even documents like the federal budget. When important documents like the National Health Security Plan were issued, the White House press office would often send an advance copy to be fully text-indexed

under WAIS and organized for Gopher access via SunSITE. When the printed document was officially released, the electronic document would be put online and often, once the initial printing was distributed, SunSITE was the only means for obtaining timely access to these documents.

On February 8, 1994, the White House Press Corps released an announcement about the national budget being made available electronically, "Today, for the first time ever, a CD-ROM disc with the complete electronic version of President Clinton's 1995 plan for the U.S. Government Fiscal Year 1995 Budget, as transmitted to Congress, was offered to the public…. The four budget documents that are available immediately will be cheaper than the printed versions and they will be more versatile: users will be able to display exact images of the printed pages, search for key words, and copy or print portions for further reference." [17] In addition to the CD, which sold for a list prince of $30.00, the budget was also available via the Internet at SunSITE and for free.

In 1994, an article on SUN Microsystems's site described the government archives:

> This system is believed to be the first time all White House material has been published electronically, including speeches as they were written and as they were actually delivered by the President and Vice President. On average, SunSITE also includes about 200 new press releases issued by the White House Press Office each month.
>
> Some of the White House-related topics found in SunSITE include speeches from candidates Bush and Clinton during the 1992 Presidential race; official statements from the Executive Office of the President for 1993; the Clinton Administration's initiative to develop the Information Superhighway; NATO communiques, press releases and speeches; the Clinton Administration's statement of intent for overhauling health care in the United States; and a collection of White House and Washington photos from the Smithsonian archives.
>
> "This collection is on the cutting edge," says Jones. "We even have government officials outside the country accessing SunSITE just to read a White House press release relocate other information. Although they can get what they need directly from the White House, they say it's much easier and quicker to just log on to SunSITE."
>
> Because all information from the White House is unedited, and not interpreted by an evening newscaster, any citizen can have access to how government is really working. To allow users to find specific information, UNC-CH created a "quick search" command that works across its entire library of information. For instance, if users want to know what President Clinton said about NAFTA, count the number of times he's used the word tax, or just read about press conferences, they can type in the key word or words related to their request and press return. [18]

The response to the availability of information electronically was generally very positive, as this letter attests:

> My major comment is that I am very excited that it has been made readily
> accessible via the Internet. I am a believer in the concept of freely
> available information at low cost to as many people as possible. The
> current explosive growth of Internet use (as well as commercial services
> such as Compuserve) attests to the widespread agreement with this concept.
> Government information has not been readily available until recently, and
> it is good to see another source.
>
> I was also pleased to find the text of the "Government Printing Office
> Electronic Information Access Enhancement Act of 1993" (Senate Bill 564)
> in the same archive (pub/academic/political-science, at sunsite.und.edu).
> Requiring the government to provide electronic access to at least some
> portion of Federal information will hasten the time that ALL information
> is available.
>
> I encourage you to continue providing access to as much information as
> possible. The National Performance Review, Federal Budget, and other
> areas that you currently have are an excellent start on a much more
> comprehensive archive. I don't, however, want to put undue emphasis on
> the political-science archive. All of the other information that you
> archive is equally valuable.
>
> Thank you for your service.
>
> --
> Steve Messinger   stevem@hpdml90.boi.hp.com
> Hewlett-Packard
> Boise, Idaho
> 19

Alternative Technology and Other Archives

A Chatham County farmer, Larry London, moved the archives of several gardening groups to SunSITE, establishing an archive of sustainable agriculture which eventually grew and expanded to become the diverse, multi-layered EcoLand tech Other early archives included an Internet Talk Radio, pioneered by Carl Malamud, which was a professionally produced radio show in Sparc audio format; archives for many groups including welsh language, the Yiddish Mailing List, personality typing systems; multiple economics groups and resources, including the full text of the North American Free Trade Agreement; Dr. Fun, the first web comic; an archive of traditional Chinese music and the OTIS on-line art gallery- "experiments in graphics and collaborative".

Cisco Educational Archives

Enabled by a grant from CISCO networking to explore the role of the Internet in education, in December of 1993, The Cisco Educational Archives (CEARCH) began. It contained a links collection for the kindergarten through 12th grade classroom and teacher, as well as original material explaining how to physically, politically and financially create an internet accessible computer network within a primary or secondary school. CEARCH was an innovative service for the educational community;
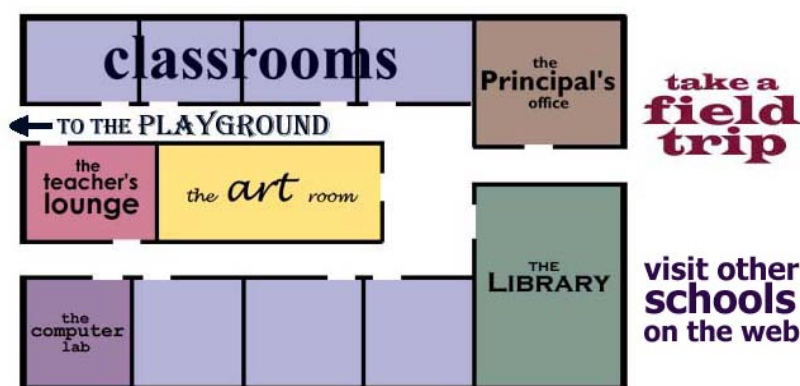

Figure 3-1   CEARCH's virtual schoolhouse

demonstrating and expanding the ways that technology and the Internet could serve and

expand the classroom educational experience. For organizational structure, the site used a metaphorical 'Virtual Schoolhouse' with virtual classrooms, library, art room, principal's office and field trips. Each metaphorical setting corresponded to the actual place within the teacher's frame of reference. For instance, the principal's office held information about administration concerns in light of online issues and resources about how to get your school 'online'. There were lists of schools that were joining the online community and what they were accomplishing. Additionally, from 1995 until 1999, there was a weekly page of "cool pointers, information on your school, or comments in general." [20] For teachers who were unsure of where to begin on the Internet, CEARCH provided a familiar setting to begin their explorations.

The rapid acceptance and popularity of SunSITE was surprising to Jones and Knott. "The initial objective chosen as a success indicator was to achieve an access rate of ten thousand file transfers per month by January, 1993. However, SunSITE

quickly became a popular destination on the Internet and by early October, 1992 was transferring over five thousand files a week." [21]

Outreach to other information providers has helped to build the SunSITE information archives. SunSITE staff began to discover individuals and institutions with valuable information resources, but who lacked either the technical knowledge or the computer resources needed to prepare and distribute their information electronically. SunSITE began a policy of outreach, offering technical consulting services and access to SunSITE resources to assist these fledgling or disenfranchised electronic publishers.

Chapter 4 – 1993: Organizing Open Source

IAFA - Beginnings of Metadata, the Data about Information

Internet Engineering Task Force (IETF) was a group of networkers, scientists, and researchers that had come together to attempt to define protocols, standards and best practices for the smooth operation of the Internet. UNC, prior to the evolution of SunSITE, had been serving as an archive for some of the resources of the IETF. A separate group, The Internet Anonymous FTP Archive Working Group (IAFA-WG), was formed under the auspices of the IETF to study a better way to provide information resources using anonymous FTP, and to create an archive mechanism for sharing information on the Internet. In a draft proposal released June 10, 1992, IAFA-WG noted several of the driving forces behind the effort, including:

> Over the past several years, Anonymous FTP has become the primary method of publishing information in the Internet environment. Anonymous FTP is an application-level service that makes use of the File Transfer Protocol [1], one of the principal protocols of the TCP/IP suite. A well organised and well maintained Anonymous FTP archive (AFA) can provide a cheap and simple way to distribute the programs, documents and datasets that are produced for general availability on the network today. 22
> Anonymous FTP archives (AFA) are not the ideal method for publishing information on the Internet, however they do have the advantage of being relatively cheap and easy to establish and provide near universal access to their contents. With proper attention by archive site administrators, they provide a relatively simple way to distribute information. 23

The IAFA documents were issued in two sections. Part I contained the services provided by the institution, document abstracts, software descriptions, administrative contacts, local time zone and other site-specific details. Part II contains the recommended encoding procedures for the information outlined in Part I.   In essence, this was the beginning of metadata, data created for the purpose of describing data. (See Appendix 5-A)

SunSITE's IAFA Physical Site Information

Following the IAFA guidelines, SunSITE created the following descriptor for the archive, which was stored in the public ftp area:

Filename: IAFA-SITEINFO
Name: calypso.oit.unc.edu

```
Cname: sunsite.unc.edu
Postal-Address: 310 Wilson Library CB# 3460
Telephone: (919)-962-9107
Organization: University of North Carolina at Chapel Hill
Organization-Type: University
Electronic-Address: ftpkeeper@sunsite.unc.edu
Contact: Jonathan Magid
Location:  Chapel Hill, NC 27599-3460 United States
Latitude-Longitude:  35 54 47 +  79 03 22 -
Timezone: -0400 (Eastern Standard Time)
Written-by: Jim_Fullton@unc.edu; Tue Jul 21 09:29:44 EDT 1992
Frequency: 5
Access-Times: 0000 / 2359

Policy:  This site will accept only non-proprietary data.  Please place
it in the "incoming" directory, or send mail to ftpkeeper@sunsite.unc.edu.
This anonymous ftp service is not designed for general data access; all
materials provided through this service will be catagorized and tested by
the site administrators.  Software and documents found suitable for inclusion
in the SunSITE program will be made available for general access.  Documents
describing acceptance criteria are not yet available, but can be requested
from ftpkeeper@sunsite.unc.edu.                    24
```

IAFA Descriptor Set

Similarly, IAFA created a descriptor set that could be applied to each of the resources. This would be placed in the directory with the actual documents (or whatever was being archived) and serve as a standardized identifier from which data could be extracted and used to build resource retrieval databases.

In the actual submittal practice, many archives never had an IAFA identifier and those that did had partial ones. But it was a beginning step toward a standardized index of the information available. A typical IAFA that would be submitted might resemble this one from the rec.models.rockets archive:

```
Title:        ROCDRAG
Version:      4
Description:   Program to estimate Cd for model rockets, with
          altitude simulation.
Author:       John Kallend, METMKALLEND@MINNA.IIT.EDU (312) 567 3050
Maintained by:  John Kallend
Platforms:     PC, (MS DOS)
Copying:       Freely redistributable
Keywords:      DRAG, MODEL, ROCKET, ALTITUDE
```

The Linux Software Movement

The Linux software movement has demonstrated one of the greatest strengths of the Internet. Linux software is a part of the Open Source software movement, which means that the users have access to the source code of all of the software. One of the greatest strengths of open coding is that it allows collaborative development, which can take place in a global Internet arena. Developers can collaborate on a project in an asynchronous manner, and can also work in parallel. The process encourages communication toward a common goal as well as encouraging a peer-review process. Linux developers tend to communicate often and openly in the email list environment.

Chaos in the Linux Archive

SunSITE had been hosting the Linux archive since late summer of 1992. Eric Troan had begun the organization of the Linux archives, but much of the essential material was missing or included in a haphazard manner. A major step toward organizing the software produced by Linux programmers and archived in the Linux archive was taken by on Jeff Kopmanis, a frustrated Linux user. On March 12, 1993 Kopmanis complained in a letter to the Kansas City Linux Users Group about the problem of spending lengthy periods of time downloading software (via a 2400 baud modem), without knowing the total size of a software package. Kopmanis sent the following request to the group:

> For the record, I'm going to start up this version map myself. I'll be
> posting a new topic in the news for this, but in case you miss it:
>
> If you are a developer, please email the following information to
> jeffk@cyberspace.org concerning your package:
>
> package name
> brief description of its purpose (5 lines or less)
> approx size (to nearest K)
> current version
> version numbers of last 3 releases
> required software for your package to run (OS versions, libs, etc)
> version numbers of required software
> FTP address, directory and filename where you distribute your package
> BBS (if any) phone number where you distribute your package
> your email address(es)
> Any other notes you see as important (please be brief, if possible)
>
> There. I've done it. The Linux Third-Party Software Map is officially
> established. The ball is now in YOUR court, developers. [25]

LSM Strategy Proposed

On March 28, Kopmanis had received a sufficient response and he decided to proceed with the project. He sent out another email to the Linux groups: "I have decided to act as a clearing house (collection site) for developers to submit information concerning their latest releases of software. The information collected will be neatly put into a file and uploaded as needed to the tsx-11.mit.edu and sunsite.unc.edu Linux sites. In turn, when users want to find out about versions, sizes, etc of various packages, they will have a resource to turn to." [26] In this early version of his proposal, he requested that developers send him the following information:

> package name
> brief description of its purpose (5 lines or less)
> approx size (to nearest K)
> current version
> version numbers of last 3 releases
> required software for your package to run (OS versions, libs, etc)
> version numbers of required software
> FTP address, directory and filename where you distribute your package
> BBS (if any) phone number where you distribute your package
> your email address(es)
> Any other notes you see as important (please be brief, if possible)
> 26

The First Linux Software Map

By April 1, 1993, Kopmanis had a workable version of the software map designed (see Appendix 5-A). Oral history of the project recalls discussions from and between multiple developers as to what the LSM should include. Kopmanis noted at one point that he had over 60 volunteers willing to assist him with the project. The version that he released on April 1 called for specifications adapted from the AIFA draft guidelines that had already been implemented at SunSITE and other archives: "Basically, the LSM submission format is an extension of the IAFA-PACKAGES format documented in /pub/docs/iafa/draft.part.II on sunsite.unc.edu The extension amounts to the addition of 3 fields: "Approx-Size", "Last-3-Releases", and "Comments" They follow the "Keywords" field in the IAFA-PACKAGES format. Also, the "Platforms" field content

definition has been amended to include information for the LSM's purposes. " [27] The proposed first LSM form looked like:

```
====================================================================
Software Packages Information
----------------------------

Filename: IAFA-PACKAGES

This file contains records with the following fields. The record is
started and delimited from other records by the "Package-Name" field.

Any number of these files may exist in the archive.


Package-Name:   Name of the file or directory containing the package

Title:        Title of the package

Version:        This field can be used if a version number or string is
                associated with the package

Description:    Description of the function of the programs in the package

Author:         Name and Email address of authors if available

Maintained-by:  This field should be included when the current maintainer
                of the package is known. Contact information should be
                included

Maintained-at:  Host name of the "home" of the package if known. This is
                the site at which the most uptodate version of the package
                would be expected to be found

Platforms:      Any requirements or restrictions that the package may
                have in terms of hardware or software (OS) platforms.
                The programming language the package is written in should
                be included.

                (LSM) In the above description, "requirements" refers to
                items such as the Linux version required or any utilities,
                libraries or other packages needed for this package to
                run.  Version numbers should be included for any non-OS
                requirements to ensure that the package will run.


Copying-Policy: The status of the package for copying purposes.

                The most common entries for this field would be
                "Public Domain" or "Freely Redistributable" or
                "Voluntary Payment" (shareware). However since a record
                may exist for software packages not resident on the AFA,
                this field may be entered as "Proprietary" or some other
                form of restricted access.
```

Keywords: Keywords appropriate for users trying to locate the
              package

Approx-Size: (LSM) Size of archive to the nearest K. Size in bytes is
              helpful, but not required.

Last-3-Release: (LSM) The version numbers of the last 3 releases of this
              package.

Comments: (LSM) Any other pertenant information. PLEASE BE BRIEF.

Example record for the LSM:

Package-Name: linux.tads2exe.tar.Z
Title: Text Adventure Development System (TADS)
Version: 2.0.13 (first release, ported)
Description: C-like object-oriented programming language designed to
              make the design and implementation of classic text
              adventures (a la Infocom) a fairly straightforward
              process
Author: Unknown, High Energy Software
Maintained-by: Adam Justin Thornton (adam@rice.edu)
Maintained-at: ftp.gmd.de: /if-archive/programmings/tads
              wuarchive.wustl.edu: /mirrors/if-archive
Platforms: Linux 0.99 (early March, 1993 SLS release); GCC 2.3.3;
              lib4.3
Copying-policy: Freely Redistributable. Copyright held by author.
Keywords: tads, linux, text adventure, High Energy Software
Approx-Size: 631K (631245 bytes)
Last-3-Release: n/a
Comments: Many TADS games are available at ftp.gmd.de in
              /if-archive .
--- 27

Acceptance and Implementation

This format was accepted and a close variation on this form has remained in active use

at the Linux archive. On April 18, Kopmanis counted 837 packages to be identified and

placed into the archive and made a plea for more volunteers to contribute 3 to 5 hours a

week to the effort. [28] The announcement of the release of the LSM-organized Linux

archive came on May 22. The release contained the database of information and a set

of documentation, which consisted of a brief synopsis of the LSM project, instructions on

how to fill out the LSM entry.

The original LSM archive included 920 software packages or files documented in the LSM. Of those, Kopmanis estimated that "1/3 of the entries have full documentation, 1/3 have an average amount of documentation and the other 1/3 are sparsely documented." 29.

The Linux archive was far from being a perfect library cataloging effort, but it was a successful beginning, and one that won support from the Linux community itself. It is also worthwhile to note that while SunSITE was providing the archive space, they were not establishing any requirement or restrictions on the management of that space. Users were allowed to, and in fact, encouraged to experiment with their Internet development and usage.

Chapter 5 – Collections and Policy

Early Collections

From 1992 until 1994, SunSITE was primarily utilizing WAIS for organizing and Gopher for access and transfer of the information. A web browser could view both WIAS and Gopher schemas, so SunSITE stayed accessible for all. WAIS had been the dominant platform for accessing SUNSITE. A partial list of the collection index from 1994 reflected the diversity and eclectic nature of the collection:

> * The American Music Resource
> * Community IdeaNet
> * Dr. Fun!
> * The Fascism Archive
> * Text of the Proposed Health Security Act
> * Jainist Texts
> * Al Gore's National Performance Review
> * Index and abstracts from POETRY magazine
> * Tantric News
> * Phone and Fax Numbers of members of the US congress
> * The Welsh language mailing list
> * Press Releases and Speeches from the Clinton Whitehouse
> * Campaign material from the '92 Bush and Clinton Campaigns
> * The "Cisco Packet" Newsletter
> * ERIC Digests from the US Department of Education
> * The Proposed National Information Infrastructure(html)
> * Search the Linux FAQ
> * Search the Linux Software Map
> * North American Free Trade Agreement
> * All NC laws and regulations
> * All NC Supreme Court cases since 1992
> * Oral Histories from the American South Project
> * Sun Technical Whitepapers
> * Archive of the SunFlash Mailing List
> * Sustainable Agriculture and Organic Farming archive
> * The Winsock Mailing list [30]

Early Digital Libraries

SunSITE began as one of the few large-scale informational sites on the Internet. In 1992 most all Internet sites were non-commercial and generally academic. The majority of the files that moved through cyberspace were text files being retrieved using ftp protocol. As web technology caught on, the participants in the web also began to change. Commercial enterprises began to move to the web. Information began to be

less academic. Companies, such as America Online and CompuServe sought ways to sell their web content rather than allowing full access. The chart below illustrates the rapid growth in the Web. The percentage of commercial sites in the two years from June of 1993 to June of 1996 grew from 1.5 percent to 31 percent

Web Growth Summary		
Month	# of Web sites	% .com sites
6/93	130	1.5
12/93	623	4.6
6/94	2,738	13.5
12/94	10,022	18.3
6/95	23,500	31.3
1/96	100,000	50.0
		31

Figure 5-1 Web Growth Summary

What constituted a digital library has remained vague since the beginnings of the Internet. With the possibility of charging for content, the issue of what an academic digital library could be also came into question. For example Johns Hopkins University funded Project Muse, their digital library, in 1995. The goal of Project Muse was the digitization of 43 scholarly journals published by the John Hopkins University Press, which they would then make available by electronic subscription to participating institutions. In 2002, the prices for a year's subscription to this archive ranged from $3,500.00 (for a high school library) to $14,000 (for an academic institution). Other early academic digital libraries, such as Cyberstacks at Iowa State University of Science and Technology which operated from 1994-1998, were essentially links collections that organized the web, but provided no original content.

Collection Policy in the Early Years

The goal of SunSITE was to provide quality information that would enhance the educational mission of the University of North Carolina at Chapel Hill. An issue facing SunSITE was what projects to support and how to define what projects were worthy. An early decision was made to only archive Internet materials that did not compete with the private sector 'for profit' web sites. But for lack of a clearly articulated collection policy, these acceptance guidelines were loosely structured.

In a 1995 article with SUN which probed the issue Paul Jones commented, ""There are no principles on what is and isn't on the UNC SunSITE server other that what is prohibited by law or Internet "netiquette" For instance, the site is barred from competing with the private sector for profit, so any of the ftp files or Web page must be accessible at no cost." [32]" Basically we get four people in a room and decide if any of us would want any of it," Jones said. "We want cool stuff that people want that's copyright free." [32]

A Formal Collection Policy

SunSITE adopted a formalized collection policy in about 1995 or 1996. The impetus for creating the policy wasn't simply to establish guidelines for what materials could be accepted, but rather to have a defensible position when challenged as to why certain materials might be in the archive. The policy states:

SunSITE Collection Criteria

The SunSITE Project at UNC-CH is the primary site in a network of information services provided by key universities around the world. SunSITE operates as a library, a publishing house, a distribution center and a technology showcase. The materials available on SunSITE represent a diverse community of information providers who obtain space on SunSITE only after meeting the following criteria:

- Do the materials further the teaching, research, or public service mission of UNC?

- Does the collection use technology in innovative and unique ways? Every collection need not be innovative, but it should use up-to-date technology.

- Does the collection add synergetic value to other SunSITE collections? Does it compliment or contradict other collections of music, agriculture, politics, religion, software, etc? An answer of NO to this question should not necessarily disqualify a collection; we may want to begin a new collection area.

- Is all of the material copyright clear and otherwise legal? Exceptions for "fair use" may apply.

- Can and will the keepers of the collection operate in a self-sufficient manner or provide requisite support funding? Exceptions can be made for especially important collections.

- Is the collection non-commercial or operated by a not-for-profit organization or individual?

- Are the materials of national or international interest? Individual pages are permitted for those who are contributing to other broader collections, but parochial materials should be kept on the campus web server.

This new collection policy retained the openness of the earlier collective decision-making process, but articulated the unwillingness to accept everything that was submitted. It encouraged exploration of the possibilities of the web medium, while discouraging web sites of personal information or other items of 'parochial' interest.

Reflections by Early SunSITE Contributors

In January of 2002 as the ten-year anniversary of the launch of SunSITE approached, Jones sent out an announcement to the webgroup, the email list of web site contributors at SunSITE requesting recollections of the beginnings of their sites at SunSITE – MetaLab – ibiblio. Below are a few selected responses that reflection the variations in types of sites that came to the archive and reason why they came.

Dr. Fun, the web's first online comic arrived at SunSITE. David Farley describes the site's history:

> Doctor Fun started at the University of Chicago Library in September, 1993 on a server auspiciously named neuromancer.lib.uchicago.edu. It was also running on another machine called microserver.lib.uchicago.edu, but I administered that one so it crashed all the time. At that time, running a web server was an experimental thing, and the neuromancer experiment went down indefinitely sometime in January 1994.

> Somebody named Ed Statsny, who ran a collection of online artists called "The OTIS Project" (now called "SITO") contacted me and generously offered me a space on OTIS for Doctor Fun. This was good, because I already knew I would have to move Doctor Fun off of a work web server eventually even when things were fixed, so Doctor Fun moved to SunSITE in February, 1994. Soon after I moved things to OTIS, somebody (Paul, Ed, Jonathan?) decided Doctor Fun should really be on its own space, and that's where it's been ever since. [33]

Icons – the early images of the web.

> http://www.ibiblio.org/gio/iconbrowser/ was the first icons collection with search engine. Available since late 1993 it moved to Sunsite in 1995 and became accessible May 21th.[34]

Human rights in Burma

> i believe freeburma, online in 94, was the first human rights website. there was nothing else i could find at the time. of course, there were no search engines either. but i did lots and lots of "[hot links]" hopping from list to list. (how does one "prove" something like that?) – glen [35]

Elvis – the King lives on

The Original Unofficial Elvis Presley Web site went online sometime in the Spring of '94 and was adopted by the lovely people of SunSITE/Metalab/ibiblio sometime in the Summer of '94. As far as I know, it was the first Elvis site online and the first listed on Yahoo!, way back when. Due to my lack of work on the site, it's also a relic of old-school Web design... [36]

Herbs and alternative therapies

Henriette's herbal page http://www.ibiblio.org/herbmed is one of the oldest herb sites on the web - it's been online since July 1995. The beginnings were small -the main page, the herbfaqs, and the plant name database. Later I added perhaps 50 pictures, and still later the first monthly archive files.
Nowadays it's 32+ MB of herbal .html files, most of which are classic texts, 2900+ herbal pictures, continuous archives from 1992 (alt.folklore.herbs) or 1994 (the medicinal herb - mailing list), popular links, the faqs, the plant names, articles... lots.
When we first got stats my hits could be counted in the hundreds, per year. These days my stats show about 44000 hits per day, on average. [37]

Chapter 6 - Bringing Order to the Web and to SunSITE

Early attempts at Organizing the Web – Treasure Hunts

In the years 1992 and 1993, as the Internet grew, rumblings began to be heard – too much information, too difficult to locate what you needed. In a mood of whimsy, Internet treasure hunts were launched. The first Internet Treasure Hunt was sent out by email on August 31, 1992 to the lib-ref (library reference) listserv and several other listserv groups. The framework of the Hunt had participants use all of the available search technologies to solve information retrieval puzzles, thereby building their skills and becoming better Internet information searchers.

The hunt lead searchers to the SunSITE presidential archives: question "6. [6] Bill Clinton made a speech somewhere on Earth day this year. Where can I find the text of it?" [38] The correct answer: [6] Via WAIS (quake.think.com), I searched Clinton speeches at sunsite.unc.edu. - clinton-speechess. Bill Clinton spoke at Drexel University in Philadelphia, PA on Earth Day in April 22, 1992." [39] The Internet Treasure Hunt was novel and humorous, but it pointed out the need to find a way to create a globally accessible, platform compatible data resource.

Organizing the Internet Through Index Trees

Tim Berners-Lee and CERN attempted a remedy to the situation by creating three index-trees of sites on the web. On the CERN homepage was a series of links to major resources. There was one link entitled "other subjects", which linked to a master page of sites classified by subject. "From that I have links to individual servers of all kinds (W3, WAIS and Gopher), and in cases where there are a lot like physics and biology, a link to a page about one specific subject. In this way you can browse the web by subject like a library." Berners-Lee commented. [40]

The other index-trees were servers sorted by organization or server-type. On these, Berners-Lee commented,

> "The list by server type is easy, because the people responsible for each protocol keep a list of the servers using it. That is, there is a tree of gophers, and there is an index of

WAIS indexes. There is the W3/WAIS/Archie server for FTP sites. This tree isn't so useful unless you know what sort of a server you are looking for, but it tends to be more up-to-date than the subject index. It also has things in which aren't just about subjects. The third tree was going to be a geographic tree of organizations, but that isn't at all up-to-date." [40]

Finding Information in the SunSITE Archive – Automated INFO Response

Upon logging into the SunSITE archive, users could access a documented entitled "Automated INFO response for SunSITE".(see appendix 6-A) This document, when printed, came to 15 pages. The title of the first section was "a Short Guide to Anonymous ftp on SunSITEsite.unc.edu 152.2.22.81" It was a full users manual for entering the archive, moving around between the directories, using the WAIS databases, and submitting software, documents, datasets or other digital objects to be archived. It assumed a strong comfort level with UNIX and command line interfaces.

SunSITE's First Hyperlinked Homepage

Berners-Lee set up a homepage for UNC which was SunSITE's first hyperlinked homepage on the World Wide Web.
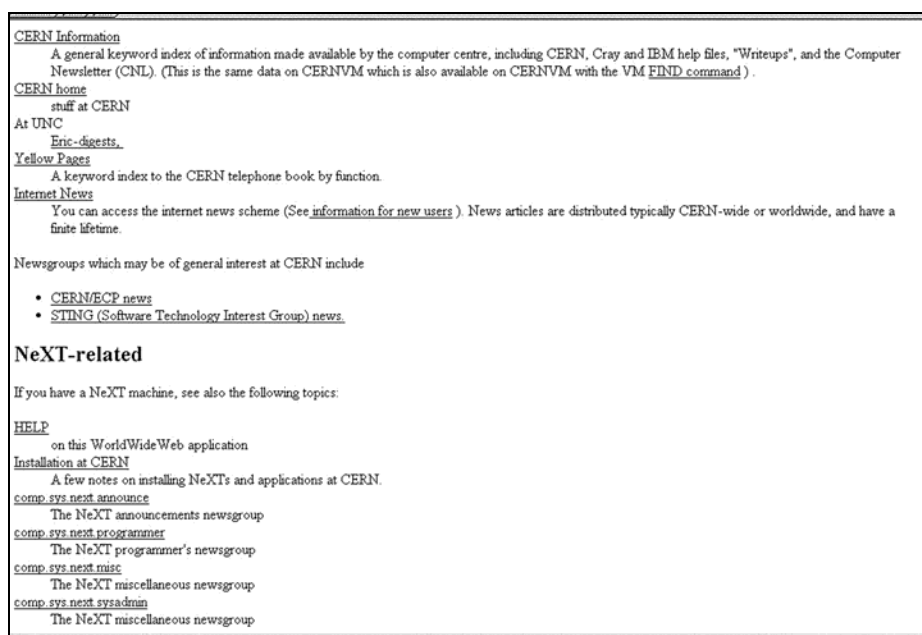

Figure 6-1   First SunSITE homepage circa 1992-93

All of the methods of organization seem simplistic and naïve by today's standards, almost a decade later. However they are significant since they were the beginnings of what would grow into the indexes, search engines and metadata in the next iteration of the Internet.

It is worth noting that many of the problems that users were encountering in 1992 and 1993 on the Internet were extremely similar to the issues that had been observed in the 1980s in the development of freestanding or limited hypertext systems. Theses problems had been documented by Jakob Nielsen in his *Hypertext'89 Trip Report* [41] and Tim Berners-Lee observed in his notes for *ECHT 1990* [42]: linked information systems were exciting, dynamic space, but the paths could not be predicted. Even though many users were navigating using WAIS, gopher or archie, they were experiencing the problem that Botafogo had dubbed the "lost in hyperspace" [43].

Organization and Chaos at SunSITE

As the web grew at a thunderous rate, people felt overwhelmed by the amount of information and their inability to locate the information that they needed. Once they found the right site, they weren't always able to locate the information within the site.

Within the first year, SunSITE's rapidly expanding collections were becoming unwieldy, as well as difficult to navigate and difficult to retrieve information from. This was not just an issue at SunSITE, but also a problem at virtually every Internet repository. Part of the problem was the diversity of formats used in accessing the information: WAIS, gopher, Veronica, archie, ftp, telnet, etc. There were no standards in 1993 and 1994 that could serve to clarify what information structures and standards would best serve the goals of the Internet Community. In early 1993 two www browsers were released that won acclaim form the Internet community: lynx (a text based browser) and Mosaic (a graphical browser). (CERN had released a browser earlier, in 1991, at almost the same time as gopher was released. But the CERN browser never gained popularity.)   It wasn't until March of 1995 that WWW browsing would surpass ftp-data as the service with greatest traffic on NSFNet, based on packet count. [44]

Annotated Lists on the Web

In March of 1993, Mosaic began an annotated WWW 'what's new' list. The list was a combination of announcements of new web servers that had been added to the network, usually announced at the rate of one or two per day. In a way, this was an adaptation of the gopher information lists, but the descriptions could exceed the 70 character Gopher limitation.

 1994-95 were years of innovation in improving resource location on the web and examining web sites for usability. In April of 1994, two resources emerged to aid in resource discovery on the Internet: Web Crawler and Yahoo.

Yahoo – What's Cool | What's New |

Initially it was reported that Yahoo was an acronym for "Yet Another Hierarchical Officious Oracle". In recent years this version of the name's origin has been denied, but the story persists. What is known is that David Filo and Jerry Yang, Ph.D. candidates in Electrical Engineering at Stanford University, started their guide as a way to keep track of their personal interests on the Internet. In April of 1994, the index was released to the general public. Yahoo was a time intensive, hand picked index of what's new and what's best on the internet. Soon the Stanford network was overwhelmed by the bandwidth usage caused by Yahoo. Filo and Yang in 1995 moved their files over to the larger servers housed at Netscape.

Figure 6–2   Yahoo's homepage in 1994

The screenshot (above) of the Yahoo homepage in 1994 [45] shows a mostly plain text interface with 19 major categories and 23,386 entries. Across the top, are the links that defined the early phase of Internet classification schemes: *What's New? | What's Cool? | What's Popular?*

Webcrawler – Automatic Indexing

WebCrawler, the first search robot, was made available on the Web in April 1994, the same month as Yahoo. WebCrawler's original database contained pages from just over 6000 different Web sites. [46] When Pinkerton presented his paper on web searching at Second International WWW Conference in 1994, the index had grown substantially. "The index covers nearly 50,000 documents distributed on over 9000 different servers, answers over 6000 queries per day and is updated weekly." [47] Pinkerton describes the WebCrawler operations as follows:

> The WebCrawler is a tool that solves these problems by indexing and automatically navigating the Web. It creates an index by an incomplete breadth-first traversal, then relies on an automatic navigation mechanism to fill in the holes. The WebCrawler indexes both document titles and document content using a vector space model. Users can issue queries directly to the pre-computed index or to a search program that explores new documents in real time. The database the WebCrawler builds is available through a search page on the Web. Experience with this index suggests that the WebCrawler is collecting enough information from the Web, and that while the current query interface is adequate, more sophisticated retrieval tools are necessary. Real-time searches, although

not available to the public, are fast and seem well targeted towards what what the user wants. [47]

WebCrawler did not use the *What's New? | What's Cool? | What's Popular?* format, but relied on searching, indexing and the ability of users to actively seek out content.

Creating Usability on the Web

While one group of web users was trying to organize the entire web by making it easier to locate information. Another group was working at establishing guideline for creating and organizing websites so that the information contained within them could be located.

Heuristics

Jakob Nielsen in the 1980s had begun a system of usability analysis for hypertext. In 1990, Nielsen created heuristics, the a standard for what he labeled discount usability studies. By 1994, Nielsen had refined this set of heuristics to be:

- Visibility of system status
- Match between system and the real world
- User control and freedom
- Consistency and standards
- Error prevention
- Recognition rather than recall memory
- Flexibility and efficiency of use
- Aesthetic and minimalist design
- Helping users recognize, diagnose and recover from errors [48]

That same year, Nielsen began applying these hypertext heuristics to the Web. One of Nielsen's earliest web heuristics investigations was a critical usability analysis of the SUN microcomputer homepage (see Figure 6-3). In 1994 and 1995, the SUN Microsystems web site and SunSITE shared similar icons and outlook. The review was quite critical, but at that time, Jakob Nielsen lacked the authority to direct standards of web usability and accessibility, as he would in later years.

Figure 6-3 Sun home page from Nielsen, 1994

The heuristic responses to the SUN homepage were:

> One user commented that the Sun home page "contains to much stuff that I will just ignore anything that is not clear." As an example of something that was not clear and would be ignored, this user pointed to the SunSITE entry… On the Sun home page, all users clicked on the Santa Claus icon first due to its prominent position at the top left corner of the page. Also, several users expressed a liking for the untraditional icon design with the button graphically protruding from the image. All users overlooked the row of flat buttons on top of the Santa button. Several users were actively looking for a "new at Sun" button and a search feature but did not find them. The design used for the Sun home page makes the button bar look too much like captions or headers for the top row of much more prominent icons. [49]

Defining the Digital Library

Beyond the confusion about how to best organize and display information on the Web, SunSITE was faced with a lack of clarity as to what constituted a digital library. At the First Annual Conference on the Theory and Practice of Digital Libraries (June 1994), the conference's opening address began:

> "At this time, no one knows what a "real" digital library is, or what a concise definition would be, or what the boundaries of the field are. We do know that one must have a very long term commitment and that this field will take decades to mature. We must not reinvent prior research results from constituent sub-fields. This places a heavy burden upon us because digital libraries research requires the integration of results from many

diverse fields. We must conduct broad-based research set in a clear understanding of its context." [50]

Throughout this period of early web organization between 1992 -1995, SunSITE's organizational structure remained somewhat static. The main effort at SunSITE was focused on maintaining and strengthening the collections and WIAS archive accessibility. It wasn't until 1995, when web browsing and the image became standard fare on the Internet that SunSITE began to seriously investigate homepage design and web indexing practices.

Chapter 7 – Evolution of the SunSITE Index and Homepage

SunSITE's homepages, from their inception until 1993, were text-based lists, with few images, similar to the original homepage set up by Tim Berners-Lee. The assumption in designing the SunSITE homepage was that most visitors were using WAIS or other telnet/ftp technologies. Therefore, images were superfluous. The information was organized in a list of major categories, with a short description and a hyperlink. Most early web pages tended to be long documents that might fill three or four screens of information.

Figure 7-1 SunSITE homepage – screenshot of lower half - circa 1993-94 [51]

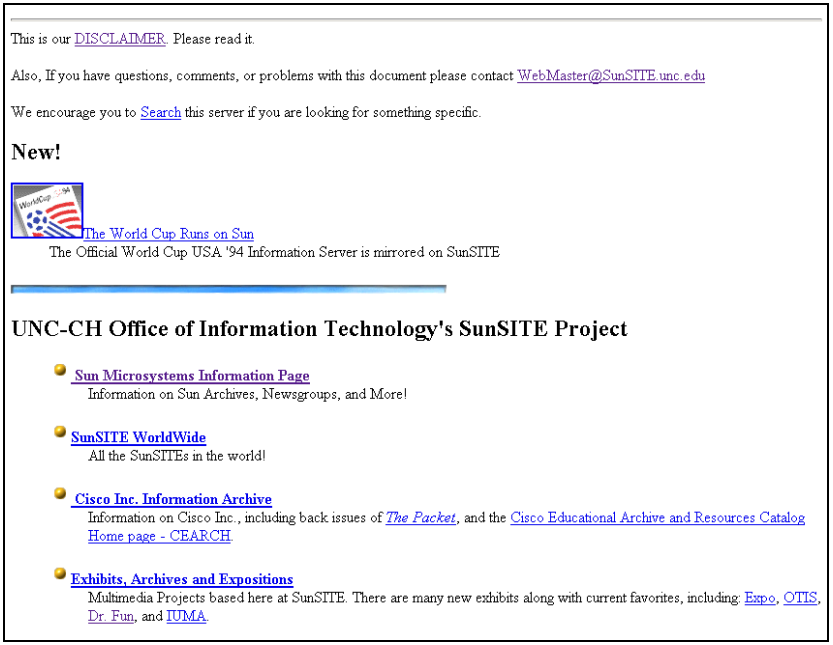**Figure 7-2 SunSITE home page 1994-1995 – screenshot [52]**

Figure 7-3 SunSITE home page 1994-1995 – screenshot scrolling down [52]

What's New | What's Cool

Design of SunSITE's homepages in 1995 – early 1996 remained similar to the preceding years, except that the location of items was flipped. Images were located on the initial screen, with text lists on the subsequent screens. "What's New" lists that were fairly ubiquitous on the web in that time began to appear on the SunSITE homepages.
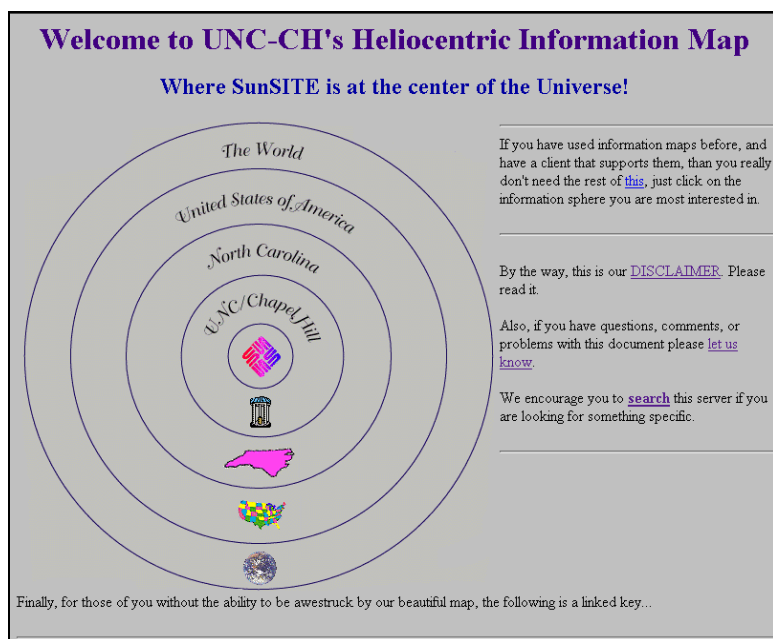

Figure 7-5   View of the World from 9th Avenue

Figure 7-4   Universe Heliocentric View of the Universe

The index pages, in contrast to the homepages, began to develop in new experimental directions. One page offered a humorous view of the internet world. It echoed Saul Steinberg's famous1975 cartoon, "View of the World from 9th Avenue", where New Yorkers see the whole world as centered around Manhattan. In the SunSITE home page, the title was "Welcome to UNC-CH's Heliocentric Information Map - Where SunSITE is at the center of the Universe!" The organization was a geographically based map with a SUN server at the center of the universe and everything spreading out from there. Below the heliocentric map was a table with text, which explained:

> "In the center of the Infoverse, of course, is SunSITE: Specializing in Sun and Cisco information, as well as any other useful or nifty things we could find to put on there. The next sphere beyond that is The University of North Carolina at Chapel Hill. Our physical location, our home, the place we lay our weary heads (other than the keyboards themselves). Beyond that lies the great state, North Carolina itself. Many interesting things are going on in the Tar Heel State, be sure to check them out.
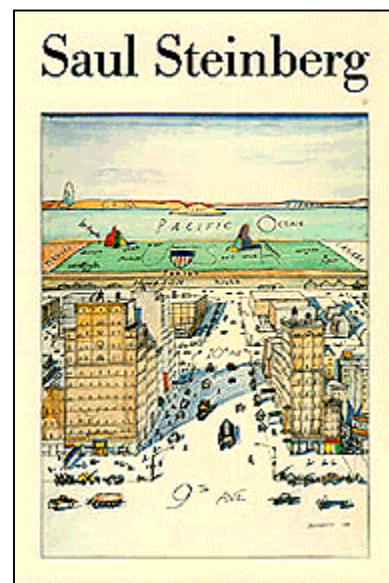
Further out along our Heliocentric Information Map lie all of the resources available in the other 49 states. Here also lie the famous, fabulous SunSITE Federal Government Exhibits, and other documents of import to the nation as a whole.
Last but not least, the final orbit carries us well beyond our egocentric info-view -- indeed, to the outer limits of the mind. Here we point to everything else in the world that is not in the United States.
If you have something cool that does not belong ANYWHERE else, this is where it goes. Also, if you are looking for something to spring you out of your staid, limited existence, this is the place to look. Here, the SunSITE team really does valiantly attempt to bring you the world. Thanks for stopping by!" [53]

Collaborative Database Indexing

One of the goals of SunSITE was to create a digital library that could be completely run by the collaborator/users in so far as it was possible. A first step toward taking this collection of web sites and turning it into a useful library was to create a method by which the site creators could supply the indexing information, much in the same way that the Linux software creators and maintainers did with the Linux Software Map (LSM).

Pratik Patel created a flat file database to be used for indexing the SunSITE collection. There was interest in the general web community as to how web pages could be generated interactively. Patel built the database, wrote the script and created an input form. Information requested was much simpler and less structured than the IAFA data categories or the categories requested in the LSM.

There were five primary categories: Computer Related, Online Publication, Politics & Society, Science and Health, Fun & Games. Contributors were asked to select a subject area for archiving and provide such basic information as title, site description and URL. The database was never well accepted by SunSITE's contributors. Few people entered data for their sites. Many sites that were indexed had index problems due to index data being incorrectly entered and therefore unable to be retrieved from the database. The database index was soon abandoned in favor of manual indexing of the sites, which would be maintained by one of the SunSITE staff. This was a labor intensive effort, but it successful for a period.

Figure 7-6   1996 Collection Database Entry Form

SunSITE's Alphabetic index

The database index was soon abandoned in favor of manual indexing. This index was similar in structure and data to the existing WAIS listings. Sites were hyperlinked by name and then a short description was provided of what might be found on the site. There was no attempt to collect or to provide metadata, nor was any effort made to review or apply qualitative judgments on the site's materials. This index was the standard for the next four years, with sites being added as SunSITE grew exponentially.

Figure 7-7   SunSITE's Alphabetic Index, circa 1996

The alphabetic index was a straightforward text-based index with tiny button icons as the only images. Each site was listed by title, with a short description of the focus of the site. Along with this alphabetic index, the home page gained a new look that combined graphical icons and short text into a browsable index that would appear primarily on the first screen viewed by the user.

1996 SunSITE Homepage: Images and Short Titles

The 1996 version of the homepage had a greater usage of images and integrated the images more successfully. The 'header' graphic integrated information about SunSITE, UNC-CH, SUN Microsystems and the winning of the NII award, all into one coherent image. Homepage section titles became more succinctly. There were ten areas selected as major areas and a link went to each: Sun Information, Free Access to Software, Government, Hypertexts, Cisco Education Archive, Weather Server, Multimedia Exhibits, Search, Heliocentric Info Map, Alphabetic Index, SunSITE Statistics. There was still the *What's New* list, as well as a link to the full alphabetic content of the digital archive.
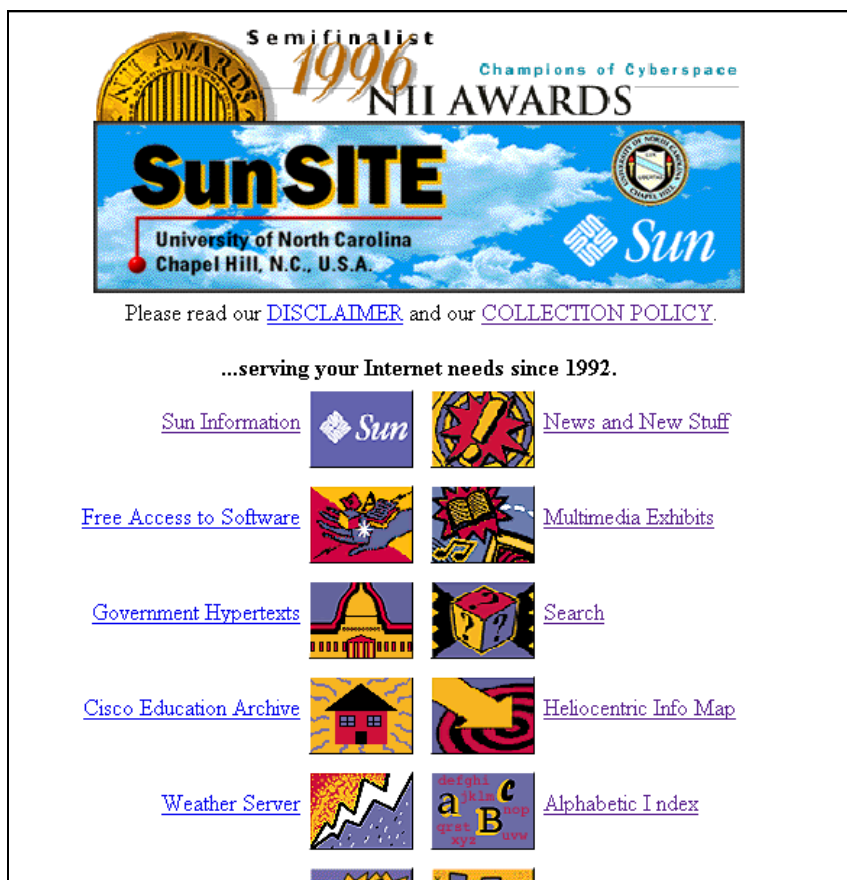
Figure 7-8   1996 SunSITE homepage

Jones Leaves and Returns + SunSITE in Transition - > Metalab

Over the years, Judd Knott had taken a lesser role in the managing of the SunSITE project. In 1995, SunSITE was headed almost entirely by Paul Jones, with a rotating staff of student assistants. In 1996, Jones accepted a job as Technical Director for the Institute for Advanced Technology in the Humanities at the University of Virginia in Charlottesville. Knott assumed the daily operation of SunSITE, with the help of another OIT employee, Mary Overby, and a full staff of graduate students.

When Paul Jones returned from Virginia in 1997, he stepped in as the leader of SunSITE. During his absence there had been some changes in the funding structures. SUN Microsystems was no longer the sole (or even a primary) funding source for SunSITE. The archive had expanded far beyond the original goal of being a software

repository for SUN. The Linux archive was growing rapidly. Linux, in a way, was in direct competition with SUN. Linux provided a free operating system, which could perform most of the same tasks that previously been proprietary to SUN. SunSITE had been such a success for SUN Microsystems that it had replicated the project around the world. By 1996, there were more than 20 SunSITEs in operation globally [21]. The UNC-CH SunSITE was flourishing, "SunSITE at UNC currently averages about half-a-million information requests a day and holds over 47 billion bytes of information and software for free distribution via the Internet.[21]"

Ultimately, it was decided to distance the archive from the SUN brand and change the name of the project to MetaLab (metalab.unc.edu). The transition to MetaLab was begun in May of 1997. An announcement from Academic Technology & Networks of UNC announces the receipt of a grant and uses the new name, MetaLab:

> "UNC-CH's MetaLab was recently awarded an Academic Equipment Grant from Sun Microcomputer Systems in an international competition for excellence in information science research. The gift of over $170,000 worth of the latest Sun server technology, including an Enterprise Server 4000 with four 250Mhz processors and a gigabyte of RAM, will be used to update and improve the popular SunSITE program at UNC-CH….
>
> The current UNC-CH server, SunSITE.unc.edu, handles over a million requests every day for information of all kinds, from free software to information on herbal medicine, electronic music, and the history of the American South. References to SunSITE.unc.edu collections have appeared in many publications, including Education World, JavaWorld, Salon, and the New York Times, as well as in Ann Landers' syndicated column." [54]

SunSITE becomes MetaLab

Almost 18 months elapsed before the changeover from SunSITE to MetaLab would become official. On November 4, 1998, Jones sent a notice out to the SunSITE contributors entitled, 'The End of SunSITE." The first part of the message was informing SunSITE contributors what the change would mean to them. The second part was a forwarded copy of the message that Jones had sent to SUN regretfully terminating the alliance.

> On Wednesday afternoon (EST), I sent the message below to folks we have dealt with at Sun (those that still work for Sun). The point is that Sun is no longer a major contributor to sunsite.unc.edu and that for us to be able to serve you and our clients (aka the Net) we will need to seek new partners.
>
> To do that means that we will need to shed the sunsite.unc.edu name and pick up with a new vendor-neutral name, MetaLab.unc.edu. The plan is to

begin this transition on December 1. http://metalab.unc.edu/ is already in
place and working. Suggestions, Comments, etc are welcome. [55]

In a message sent to SUN four days prior to the general announcement, Jones
was more explicit about the financial issues that were at the heart of the name change:

> In 1992 Sun and the University of North Carolina at Chapel Hill created
> the first SunSITE (SunSITE.unc.edu). Over the years, that partnership has
> served both parties and the Internet community exceedingly well. As a
> measure of that project's success, SuSITE.unc.edu has appeared in over
> 125 major media stories in the past 12 months; it has served over 40
> million files by httpd in the past month (October 1998) and over 10
> million files by ftpd. SunSITE.unc.edu has served as the model for Sun's
> extension of the SunSITE program to over 60 new sites in the past 5 years.
>
> The cost of the partnership has been borne jointly by UNC and Sun. Sun
> made incremental hardware contributions and some financial support--that
> financial support amounting to less than the cost of a banner ad on any
> site of the same activity level as SunSITE.unc.edu. UNC contributed
> salaries, networking, office space, software and support contracts, and
> supporting hardware. More importantly UNC contributes to the North
> Carolina Educational Internet connections where SunSITE.unc.edu
> constitutes over 1/2 of all .edu traffic leaving the state.
>
> SunSITE.unc.edu has taken a major role in Sun initiatives including work
> at the Mars Landing, with the Nobel Prize, with StarOffice, with Java, and
> with Solaris announcements.
>
> In order to serve our clients better, UNC has secured additional partners
> for the project. This leveraging was encouraged by Sun from the beginning
> and was further encouraged at the most recent SunSITE Managers Meeting.
> Currently SunSITE.unc.edu's partners include: Cisco, Real Networks, Red
> Hat, and Educause. Let me stress that these have been minor partners in
> relation to Sun in the project.
>
> However with the new financial year in July, Sun is offering to contribute
> only 1/10 of the amount previously contributed with an offer even lower
> for the following year. This would make Sun by far the most minor
> contributor to the project. Further UNC and Sun have not had a contract
> since 1996.
>
> It is clear that to serve the clients we have developed through our
> partnership we need to revise our relationship. It is in no way fair, to
> our clients and our other contributors for Sun to reap the major benefits
> while making the smallest contribution. And more importantly, we cannot
> continue to serve our clients with such reduced support.
>
> While we welcome and encourage Sun's continued participation in the
> project, UNC must seek another partner for the project immediately.
>
> Beginning December 1, 1998, UNC will begin the process of transitioning
> from SunSITE.unc.edu to MetaLab.unc.edu as described in our older contract
> with Sun.

> There is no joy on our part in making this decision, but it seems to be the only option that we can take in order to continue to serve our clients, be fair to our partners, and to be even handed in our dealings with Sun. [55] (see appendix 7-1)

The name change was marked by home and index page redesigns that featured an animated gif that displayed the name SunSITE peeling off to reveal the new MetaLab name beneath.

Figure 7-9   1998 Index Page with Peeling Logo

Chapter 8 – Diversification and Exploring New Technologies

The late 1990s saw a surge in multi-media experimentation at MetaLab. Multimedia had been a tantalizing possibility for Internet publishing. For photographs, drawings and other still images, compression technologies had developed quickly and web usage had become a reality fairly early on. Audio and video were more difficult because of the larger file sizes and because of the difficulty of producing compression methods that did not result in unacceptable levels of distortion. Apple had introduced hypermedia projects using HyperCard in 1997. Hyperlinked multimedia games, such as Manhole, Cosmic Osmo and Myst existed for stand-alone computers. The next level of challenge was to bring full multimedia to the web.

Mbone – Tunneling Through the Internet

Mbone was the first attempt to deliver music through the Internet. It used a protocol that could 'tunnel' through the IP protocol and thereby provide distribution for real-time audio, video, and file transfer. With mbone, "distribution occurs through the use of multicasting technology. Unicasting is the normal way we use the Internet. One host sends a packet of information to another host. Multicasting occurs when one host sends a packet of information to a group of hosts. The packet is only replicated when it hits a designated branch point, in other words, someone tuned in to receive the packet." 56

The first mbone broadcast occurred in March of 1992, it was an audio broadcast of the IETF meeting. On November 18, 1994 a concert of the Rolling Stones was broadcast live from Dallas. During the UNC Bicentennial in 1994, SunSITE organized an MBONE broadcast of the visit of President Clinton to UNC for the celebration, which allowed UNC Alumni and other interested parties to participate in the events from anywhere in the world. But these remained landmarks for a largely unaccepted technology. In 1996, Richard Dean, producer of New Media Services at National Public Radio explained the problem with mbone cybercasting (as putting radio over the internet was called):

> "Here's my problem with the MBONE: Nobody's on it. I think MBONE is a really neat technology, but it doesn't reach who I want to reach; it reaches a bunch of guys with

UNIX boxes. With NPR, our goal is to get our content to the end user. That end user is using a dial-up Internet account, he said. RealAudio technology may not be the ultimate in Internet multimedia, but for NPR, it's the best of the technologies available today. "I look at things two ways one is what's now, what can I do today, and then, where is the technology going?
As faster connections become available to the masses, "the MBONE will become something more important than it is right now. But I need to reach listeners now, not in years. He equates today's audio technology to the hand-crank phone: "It may not be the right solution in the long run, but in the short run it is. People don't want to wait. When the technology reaches the next level, you jump to the next level." [57]

Despite problematic technology, SunSITE staffers had a strong interest in cybercasting. Mark McCarthy created a resource repository on cybercasting technologies Mbone. Also SunSITE hosted the archives of several Usenet groups devoted to exploring the applications for mbone and other cybercasting technologies. But there was no actual mbone broadcasting done at SunSITE.

Internet Talk Radio

One of the earliest Internet audio broadcasting experiments at SunSITE was Internet Talk Radio (ITR). This was a very different approach from what the mbone was attempting. ITR was an attempt to distribute broadcast file using the existing technologies and minimum bandwidth. Essentially Internet Talk radio consisted of a regularly updated set of digital recordings that were encoded using SUN's .au format. These recordings were distributed over the Internet and could be retrieved by ftp. They were then played back on the listener's computer.

The Internet Talk Radio README describes the project this way:

We're the first station in cyberspace, a source of news and information for the Internet community. We run two channels: Internet Talk Radio is a science and technology channel and the Internet Town Hall is devoted to public affairs….

"Station in cyberspace"? Get real!

No, really. We run a "radio" station, publishing sound files which you listen to on your personal computer. You know that funny sound your Mac makes when you start it up, quacking like a duck or talking like Bart Simpson? Think of our programs as a very long version of that duck…

Our flagship show on Internet Talk Radio is "Geek of the Week" featuring in-depth interviews with members of the technical community. We've talked to all sorts of famous engineers about topics such as the next generation of TCP/IP, resource discovery protocols, network security, and how to put toasters on the Internet. [58]

IUMA and other Music Distribution Archives

In 1993 and 1994 SunSITE supported a number of experimental multimedia groups including the Internet Underground Music Archive, N.C.Meme (an acronym for North Carolina Multimedia Exhibition of Musicians and Entertainers), the SUN sound archives (a collection of miscellaneous sounds), and the Chinese Music archive. Internet Underground Music Archive (IUMA) was essentially a computer bulletin board that functioned as an alternative to existing methods of music distribution. Two students at University of California Santa Cruz began IUMA in 1993. They were in a band and were looking for a new way to distribute their music. In 1995, Rolling Stone Magazine ran an article on IUMA, explaining the new technology and how the creators see Internet distribution

> "as the start of an evolutionary change that will eventually see full CDs fed over fiber-optic lines to home audio systems via computers. Geffen Records is already preparing to offer a full Aerosmith song on line. This type of music distribution is part of everyone's future," says Jeff Gold, a senior VP at Warner Bros. Records. All IUMA requires is access to the Internet and a sound card, an inexpensive piece of equipment built into most Macintosh computers, After downloading, users can transfer the tunes to an audiotape and play them on a home stereo." [59]

WXYC – First Live 24 hour Radio on the Internet

David McConville was the SunSITE employee who was a driving force behind much of the sound and video experiments that occurred at SunSITE. He recalls his time at SunSITE:

> "I was fortunate enough to meet Paul and Doug Matthews (founder of alt.music.chapel-hill) shortly after my arrival there (Chapel Hill – ed). This was a dream for me, since I'd specifically moved to Chapel Hill to learn more about audio/video compression and distribution technologies, hoping to find a way to distribute audio/video files quickly and easily around the globe.

> While I was there, my official job (with Doug) was to maintain the Cisco Educational Archives (CEARCH), but I ended up focusing more of my time on trying to figure out how to get interesting audio/video material online. Since we had a lot of bandwidth, we were able to run a CU-SeeMe reflector that we used to mirror many live events (including Burning Man). I also liked to check out various happenings on CU-SeeMe and the MBONE (the multimedia backbone - the very early high bandwidth audio/video network inside of the Internet) at all times of the day and night. I managed to get on Japanese TV by being projected onto a wall at a big rave while teleconferencing and witness the IRA calling a cease-fire, even though no news cameras were allowed to film the event (another CU-SeeMe user just pointed his camera out of his windows to let me see it)."

In 1994 SunSITE began experimenting with digital re-broadcasting of radio. A new Internet audio/video broadcasting technology called CUseeMe was being developed at Cornell University. Cornell's goal in developing the CUseeMe technology was to create a web based collaborative work environment that would allow real time conferencing across the Internet. However, experimentation at SunSITE with the CUseeMe technology resulted in a technology that allowed live radio signals to be broadcast thorough the Internet, at a low bandwidth demand. SunSITE applied this discovery to digitally re-broadcast the live signal of WXYC, the University's student-run radio station, in real-time over the Internet. On October 12, 1994, SunSITE and WXYC radio became the first 24-hour live streaming radio station on the Internet.

Publishing - Unix Server Book

In 1996, three of SunSITE's team, Jonathan Magid, Paul Jones and Michael J. Hammel collaborated to create a book entitled, *The UNIX Web Server Book*. The goal was to create in a single source the software and the documentation that would enable individuals and grassroots organizations to be able to publish on the Web. It also came with a CD that supplied all of the necessary software. The book, published by Ventana Press was so popular that it was reprinted two years after the original edition, under the title *The Web Server Book.* Linux Journal reviewed the book shortly after its publication and commented:

> "Setting up a web server is not a difficult task (given a stable Internet site), but it is one that requires a unique set of tools. The Unix Web Server Book was written to allow someone with little or no experience with the Web to set up a server. This is perhaps something of an anomaly, since anyone capable of administering a Unix system is likely to have enough basic knowledge to configure a web server using software documentation. Still, there is a necessary background that this book attempts to fill.
>
> The Unix Web Server Book assumes no prior knowledge of the World Wide Web and so begins with a brief history. Chapter two, ``The Basic Pieces'', is summed up by the chapter legend ``What is this http:// stuff?''. Chapter three compares servers and recommends Apache, a freeware server available for many platforms. An introduction to web security follows. In the next ten chapters, introductions are given to the elements of developing web pages, including HTML, importing documents, images, multimedia, searching, forms, advanced CGI, Java and HTML checkers. Finally, ``Fitting in: Joining the Virtual Community'' finishes up the book. It's all at a pretty basic level, but the sum is a hefty book with voluminous references to on-line resources. " [60]

Documenting the American South

The UNC Academic Affairs Library in 1998 received a National Digital Library award from a competition sponsored jointly by the Library of Congress and Ameritech to digitize existing historic, first person narratives and other primary source materials documenting the cultural history of the American South. The digitized documents were then compiled to create a digital archive, "Documenting the American South: The Southern Experience in 19th-Century America", "First-Person Narratives of the American South, 1860-1920".   SunSITE's role in the project was to provide them with full Internet services, a stable storage space and database support. The original documents were text and graphics from the original books and pamphlets. Later archives encompassed the musical collections of UNC's Southern Archives, such as the original Goldband recordings, a Lake Charles, Louisiana record company active in the 1940s and 1950s.

'DocSouth' as SunSITE's staff refers to the project was an interesting project for the scale of the database that it required. The initial part of the archive included more than 18,000 pages of information. Additionally it was a project that allowed some development with the university library system and strengthened SunSITE's role as a digital library of academic interest.

Digital Video and Education

In 1998MetaLab began a collaboration to produce digital video, in conjunction with the UNC School of Journalism. Todd Stabley, a graduate student in English, was pursuing a study of the films of Italian filmmaker Michelangelo Antonioni. Instead of writing pages of analysis, Stabley wanted to create tools that would allow students to flip through scenes in a movie in a manner similar to flipping through pages in a book. Stabley received a $29,000 chancellor's grant to experiment with the technology. Working with Metalab, Stabley created an "archive of film stills and sequences complete with sound so that a student can analyze particular scenes and incidents without the distraction of fast-forwarding, rewinding or pausing."[61] The ultimate goal of the digital video experiment was to create a new way of doing film research in the English department. In following semesters, the students in Stabley's section of English 42

wrote on-line term papers. "Instead of describing scenes, as Stabley had to do in his thesis, they illustrated their papers with film frames and sequences." [61]

Streaming Video

In 1999 Digital Bitcasting Corp. and MetaLab formed a research partnership to explore uses of the international standard MP3 audio and MPEG video formats. The initial project was to stream MP3 audio at the Folkden, the Web page of folk-rock artist Roger McGuinn. The technology that made this possible was the development of CD quality audio through the use of RealSystem technology. Prior to the development of RealSystems, "Internet audiophiles had to choose between a time-consuming download of high-quality audio or streaming lower-quality audio. Bitcasting's RealSystems G2 upgrade eliminates that trade-off, streaming high-quality MP3 audio and MPEG video directly to RealPlayer G2 users." [62] The long term goal was to utilize Bitcasting's technology in distance education applications at the UNC-CH School of Public Health. [62]. This technology represented a break in tradition for MetaLab.

Streaming Radio Stations

The RealStreaming software permitted MetaLab to experiment into the potential of live broadcasting on the Internet. In 1999, when there were hearings about the impeachment of President Clinton, MetaLab began live Internet broadcasts of WUNC's National Public Radio coverage of the hearings. In a newspaper article, Jones explained the process: "We set up a computer that takes the on-air feed from the satellite to convert it to a digital format… Up in Washington, they produce a radio program along with live camera in the U.S. Senate chambers. NPR squirts it to its satellite. NPR stations, like WUNC receive it. From their receiving, we take it into our computer, and we convert it into a digital format. Anyone who wants to hear the impeachment can go to our home page or WUNC's home page and listen." [63] This was the beginning of a permanent role of broadcasting WUNC's signal via the Internet that would last long after the hearings had ended. Between 2000 and 2002, MetaLab added two more radio stations to their streaming rosters: WCPE, a 24 hour classical station and WXDU, Duke University's station.

Chapter 9 – Organization and Transformation: MetaLab -> Ibiblio

In August of 2000, MetaLab received confirmation of a $4 million donation from the Red Hat Center, a philanthropic program created by Bob Young and Marc Ewing, co-founders[1] of Red Hat Linux software. The $4 million would be given to MetaLab over four years to employ more people and expand its bandwidth. [64.] In conjunction with this donation, MetaLab received a new name, ibiblio. *Biblio*, taken from the ancient Greek word for *book*, served to reinforce the electronic library nature of the archives.

Both Ewing and Young applauded the openness and the eclectic nature of the MetaLab archive. ""ibiblio is to the Internet what the public library is to Barnes and Noble." Young said he and fellow Red Hat co-founder Mark Ewing set up the non-profit Red Hat Center last year to spread the free flow of ideas. Red Hat packages and sells its version of software products based on the Linux operating system, whose computer code is freely shared by users worldwide." [64] As part of the site renovation, Paul Jones articulated a vision that encompassed the past of SunSITE - MetaLab and expressed the future of the ibiblio collections:

> Home to one of the largest "collections of collections" on the Internet, ibiblio.org is a conservancy of freely available information, including software, music, literature, art, history, science, politics, and cultural studies. ibiblio.org is a collaboration of the Center for the Public Domain and The University of North Carolina - Chapel Hill.
>
> It's Not Your Average Library
> The evolving Internet has created new opportunities to share knowledge. Imagine being able to walk into your local library and view, on demand, and without charge, not only every imaginable written text, but also music and poetry archives, African American authors, American history, sports statistics, philosophy of religion, Italian literature, large text database projects, software archives, and more.
>
> Then imagine, in addition to being able to view the collection, you have the opportunity to critique it, expand it, or to create and manage a new collection in your own area of interest. While such a feat is physically and fiscally beyond the scope of even the largest and most extraordinary physical library, ibiblio.org achieves just such breadth and depth on the Internet.
>
> Users in China studying American poetry or folk music can access ibiblio.org and find a vast collection of poems or songs. They can listen to the artist perform, read the biography, download sheet music, reference secondary criticism, and submit their own research papers to the collection. A veteran of World War II from anywhere in the world

can access and contribute to the enthusiast-managed Pearl Harbor archives, joining the virtual dialogue on the topic.

The average municipal public library receives a few hundred visitors a week. ibiblio.org averages 3 million information requests per day, and the contributor-maintained collections are expanding daily. A free and vibrant exchange of ideas among a large community of contributors who share their knowledge across disciplines, ibiblio.org uses the open source model to encourage users to help shape the way information is managed and accessed in the 21st century. [65]

The goals of ibiblio were to remain unchanged, but it was anticipated that the collaboration would strengthen ibiblio's abilities to:

- Expand and improve the distribution of open source software;
- Continue UNC's programs to develop an on-line library and archive;
- Host and foster projects that expand the concepts of transparency and openness into new areas;
- Create, expand, improve, publish, and distribute research on the open source communities;
- Expand and improve the creation of and distribution of open source software and documentation;
- Serve as a model for other open source projects. [65]

Changes and Time Constraints

Informal confirmation of the donation was received only about three weeks before the formal announcement of the changeover to the new name of ibiblio and the launch of the ibiblio website. Additionally, the Red Hat Center requested that certain changes be accommodated upon receipt of the funding. Most of these were requests to improve the organization and navigation of the MetaLab collections. A key point was that a full operational index needed to implemented immediately.

Index considerations

Some of the discussions centered on whether MetaLab was an archive or a library and what implications there might be for each category. If the collection was a library, should library organizational systems, such as the Dewey Decimal System be used? If the collection was an archive, could a computer standard be applied? Most of the ibiblio staff had a strong background in computer science and/or journalism.   The ibiblio office space was housed within the building for the School of Information and Library Science. Paul Jones taught at UNC in both the Journalism School and the School of Information and Library Science. There was some reluctance from the ibiblio staff to accept the title 'library'. 'Archive' held images of dusty, mysterious places with

intriguing documents. 'Library' evoked images of sour-faced women with gray hair buns and overdue stamps. 'Archive' seemed flexible; 'library' seemed rigid and rule bound.

Much of the direction in the initial stages was dictated by the sheer size of MetaLab's collections. The collections index had been poorly maintained, over the years, with only a percentage of the new collections having been added to the index. Within MetaLab's 'web' archive (/public/html), there were around 900 directories.   The collection index listed only about 160 projects the ibiblio indexing effort. Also, there was the anticipation that with additional funding to increase hardware, staff and bandwidth, more collections would be housed on ibiblio.   This forecast proved to be correct. Within 18 months, web directories increased to 1,139 directories, which contained a total of 1,112,503 files.

Audience definition

MetaLab handled an average of 1.5 million transactions daily. Not all of these were in the 'web' section of the collections. Much of this traffic was visiting the Linux archives, Linux documentation project, EcolandTech, and other projects that were accessible by html and ftp both. The interest of the audience could not be defined into one clear-cut genre. The collections were spread out over a wide the comprehensive spectrum of interests. Some of the least visible collections often proved to be the most popular. The Chinese music archive (http://www.ibiblio.org/chinese-music/), which was a small simple site constructed as a student's hobby, was consistently in the top ten most visited sites of MetaLab. It's popularity arose from the linking and mirroring of the music by sites in Asia which were bound by political or bandwidth constraints and therefore unable to host traditional and contemporary music on their own sites.

The new ibiblio site had to be flexible enough to accommodate all existing or potential users, regardless of access mode. MetaLab/ibiblio also had Linux and Unix users who still used the lynx browser, which is a non-graphical text web browser. All it requires is a terminal window and a telnet connection to the Internet. Lynx remains a standard program in both Linux and UNIX. While it is not the graphical excitement that most of us associate with the Internet, it is a fast connection.

Existing Content Management:

There are many driving forces behind every library, the same being true of every digital library. The 1995 SunSITE collection policy, which was still in place, remained a primary articulation of the goals of the archive. The first two lines of the collection policy had stated clearly the two dominant forces: "Do the materials further the teaching, research, or public service mission of UNC? Does the collection use technology in innovative and unique ways? Every collection need not be innovative, but it should use

Figure 10-1 Categorical Index

up-to-date technology." [66]

Since 1996, ibiblio had had exuberant, but eclectic growth. The initial alphabetic index, which had been created in 1996 had not been updated regularly and was unusable. It had largely been replaced by the categorical index, which had non-standard categories that were more of a poetic nature than a hierarchical. The top-level categories were: arts & culture, entertainment, education, athletics, moving-on, e publications, computer info, social & political, bon voyage, science & health (see figure 10-1).

Figure 10-2

Second Level Index Page

The second level categories were in alphabetic order, with no sub-categorical grouping(see figure 10-2) This was effective for the most part, since the majority of

pages had 20 or less site titles . Other than putting software into computer info, there were no obvious or stated rationales for the placement of a site within a category. The 'Himachal Pradesh' site, which contained primarily tourist information and some history was placed in 'social-political'; 'bon voyage' contained various global picture archives including Cuba, Russia, Hong Kong and Vietnam (many of these same archives also contained history). Selected sites would be highlighted each month on the homepage. (see figure 10-3).
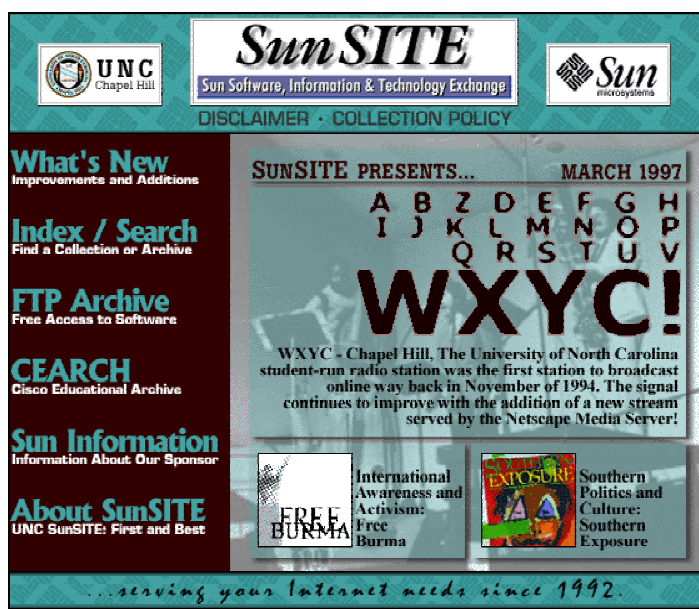
Figure 10-3   Homepage – March 1997

Content Management Vision

Paul Jones, at the time, was often heard uttering the phrase, "The Inmates Running the Asylum". It was the title of a newly released book by Alan Cooper, critical of the software design process. The subtitle was "Why High Tech Products Drive Us Crazy and How to Restore the Sanity." To Jones, however, the inmates running the asylum was a wry observation of how an enthusiastic 'Jeffersonian' democracy should operate – and how the ibiblio archive should also operate.

The process that helped Linux and Open Source software survive against the corporate empires of Microsoft and Apple was the stimulating atmosphere that Jones wanted to engender in all of the collections of SunSITE-MetaLab-ibiblio. It was an atmosphere where collections could be introduced and hopefully those collections would (metaphorically) get used, examined, shaken upside down, and reassembled by the users in new and inventive ways.   Jones wanted the library part of the digital library to remain in the background and allow the contributors to control their works as much as possible.

In an article Jones wrote entitled "Open[Source]ing the Doors for Contributor-Run Digital Libraries", he makes clear the connection between the creativity and collaboration

of Open Source software and that of an open digital library. "The first step in contributor run libraries is to allow people to contribute. This may sound obvious, but many collections try to control or 'gatekeep' from the onset."[67] The ideal for the open source library is that there would be no criteria requirements from the outset, but that a karma system, such as exists in the commentary rating of 'slashdot' or in the book rating/reviewing system of 'amazon.com' would essentially create a hierarchical system that would be more egalitarian than the librarian/critic model. "By giving contributors and readers access to tools for evaluation, ranking and managing the collections, we are not just off-loading work; we are building communities of intellectual discourse. Strong community members are recognized by reputation capital and trust metrics and are rewarded."[67] Then ultimately, the contributors will respond by producing higher quality. "Digital libraries can give back to contributors as well. By sharing collected information, contributors can see which items (manuscripts, songs, and software) are most in demand in the form of top ten lists or most recommended. This enhances not only the referral services, but helps new contributors understand what is considered a 'good' item."[67]

The conclusion of this is that the ultimate goal (though not the immediate step) would be to create an index that would empower the contributors to create and manage all of their own information for their web site. "Our experience with the Linux Software Archive at MetaLab.unc.edu which began in 1992 was that by removing nearly all barriers to submission and instituting instead some simple verification procedures, we were able to accept (and later distribute) some very high quality software with a very low rejection rate."[67] Whatever mechanism was selected would be modeled on the clear, simplistic data in the Linux Software Map. However, with an eye to the future, it should also be compliant with the most common metadata standards of the day, which would be: Dublin core and OAI (Open Archives Initiative) Protocol for Metadata Harvesting.

Ibiblio as a Vision of the Future of Collaborative Digital Libraries

When all of these considerations came together, a clear vision emerged for essential elements, "The goal on the structuring of Ibiblio was to create an accessible digital library. This access should be composed of three key elements: a search engine, an XML based metadata archive (which can be openly accessed for research), and a

browsable index (ever found that really great book while wandering down the library shelves?). To this end, Ibiblio has enlisted google as the search interface. We are presently building the metadata archive. The index is in its formative stages." [68]

Additionally, it was decided that the homepage would serve as an information port to all of the collections at ibiblio. The primary focus of the homepage would be the web collections. The logic behind this was the fact that most Linux developers and users connected to the Linux archive site using a terminal based technology, such as ftp or telnet; hence the homepage would not be their primary destination. The homepage would have pointers to the software archive, and the software archive could maintain a homepage of its own that served that community.

Implementation: Design Principles / Information Architecture

The idealized hope of this re-organization was to use the current theories and technologies from the field of information architecture to create an analysis of the web collections' existing materials then extract hierarchical mappings of the entire archive. This would allow a visualization of what was contained in the nearly 1,000,000 files and how all of the various collections related (or didn't relate) to each other. It was hoped that a diagrammatic mapping (as illustrated in Figure 9-1 below) of the MetaLab archives would help clarify the complex and intertwining paths of the collection.
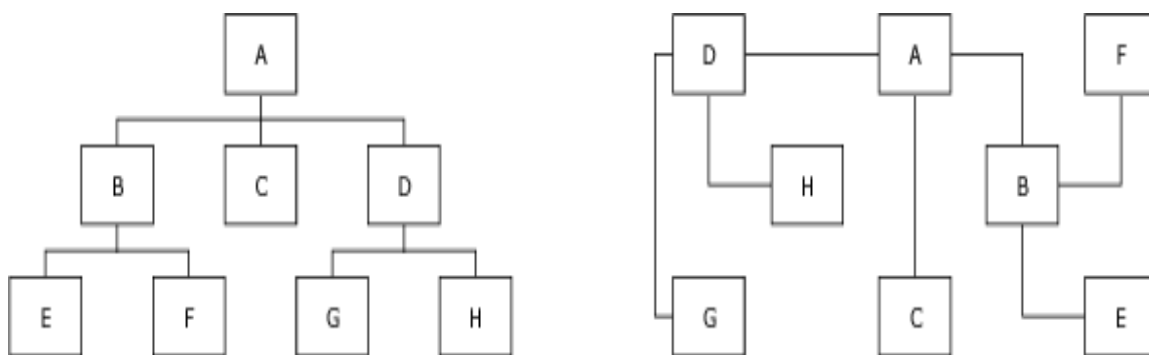
Figure 9- 1   Web Site Diagramming Schemas [69]

But, the sheer size of the collection (close to 1,000,000 files – html, image, sound, etc) made it impossible to manually map the entire web collection. Visio is a software product, produced by Microsoft that, among many other assets, creates site maps of web sites. So, Visio was installed on a machine running Windows'98, with almost 3 gigabytes free storage space. Visio was pointed at the top-level directory of metalab.unc.edu and requested to create a basic site map. Visio ran for 16 hours straight, filled 2.9 gigabytes of disk space and crashed the computer when there was no more available hard disk space. It became apparent that even if it were possible to map the site, the data generated would be too large and complex to deal with in a graphical manner.

The Digital 'Library' Becomes More Than a Metaphor

With the time deadline for the public launch of ibiblio coming closer, it was decided to adapt an existing hierarchical system for the collection indexing process. The arguments for using it were that many had been in existence in libraries for over a century and were understood by patrons. Additionally, there were guidelines for relationships between subjects ('politics' is a 'social science', not a 'philosophy').

A quick analysis was undertaken of what classification options would be available. A comparison was made between several major classification systems. The fours system examined most carefully were Yahoo, Universal Decimal Classification, Dewey Decimal Classification, and the California Digital Library's system.

Yahoo was selected as one of the four systems to receive closer examination because it had demonstrated that it functioned extremely successfully on the web and that it organization was one that web users could understand. Dewey Decimal System was and remains the standard for American children's libraries in the public library sector. Therefore, Dewey Decimal Classification was investigated because of its familiarity to American users.

ibiblio	UDC	Universal Decimal Classification	DDC	Dewey Decimal Classification	Yahoo	California Digital Library
journals					News & Media	
reference	0	generalities	000	computers, information & general reference	reference	general interest & reference
humanities	1	philosophy &	100	philosophy &	arts &	interdisciplinary

					Humanities	& ethnic
		psychology		psychology	Humanities	& ethnic
religion	**2**	religion	**200**	religion		
social sciences	**3**	social sciences	**300**	social sciences	Social Science	social sciences
					Education	
					Government	
					Business & Economy	
					Society & Culture	
language	**4**	linguistics	**400**	language		
science & math	**5**	natural sciences, and math	**500**	science	science	physical sciences
						biological & medical sciences
technology	**6**	technology	**600**	technology		engineering & technology
					Computers & Internet	biological & medical sciences
					Health	
creative arts	**7**	arts	**700**	arts & recreation		
					Entertainment	
					Recreation & Sports	
literature	**8**	language and literature	**800**	literature		
geography & history	**9**	geography, biography, &history	**900**	history & geography	Regional	geographical
						history
archives						
historic						

Figure 9- 2 comparing four hierarchical indexing systems, with ibiblio's categories (67)

"The Universal Decimal Classification was an offspring of the Dewey system and became the preferred classification system by scientific and technical libraries. Its decimal basis and attempts at hierarchical range underscore its theoretical origin in Dewey. UDC is in worldwide use, and has been published in whole or in part in 23 different languages." [68]

California Digital Library was one of the first online libraries with a classification system. Though not as popular as Yahoo, California Digital Library has proven to be a logical and easy to use system for web patrons.

The UDC was selected as the primary source model for ibiblio's classification system for two reasons: first, UDC is similar to an Open Source solution, meaning that the information that structures the classification system is freely and openly available;

Hybridized Vigor

Dewey Decimal Classification (DDC)

000 computers, information & general reference
100 philosophy & psychology
200 religion
300 social sciences
400 language
500 science
600 technology
700 arts & recreation
800 literature
900 history & geography

Universal Decimal Classification (UDC)

0 generalities
1 philosophy & psychology
2 religion
3 social sciences
4 linguistics
5 natural sciences, and math
6 technology
7 arts
8 language and literature
9 geography, biography & history

ibiblio classification

journals
reference
humanities
religion
social sciences
language
science & math
technology
   includes computer-tech
creative arts
literature
geography & history
archives
historic

Yahoo

News & Media
Reference
Arts & Humanities
Social Science
Education
Government
Business & Economy
Society & Culture
Science
Computers & Internet
Health
Entertainment
Recreation & Sports
Regional

key
multifaceted association
simple association

California Digital Library

General Interest & Reference
Area, Interdisciplinary, & Ethnic Studies
Arts & Humanities
Biological & Medical Sciences
Engineering & Technology
History
Physical Sciences
Social Sciences
Geographical (By Place)

and the system can be used without payment (Dewey Decimal System is owned by the OCLC and charges a substantial annual licensing fee); second, the international acceptance of the UDC made it more appropriate for a digital library, since it is used worldwide. Dewey tends to be more of a North American solution. But the system that was utilized at ibiblio ultimately was a combination of the four investigated systems, as illustrated in Figure 9 -3.

Figure 9- 3 Hybrid Indexing Systems

The page explaining the indexing states:

> "In creating our index categories, we looked at major classification systems: Universal Decimal, Dewey, Yahoo & the California Digital Library. Each system had its own strengths and weakness. Both decimal systems are around a century old - neither scales quickly or easily to the digital changes of the past 20 years. Most web indexes have a somewhat specialized focus.
> Ibiblio's largest collections are the Linux archives and the Linux documentation project - both are key elements in the Open Source Community. Another major ibiblio resource is the "Documenting the American South Collection." In creating an ibiblio index we have come up with a hybrid index. Like Mendel's hybrid corn, this index draws vigor from its predecessors. However, the ibiblio index allows the structural freedom to acknowledge the significance of major collection elements."   68

Design Using a Template

The graphical design of the homepage was provided by Red Hat Center, which hired a graphic designer to create a site format for the ibiblio. The information organization system then had to be meshed into the design created by the graphic designer. The design that was presented was an open, malleable design, so it was a fairly simple matter to drop elements into the modular areas that the graphic designer had provided. (see Figure 9-4).

There was a single link on the home page that led to the collection index page. The collection index page that had an alphabetic category index in the center of the page and surrounded this with the hierarchical categorical index in the sidebar (see Figure 9-4). Clicking on a category in either area would lead to a secondary level collection index page – each page covered one category from ibiblio's index system. The collection homepage and index remain fully accessible in lynx, which means that the pages are also handicapped accessible if using readers or other assisting devices.

Figure 9-4

Ibiblio's first homepage

Sept. 2000

Figure 9-4   ibiblio's First Homepage

Design Evolution after ibiblio's Launch:

The design evolved and changed after the official launch. Within a few months, it had become obvious that the collection index was overlooked if it was only identified by a single link on the homepage. The response to this was the changing the right sidebar to house the full categorical index, as had been done on the collection index pages before.

The result of that change was that users became more aware of the depth and diversity of the ibiblio collections. But some also missed the familiar 'quick links' from the homepage. In order to incorporate the categorical links, as well as the 'quick links' that regular users had come to expect (to such sites as the radio stations that ibiblio provides streaming services for or to the Dr. Fun cartoon or the Linux Documentation Project). It became obvious that the number of links had to increase, yet remain manageable somehow. A right sidebar was added which featured a relatively stable collection of links. The overall layout was simplified to avoid visual confusion.

Self-indexing:

The notion of self-indexing remains a key idea in the design of a digital library. One reason was to allow the control of a site to remain within the domain of the site's creator. The other had to do with the vigor and unpredictability of an open digital library. Some contributors might approach ibiblio's administrators with a site that was complete and simply needed a home. Others might approach with nothing more than an idea. If the site or the idea fell within ibiblio's collection guidelines, the contributor would be given a web directory and basic assistance in technological matters. To attempt indexing of all sites at the moment of releasing web space would be disastrous. Other sites would never be developed beyond the initial 'site under construction' page. What was needed was a mechanism that allowed the contributor to announce the site and index it when the contributor viewed the site as being complete.

Figure 9-5   self-indexing web form

In May of 2001, Adriane Boyd came to work at ibiblio. A talented and ambitious programmer, she built the self-indexing software in a few months. Most of the access information is self-generating. The information is provided into a form. It takes little more technological sophistication to fill out the index entry than it would to respond to a web survey. The existing collection index information was migrated into the new collection index database. Due to consistent indexing efforts since the fall of 2000, there were almost 400 entries when the self-indexing process first came online in July of 2001. Eight

months later, in March of 2002, there were 529 entries. Just by the 25 percent increase in indexed collections, the self-indexing system could be declared a success.

search

Collection Index

What's New

Arts and Recreation
Animation, Architecture, Cartoons, Crafts & Folk Arts, Fine Arts, Games, Graphic Arts, Music, Photography & Film, Recreation, Recreational & Performing Arts, Sports & Outdoor Activities

Geography, Biography, and History
Biography, Geography & Travel, Geography - Asia, Geography - Burma, History - African, History - American, History - Asian, History - European, History - World, North Carolina Culture, Southern Culture

Language
Burmese, Dutch, Gaelic, Greek, Linguistics, Russian, Yiddish

Literature
American Literature, Humor, Italian Literature, Japanese Literature, Literature, Poetry, Storytelling

Natural Science and Mathematics
Astronomy, Biology, Botany, Chemistry, Earth Sciences, Ecology, Life Sciences, Mathematics, Physics

Philosophy and Psychology
Ethics, Oriental Philosophy, Paranormal Phenomena, Philosophy, Psychology

Reference
Dictionaries, General Encyclopedic Works, General Reference, Journalism, Journals, Library Science

Religion and Theology
Atheism, Baha'i, Bible, Buddhism, Christianity,

In layout and design, the system remained much the same. The categories are generated nightly from the collections database. Any new subcategories will be added to the list of major and sub-categories that appears on the left sidebar of the homepage (see figure 9-6). The collection index presents in a variety of views: the main collection index (which can be viewed in an alphabetical arrangement or can be arranged by the Universal Decimal Classification. Clicking on a major category then leads to the category page and a display of the subcategories and sites, with their respective descriptions.

Figure 9- 6 Collection Index Sidebar

Chapter 10 – Digital Futures

July 2002, SunSITE – MetaLab-ibiblio will celebrate ten years of existence. It has been an exciting time full of innovation, risks and experimentation. At times the balance between usability and technology has leaned heavily toward technology. There have been fantastic tools and information available on the site, but it has been difficult to find the information using the list, which could only highlight a handful of sites out of the hundreds available.

Since the burst of funding from RedHat Center, the collections have expanded and diversified even further. Several archives have moved to ibiblio to explore shared and downloadable music created, recorded and made available by musicians and their affiliates, rather than being controlled by the music companies. Other new areas of exploration include art, architecture and new media. Open Source software development is moving beyond the Linux platform. There are some new archives at ibiblio developing to support non-Linux Open Source software. Open Source Book archives have developed to allow freely distributable texts. Ibiblio has become a haven of experimentation for these fledgling ventures.

The role of the staff has moved more into support for the variety of software and hardware configurations. Having a diverse and experimental archive has meant allowing contributors to run a wide variety of software packages, many of which are not stable or in full release. This can create issues and problems for the ibiblio staff. But the staff attitude is that any Open Source software package that can be accommodated will be accommodated.

The future of the archive remains dependent on funding resources. At present, most of the employees are part-time graduate and doctoral students from the School of Information and Library Science.   The budget is a shoestring, and response times can be instantaneous some weeks and agonizingly slow others, depending on class and exam schedules. There are only three permanent employees at the archive, all full-time: Paul Jones, Jonathan Magid and Fred Stutzman. Magid and Stutzman both are system administrators. Paul Jone's title's is 'fearless leader', but his role goes far beyond that. In a way, the sunsite – metalab – ibiblio archive existed, exists and will continue to exist

largely because of Jones' vision, imagination, and aspirations – aided by a lively and raucous Jeffersonian democracy of a staff and contributors.

Collectively, we never sleep.

Paul Jones
Fearless Leader

Jonathan Magid
Minister of Technology

Donald Sizemore
Should have listened
to his mama.

Jason Z. Moore
Football Captain

Serena Fenton
the Linux Diva

Andre Burton
Hard at work

John Reuning
fix it! fix it!

Fred Stutzman
You're a Frat Boy,
aren't you?

Meichun Li

Figure 10-1 ibiblio staff – Spring 2002

Endnotes

1.   Martin Dodge, *An atlas of cyberspace; historical maps of computer networks* (Centre for Advanced Spatial Analysis, 2002 [cited Feb 18 2002]); available from http://www.cybergeography.org/atlas/historical.html.

2.   Judy Hallman, "Using INFO," (comp.os.vms, 1992).

3.   Jim Fullton, "Opinions Needed," (alt.bbs, 1990).

4.   Jim Fullton, "Future of BBS'ing," (alt.bbs, 1990).

5.   Robin 'roblimo' Miller, *iBiblio Takes MetaLab Concept To A New Level* (slashdot, 2000 [cited March 23 2002]); available from http://slashdot.org/features/00/09/17/155240.shtml.

6.   Greg Easterly, "Busy computer junction at UNC-CH," *The News & Observer*, June 25, 1994 1994, B1.

7.   Steven M. Christensen, *Steven M. Christensen and Associates, Inc.* (Steven M. Christensen and Associates, Inc.

Scientific Computing Software and Consulting Services, March 1, 2002 2002 [cited March 3 2002]); available from http://smc.vnet.net/Christensen.html.

8.   Paul Jones, "what I'm sending to Graves and Groves," (Message-Id: <9202111647.AA15082@mento.oit.unc.edu>, 1992).

9.   Paul Jones, "Sun archive proposal guidelines," (<pjones@mento.oit.unc.edu>, 1992).

10.   Paul Jones, "Sun GOD needed," (support@uncvx1.bitnet, 1992).

11.   Paul Jones, "sunsite pr announcement," (MetaGroup <sungroup@metalab.unc.edu>, 1992).

12.   Jonathan Magid, "Re: Help: can I get files from sunsite using ftpmail?," (comp.os.linux, 1992).

13.   Paul Jones, "Re: Soviet archives online," (bit.listserv.pacs-l, 1992).

14.   Paul Jones, "Re: status?," (paul jones <pjones%lambada@lambada.oit.unc.edu>, 1992).

15.   Jonathan Magid, "Re: ftp sites + demise of banjo.concert.net," (Newsgroups: comp.os.linux, 1992).

16.   Trish Wilson, "President on-line at UNC-CH," *The News & Observer*, May 30, 1993 1993, A1.

17.   Jock Gill, "Electronic Budget Press Release," (<jgill@eop.gov>, 1994).

18.   The University of North Carolina at Chapel Hill Teams Up With Sun Microsystems to Launch Public Information Server on the Internet (SUN Microsystems, 1994 [cited March 24 2002]); available from http://www.sun.com/smi/Success/IndustrySpecific/Education/UnivNCarolina.html.

19.   Paul Jones, "success story," (nprdocs@unc.edu, 1993).

20.   *New at CEARCH Archives* (CEARCH, 1995 [cited March 23 2002]); available from http://www.ibiblio.org/cisco/archives/1995/july1995/whatsnew.july23.html.

21.   Judson Knott and Paul Jones, "SunSITE Serving Your Internet Needs Since 1992," *D-Lib Magazine* February 1996 (1996).

22.   IAFA-WG, *Publishing Information on the Internet with Anonymous FTP; (IAFA DOC II)* (DRAFT 92.06.10) (IAFA-WG, 1992 [cited February 15 2002]); available from http://www.ibiblio.org/pub/docs/iafa/draft.part.II.

23.   P. Deutsch et al., *Publishing Information on the Internet with Anonymous FTP* (Internet-Draft) (Internet Engineering Task Force (IETF) and Loughborough University department of Computer Studies, Mon Dec 18 12:56:01 GMT 1995 1995 [cited Feb 12 2002]); available from http://www.roads.lut.ac.uk/Reports/iafa-draft/node4.html.

24.   Jim Fullton, *IAFA-SITEINFO* (sunsite.unc.edu, 1992 [cited Feb 10 2002]); available from SunSITE server: /public/home3/old-ftp.

25.   Jeff Kopmanis, "Re: Suggestion: Current Version List," (Kansas City Linux Users Group, 1993).

26.   Jeff Kopmanis, " Announcement: Linux Software Map (preliminary)," (comp.os.linux.announce, 1993).

27.   Jeff Kopmanis, "LSM: Submission Specification (4/1/93)," (comp.os.linux.announce, 1993).

28.   Jeff Kopmanis, "LSM Volunteers linux software Map," (comp.os.linux, 1993).

29.   Jeff Kopmanis, "First release of the LSM to sunsite.unc.edu," (comp.os.linux, 1993).

30.   staff, *WAIS databases on sunsite* (sunsite.unc.edu, 1994 [cited March 24 2002]); available from http://www.ibiblio.org/jem/wais-on-sunsite.html.

31.   Matthew Gray, *Web Growth Summary* (M.I.T, 1996 [cited March 24 2002]); available from http://www.mit.edu/people/mkgray/net/web-growth-summary.html.

32.   "Intuition Proves successful at sunsite.unc," *The Sun Observer* Volume 8, no. No 5 (1995): 10, 49-50.

33.   David Farley, "[webgroup] 10 years of ibiblio/metalab/sunsite," (webgroup@metalab.unc.edu, 2002).

34.   Gioacchino La Vecchia, "[webgroup] 10 years of ibiblio/metalab/sunsite," (webgroup@metalab.unc.edu, 2002).

35.   glen, "[webgroup] 10 years of ibiblio/metalab/sunsite," (webgroup@metalab.unc.edu, 2002).

36.   Andrea Berman, "[webgroup] 10 years of ibiblio/metalab/sunsite," (webgroup@metalab.unc.edu, 2002).

37.   Henriette Kress, "Re: [webgroup] 10 years of ibiblio/metalab/sunsite," (Henriette Kress <hetta@saunalahti.fi>, 2002).

38.   Rick Gates, "First Internet Hunt !!!," (bit.listserv.libref-l 'Discussion of Library Reference Issues', 1992).

39.   Rick Gates, " Results: First Internet Hunt, Pt. 2 of 2," (bit.listserv.pacs-l, 1992).

40.   Tim Berners-Lee, *How does www keep track of the available servers?* [web page] (w3.org, 1992 [cited March 6 2002]); available from http://www.w3.org/History/1992/WWW/FAQ/KeepingTrack.html.

41.   Jakob Nielsen, *Hypertext'89 Trip Report* (Jakob Nielsen, 1989 [cited); available from http://www.useit.com/papers/tripreports/ht89.html.

42.   (attrib) Tim Berners-Lee, *Points* (w3.org, 1990 [cited March 8 2002]); available from http://www.w3.org/Conferences/ECHT90/Points.html.

43.   R. Botafogo and B. Shneiderman, " Identifying aggregates in hypertext structures" (paper presented at the Hypertext `91, San Antonio, TX,, 1991), 63-74.

44.   Robert H'obbes' Zakon, *Hobbes' Internet Timeline v5.5* (zakon.org, January 3, 2002 2002 [cited Feb. 17 2002]); available from http://www.zakon.org/robert/internet/timeline/.

45.   David Filo and Jerry Yang, *Yahoo homepage* (Bilkent University, 1994 [cited 2002 March 27]); available from http://web.bilkent.edu.tr/History/yahoo/.

46.   Brian Pinkerton, *WebCrawler Timeline* (thinkpink.com, [cited March 12 2002]); available from http://www.thinkpink.com/bp/WebCrawler/History.html.

47.   Brian Pinkerton, *Finding What People Want: Experiences with the WebCrawler* (thnikpink.com, 1994 [cited March 9 2002]); available from http://www.thinkpink.com/bp/WebCrawler/WWW94.html.

48.   Jakob Nielsen, "Enhancing the explanatory power of usability heuristics" (paper presented at the CHI'94 Conference, 1994).

49.   Jakob Nielsen, *Report From a 1994 Web Usability Study* (useit.com, 1997 [cited March11 2002]); available from http://www.useit.com/papers/1994_web_usability_report.html.

50.   Richard Furuta and John Leggett, "From the Conference Chairs" (paper presented at the First Annual Conference on the Theory and Practice of Digital Libraries, College Station, Texas, 1994).

51.   *Welcome to SunSITE!* (sunsite.unc.edu, 1993-94 [cited March 23 2002]); available from http://www.ibiblio.org/sunsite/unchome.html.

52.   *Thank you for choosing SunSITE Classic* (sunsite.unc.edu, 1995 [cited March 23 2002]); available from http://www.ibiblio.org/sunsite/sunsitelocal.html.

53.   Kellyjo Garner, *Welcome to UNC-CH's Heliocentric Information Map* (ibiblio.org, 1996 [cited March 12 2002]); available from http://www.ibiblio.org/heliocentric.html.

54.   "newsbrief . . . The biweekly bulletin of computing, instructional and information technology for UNC-CH," , 30 May 1997 1997.

55.   Paul Jones, "End of SunSITE.unc.edu - don't panic," (Web Group <webgroup@sunsite.unc.edu>, 1998).

56.   Robert Cobb, *The Mbone (or Virtual Internet Backbone for Multicast IP)* (School of Computer Science, Otto-von-Guericke-University of Magdeburg, 1996 [cited March 5 2002]); available from http://www.cs.uni-magdeburg.de/Mbone/.

57.   Kevin Savetz, Neil Randall, and Yves Lepage, *MBONE: Multicasting Tomorrow's Internet* (San Francisco: IDG - Hungry Minds, Inc, 1996).

58.   *FAQ on the Internet Multicasting Service* (info@radio.com (Automatic Response), 1994 [cited March 12 2002]); available from http://www.ibiblio.org/pub/talk-radio/faq_itr.readme.

59.   James Daly, "Music by Modem," *Rolling Stone* (1995).

60.   Gerald Graef, "Book Review: The UNIX Web Server Book," *Linux Journal*, November 01, 1997 1997.

61.   Robert O'Kelley, "**Technology grant sparks new teaching tools for film**," *Chapel Hill News*, 05/01/98 1999.

62.   Patty Kinneer, "UNC-CH's MetaLab and Digital Bitcasting forge multi-media research alliance," *UNC News Service*, March 30, 1999 -- No. 223 1999.

63.   April Simun, "MetaLab, WUNC transmit live broadcasts of Clinton's impeachment, Senate trial," *UNC-CH News Services*, Jan. 29 , 1999 -- No. 73 1999.

64.   Kirk Ross, "UNC's online library to be built by users," *The Chapel Hill News*, September 13 2000, A1.

65.   Paul Jones, *About ibiblio* (ibibli.org, October 19, 2001 2000 [cited March 10 2002]); available from http://www.ibiblio.org/about.html.

66.   ibiblio staff, *ibiblio.org Collection Criteria* (ibiblio.org, last updated: July 3, 2001 2001 [cited March 10 2002]); available from http://www.ibiblio.org/collection.html.

67.   Paul Jones, *Open[Source]ing the Doors for Contributor-Run Digital Libraries* (ibiblio.org, 2000 [cited March 10 2002]); available from http://www.cni.org/tfms/2000b.fall/.

68.   Serena Fenton, *How is Ibiblio organized?* (ibiblio, 2000 [cited March 10 2002]); available from http://www.ibiblio.org/collection/indexing.

69.   Jesse James Garrett, *A visual vocabulary for describing information architecture and interaction design;version 1.1b* (Jesse James Garrett, 6 March 2002 2002 [cited MArch 10 2002]); available from http://www.jjg.net/ia/visvocab/.

Appendices

1-A   Sun Archive Proposal Guidelines

1-B   Sun Anonymous FTP Proposal Recommendations

3-A   SunSITE PR Announcement

4-A   Automated INFO response for SunSITE

5-A   LSM Submission Specification

Appendix 1-A

Subject: Sun archive proposal guidelines
Date: Fri, 17 Jan 92 17:04:37 EST
From: Paul Jones <pjones@mento.oit.unc.edu>
To: Steve Christensen <steve@augustus.physics.unc.edu>
CC: knott@mento.oit.unc.edu, fullton@mento.oit.unc.edu, pjones@mento.oit.unc.edu

Steve,
here's a very rough first cut at answers to the questions that you posed in your proposal. The fact that I didn't answer questions 6 and 7 does not mean that I have no ideas about the long term life of the project, but that I was on vacation and have spent this week catching up on items that had even more pressing deadlines. I'd like a chance to talk to the Associate Provost before speculating on the long term or committing UNC for that matter. There a few questions that we'll have to pass on to the University lawyers.

Problem and how to deal with them is another paper in itself. Just the security angle, takes quite a bit of discussion. Of course, we would not allow uploads or downloads of untested programs because of liability problems.

I hope that this will serve as a starting point for discussion and that you and SUN will not see it as our final word on the matter.
Best,
Paul

1) Do you and your institution have experience in running widely available anonymous ftp servers? If yes, what?

We've been running an internationally accessed anonymous ftp site for over 5 years. Many days we have experienced over 100 ftp sessions in a 24 hour period. We use anonymous ftp as a way to distribute documents and software and fixes.

2) What expertise do you have to offer that is unique?

We have quite a bit of experience with the WAIS protocol and with publicly accessible systems. Our extended bulletin board (bbs.oit.unc.edu) averaged over 1200 logins per 24 period this week. We have had logins from six of the seven continents in the past month. We're still waiting for a connection from Antarctica.

ftp.oit.unc.edu is fairly well know as a place to find good documents on the Internet already. This week we've had ftp sessions from five continents (a South America connection might have happened and I could have missed it; I was away the first two weeks of the month--hence these brief and breezy replies).

3) What do you consider to be the optimal hardware and software configuration for an ftp server? What is the minimal starting configuration?

The real question with an ftp server is storage. ftpd itself doesn't require much in the way of memory or cpu power. Still I'd like to see a RISC based host and one to which memory could be added in case we wanted to add other services (and of course if SUN agreed about these services). About storage, that amount is contingent on the amount of software and documentation. Before I could propose a configuration in good faith, I would have to know more about our proposed database.

4) How would you implement the ftp software and can you provide any special interfaces or help systems for the user? Can you also provide access for BITNET users?

anonymous ftp is a fairly standard set up. The interesting problem is finding what you want once you're there. Besides the inclusion of accurate descriptive README or INDEX files at each directory level, we would use the WAIS protocol to make the paths to the programs easily tracable and the contents known. By setting up a WAIS server on the ftp machine and making available both a simple WAIS client and an X-WAIS client, actually finding the documentation or software you desire should be quite simple. Further, much of the documentation could be WAISed making WAIS an attractive way to scan documents and download them as an alternative or adjunct to ftp.
We have some experience with mailbased archive distribution of documents; that is in my mind the best way to distribute files to bitnet sites. I would not recommend putting the ftp machine on BITNET.

5) What personnel resources would you estimate would be required to maintain the server? What can you provide?

My initial guess is that we would use 1 FTE who would also become the campus SUN contact. At present SUN support is provided by personnel that also are required to support DEC, IBM. NeXT, HP, SGI, and other UNIX hosts. The ftp/SUN specialist would dramatically raise the level of SUN support available to our campus (giving SUN a greater competitive edge).
I would see the single FTE aided by 20 hours of student support. Our experience has shown that we are able to hire highly motivated students who quickly become what I call Netnaughts. Our students seem to have a higher tolerance for tasks that systems managers hate to perform such as documentation (creating README and INDEX files in this case)

6) Do you have any suggestions on the best way to run this project
   in the long term?

7) What problems do you see? How would you address them?

8) Do you believe that such a project should be limited to public domain software or could demo versions of commercial software of interest to university users also be included?

My experience tells me that demo software does very well in such an environment. SUN will have to remember that demo software often escapes. We have (for other vendors) set up demo accounts that are "captive" that is usable by a special login that requires no password. This is a better alternative than actually distributing the software by ftp or other means.

9) After delivery of a workstation, how long do you estimate it will take to have an ftp server ready to receive software?

24 hours.

Please give contact information:

Names of submitters: Paul Jones, Judd Knott, Jim Fullton

Institution: University of North Carolina

Address: Office FOR Information Technology
      Campus Box 3460
      Chapel Hill, NC 27599-3460

Phone: (919) 962-9107

Email: pjones@samba.oit.unc.edu, knott@conga.oit.unc.edu, fullton@rhumba.oit.unc.edu

Fax: (919) 962-5604

Appendix 1-B

From pjones@mento.oit.unc.edu Tue Feb 11 11:42:19 1992
Received: from mento.oit.unc.edu by augustus.physics.unc.edu (4.1/TAS/11-16-88)
        id AA06447; Tue, 11 Feb 92 11:42:01 EST
Received: by mento.oit.unc.edu (NeXT-1.0 (From Sendmail 5.52)/TAS/11-16-88)
        id AA15082; Tue, 11 Feb 92 11:47:35 EST
Date: Tue, 11 Feb 92 11:47:35 EST
From: Paul Jones <pjones@mento.oit.unc.edu>
Message-Id: <9202111647.AA15082@mento.oit.unc.edu>
To: steve@augustus.physics.unc.edu
Subject: what I'm sending to Graves and Groves
Status: RO

                    Sun Anonymous FTP Proposal Recommendations

                           Steven M. Christensen
                            Consulting Scientist
                Sun Microsystems Education and Research Marketing

                            February 4, 1992

Introduction

In January, 1992, three university sites were approached as possible sites for an anonymous ftp
server devoted to Sun and Sun-related public domain, shareware, or free commercial demo and
support software and archives. To save time, an electronic mail message was sent to the
interested parties at each place outlining the general ideas of the project and giving some
guidelines and questions for them to follow and answer. They were given only a very short time
to answer. The proposals arrived via email (UNC, Buffalo, Harvard, in that order). In this
document I will recommend a site and give a bried outline of the next steps in the project. In
order to get the project started in a reasonable time, negotiations between Sun and the selected
site should begin as soon as possible this month.

The Proposal Guidelines and the three resulting proposals are attached.

Recommendation of a Site

There are a number of key aspects to consider in selecting the best
site from the three:

1) Expertise in setting up and running such a project
2) Costs of the project
3) Viability of the worldwide network connection

4) Reputation in computing of the institution
5) Potential to support Sun at the institution and in the area
6) Future potentials

It is clear to me that the University of North Carolina best satisfies all of the above needs.

1) The UNC Office of Information Technology contains a number of highly expert communications programmers and system administrators who already provide a large and advanced network service to worldwide users. They provide support both for the university campus but also for the off campus Institute for Academic Technology and others. Their office as already done significant evaluation and testing of every major workstation vendors products. They have access to a large force of student assistance. The campus is famous for its computer science department and research into computer graphics and visualization. It has provided numerous workers for the Sun facility in the Research Triangle Park. They an also be expect to be able to provide BITNET access.

UNC is strongly supporting the WAIS research efforts and it is likely that developers from OIT will take over much of that development. They expect UNC to become "WAIS Central". This is allow the Sun archive to become a central part of any such project. Other vendors are highly supportive of this effort and have provide equipment to OIT. Sun has not.

The other institutions do not have the same connection and in one case even expressed a concern over potential network overload. They each lack the long term large project experience that UNC has.

2) Costs for personnel in New York and Massachusetts will be higher.

3) UNC-OIT has already been running a worldwide network access program with thousands of user connections. The other institutions have not. OIT was able to provide a more detailed plan based on the guidelines two weeks before the other schools. It is a very professionally run organization.

4) UNC is a well known research university in a high tech area not strongly affected by the bad economic times. The Research Triangle area has a very high presence of research and academic works with the highest density of Ph.D.'s in the country - even higher than in the Boston area. UNC hosted the first NSF conference on workstation use in the physical sciences in 1984.

5) Sun has come under heavy competition for IBM, Data General and DEC in this area with HP and NeXT also present and growing. A Sun project here will help to fend off these other vendors.

6) The three major universities (UNC, Duke, NC State) and many smaller colleges in the area, plus over a hundred high tech companies and government research facilities (EPA, US Forest Service, Glaxo, MCNC, NCSC, National Statistics Research Center, ...) can provide a large resource for Sun development and customers, influencing sites nation wide. A very large number of students in the computational sciences can provide a willing force of programmers for new public domain software.

Project Design Recommendations

The design of this project is straightforward, but certain legal and other details will need to be worked out with the site selected. The following recommendations can be carried out given sufficient support from Sun and the site. A timeline will need to be negotiated.

Hardware:

    Sparcstation 2
        32 Megabytes RAM
        19" Color Monitor
        3 Gigabytes Disk (more added as needed)
        CD-ROM Drive
        3.5" Floppy Drive
        Network connections and cables
        8mm tape backup device

Software:

    Solaris 2.0
    C/C++ Compilers
    Programming and GUI tools
    Networking

It is clear that in order to provide sufficient performance to handle a number of simultaneous users downloading large files, a Sparcstation 2 level machine is required. Further, software testing and development will require significant performance. The most important aspect will be a large amount of disk space for the storage of a very large number of programs and archives. Given the large size of the archives, an 8mm tape backup device will be needed. All software required to support software testing and development should be included on the machine with continued upgrades as needed. Many of the important public domain program were written under version of Sun OS earlier than Solaris 1.0 or OS 4.1.1 and hence testing will be required.

After delivery of the machine and software to the site, it will be placed on the network and communications from various outside sites tested and the machine registered on the net. Then the anonymous ftp software will be activated.

The next step is to determine a first listing of the software that should be placed on the machine. This would include the X-Windows, TeX, GNU, comp.sources.sun programs and other major packages of wide interest and support. In every case, the software packages will be tested to see that they currently work under the Solaris and Sun compiler systems. In those cases where SunView, OpenWindows, X-Windows R5 are included as part of the software, ports to the latest Sun GUI should be pursued. This can come from encouragement to the original developers or by programmers at the ftp site.

Once sufficient software is available, an announcement can be made indicating the availability and rules of the anonymous ftp server. Requests for contributed software can also be made.

It is important that the archive be up-to-date, indexed, and registered with the Archie service. New software posted to the net will need to be tested and added as it become available. This will require the project members to survey a large number of new newsgroups and mailing lists. A periodic listing of the server contents and rules for use can be given to Sun representatives for dissemination to the sales force and others in Sun.

The above activities will require at least 1.5 FTE's. A system administrator can be used to keep the server running properly and indexed. Date on server usage can be maintained by such a person. This can be a student. Another person should be the "overseer" of the project - the person who selects the software, tests it, adds it to the archive (perhaps in compiled as well as source form), scans the net for new software and encourages developers to submit new or updated works. This cannot be a student and will require a highly skilled as well as respected individual who can speak to the software community. It is clear from all of the applications that no institution is will to provide free labor costs to Sun.

```
-----------------------------
Original Proposal Guidelines
-----------------------------
```
Below is the description of the project and the guidelines that were send to the three institutions.
```
----------------------------------------------------------------------
```

With the crisis in funding in university and research institutions in the US, there is a major need for inexpensive, but reliable, software for Sun workstation users. There is a wealth of software in the public domain that can be made available to Sun users.

Further, software developers at universities need a visible location where their software can be tested, distributed, and improved. Informational resources such as the Sunspots, SunFlash, Sun Managers and others also need a central archive.

As a consulting scientist to Sun, I have recommended that Sun help to support an effort to collect this software in one location on a university campus. The campus selected will become a key location in the growth of workstation use. The document below is an effort to guide you in producing a proposal to Sun. If you have questions or suggestions, please email me.

We would like to have email proposals returned to us no later than January 20. Then, after review, we will select the best proposal and ask for a formal written proposal. Barring problems, final selection should be made by the end of January.

Watch for supplemental information via email.

Steve Christensen

```
========================= cut here ==============================
```

Anonymous ftp Server for Sun Public Domain Software

Project Description and Proposal Guidelines

January 1992

Contacts:

Steven M. Christensen, Ph.D.
Consulting Scientist for Sun Microsystems
c/o MathSolutions, Inc.
P.O. Box 16175

Chapel Hill, NC 27516
Phone/Fax: 919-967-9853
Email: stevec@wri.com or steve@physics.unc.edu

Annette Nemetz
Market Segment Manager, PAL-1
Sun Microsystems, Inc.
2550 Garcia Avenue
Mountain View, CA 94043
Phone: 415-336-5559
Fax: 415-336-0643
Email: Annette.Nemetz@Corp.Sun.Com

Project Description

The Sun Microsystems Education and Research Group believes that there is a need for a central location for the collection of public domain Sun software. There are many hundreds of programs already located in the over 800 anonymous ftp sites on the worldwide networks. The anonymous ftp server in this project will provide, over the next year, a new place for developers in universities to place their software. This will help these developers obtain comments about, and an audience for, their work and give Sun users a library of free software that is organized and focused. This is be a major service to the academic and research Sun user community.

It is estimated that there are several gigabytes of potential software and other archival information such as the SunSpots and SunFlash files. It is likely that several thousand Sun users will make use of the new server in the first year. Software will need to be organized in a logical fashion with continuing updates of already installed software and the addition of new programs.

Sun is soliciting proposals from a number of universities on how they might set up this project. The following guidelines and questions will help Sun and its representatives to determine which site will best serve the goals of this project.

Guidelines

1) The host institution will need to provide a network link to a Sun Workstation. The link should be maintained at all times and be accessible to the Internet. An appropriate Sun workstation will be provided by Sun. The kind of workstation is yet to be determined. So long as the workstation's primary ftp use is not affected, this workstation may be used for other purposes.

2) System administration and the setup of the anonymous ftp software on the workstation will be provided by the host institution.

3) The minimum duration of this program will be two years. The host institution will need to make a commitment for at least two years.

4) The host institution will provide logins to the ftp server machine so that Sun representatives can add, modify or remove public domain programs.

5) The host institution must allow Sun to publicize the availability of the ftp server. This will be done through newsletters distributed electronically and in print to Sun users.

6) The host institution may add its own Sun related files to the ftp server for local use.

7) Optionally, a BITNET access can be provided so that such users can receive certain portions of the archives via email requests.

8) The archive should be registered with the "archie" database of servers in Canada.

Proposal Requirements

Each proposal should address first how the Guidelines above could most efficiently and economically be satisfied by the proposers institution. Secondly, an estimate of the hardware, software, and personnel requirements to make the project work should be made. Your realistic comments on any of the elements of this project are welcome.

The following questions should be answered:

```
----------------------------------------------------
University of North Carolina at Chapel Hill Proposal
----------------------------------------------------
```

1) Do you and your institution have experience in running widely available anonymous ftp servers? If yes, what?

We've been running an internationally accessed anonymous ftp site for over 5 years. Many days we have experienced over 100 ftp sessions in a 24 hour period. We use anonymous ftp as a way to distribute documents and software and fixes.

2) What expertise do you have to offer that is unique?

We have quite a bit of experience with the WAIS protocol and with publicly accessible systems. Our extended bulletin board (bbs.oit.unc.edu) averaged over 1200 logins per 24 period this week. We have had logins from six of the seven continents in the past month. We're still waiting for a connection from Antarctica.

ftp.oit.unc.edu is fairly well know as a place to find good documents on the Internet already. This week we've had ftp sessions from five continents .

3) What do you consider to be the optimal hardware and software configuration for an ftp server? What is the minimal starting configuration?

The real question with an ftp server is storage. ftpd itself doesn't require much in the way of memory or cpu power. Still I'd like to see a RISC based host and one to which memory could be added in case we wanted to add other services (and of course if SUN agreed about these services). About storage, that amount is contingent on the amount of software and documentation. Before I could propose a configuration in good faith, I would have to know more about our proposed database.

4) How would you implement the ftp software and can you provide any special interfaces or help systems for the user? Can you also provide access for BITNET users?

Anonymous ftp is a fairly standard set up. The interesting problem is finding what you want once you're there. Besides the inclusion of accurate descriptive README or INDEX files at each directory level, we would use the WAIS protocol to make the paths to the programs easily traceable and the contents known. By setting up a WAIS server on the ftp machine and making available both a simple WAIS client and an X-WAIS client, actually finding the documentation or software you desire should be quite simple. Further, much of the documentation could be WAISed making WAIS an attractive way to scan documents and download them as an alternative or adjunct to ftp. We have some experience with mailbased archive distribution of documents; that is in my mind the best way to distribute files to bitnet sites. I would not recommend putting the ftp machine on BITNET.

5) What personnel resources would you estimate would be required to maintain the server? What can you provide?

My initial guess is that we would use 1 FTE who would also become the campus SUN contact. At present SUN support is provided by personnel that also are required to support DEC, IBM, NeXT, HP, SGI, and other UNIX hosts. The ftp/SUN specialist would dramatically raise the level of SUN support available to our campus (giving SUN a greater competitive edge). I would see the single FTE aided by 20 hours of student support. Our experience has shown that we are able to hire highly motivated students who quickly become what I call Netnaughts. Our students seem to have a higher tolerance for tasks that systems managers hate to perform such as documentation (creating README and INDEX files in this case)

6) Do you have any suggestions on the best way to run this project in the long term?

There are two ways that the project could be run. My preferred is that the project be reviewed at the end of two years and be redesigned and funded at the level that our experience tells us is necessary. At that point both UNC and SUN will have much better ideas of the cost and benefits involved. It may be that the advantage of having specific SUN support on campus combined with support for the Internet archive will be enough for SUN to want to continue funding the project at the same level. It may be that UNC and the UNC SUN owners will want to participate in the support of the project. At any rate, a review will help access this. As it now stands, it would be difficult for either institution to commit without some experience with the project. If neither institution wants to continue the project at the necessary funding level, SUN users group volunteers could run the project given that they are willing and that any liability problems can be resolved. This is definitely an "on the cheap" solution which may not be desirable to either SUN or UNC.

7) What problems do you see? How would you address them?

The major problem areas are accessibility, security, and reliability. Accessibility means not only that the host can be reached by ftp, but that someone on the net can actually find programs and documents that they want to retrieve. To this end WAIS and archie are very useful and is good documentation in the form of README and INDEX files. Security means that unrestricted uploading is not allowed and that other services are both usefully accessible and undamaging to the system. Also access to other Internet services, if provided, must be allowed in such a way as to safeguard other Internet hosts. Our extensive experience with the extended bbs will prove invaluable in this area. Reliability means that the archives must be up to date, that software provided must be clean and current, and that the host will be backed up on a regular basis. This is good systems administration for the most part.

8) Do you believe that such a project should be limited to public domain software or could demo versions of commercial software of interest to university users also be included?

My experience tells me that demo software does very well in such an environment. SUN will have to remember that demo software often escapes. We have (for other vendors) set up demo accounts that are "captive" that is usable by a special login that requires no password. This is a better alternative than actually distributing the software by ftp or other means.

9) After delivery of a workstation, how long do you estimate it will take to have an ftp server ready to receive software?

24 hours. That is to set up the machine. Of course, actually loading the software and providing indexing and access tools will take longer.

Please give contact information:

Names of submitters: Paul Jones, Judd Knott, Jim Fullton

Institution: University of North Carolina

Address: Office FOR Information Technology
    Campus Box 3460 Chapel Hill, NC 27599-3460

Phone: (919) 962-9107

Email: pjones@samba.oit.unc.edu, knott@conga.oit.unc.edu, fullton@rhumba.oit.unc.edu

Fax: (919) 962-5604

Appendix 3-A

Subject: [aluminums] sunsite pr announcement
Date: Tue, 8 Jan 2002 14:25:06 -0500 (EST)
From: Paul Jones <pjones@metalab.unc.edu>
To: MetaGroup <sungroup@metalab.unc.edu>

draft with comments and questions from Sept 9 ,1992
one famous comment:
Please use SunSITE instead of SunSite. SITE stands for:
Software, Information, and Technology Exchange.
=   =   =   =

SUNSITE RELEASE

FOR MORE INFORMATION:

Sun Microsystems Computer Corporation
Robert Manetta (415) 336-0979

University of North Carolina

SMCC AND THE UNIVERSITY OF NORTH CAROINA ANNOUNCE AVAILABILITY OF
FREE OR LOW-COST SOFTWARE FOR EDUCATION AND RESEARCH

CHAPEL HILL, North Carolina - October XX, 1992 - Sun Microsystems Computer Corporation
(SMCC) and the Office of Information Technology at the University of North Carolina have
joined together to provide a dedicated anonymous-ftp server, "SunSite." SunSite will act as a
central respository for a collection of public-domain software, shareware and other electronic
material such as research articles and electronic images.

Academic software developers who create software destined for the public domain will be
encouraged to submit their creations to the program at **waiting to get e-mail address from Judy
Gallegos; she needs to do a test first**. Those who chose to do so will be taking advantage of a
free, easy distribution channel as well as increased exposure to every user on the Internet. In
addition, much like having research documents published, posting information on SunSite may
also lead to additional prestige and tenure credit.

The 200-year-old University of North Carolina at Chapel Hill was the nation's first state
university. Since the nineteenth century, it has offered distinguished graduate and professional
programs. In February of this year, SMCC donated a SPARCstation 2 through the company's
Academic Grant Program to the university. The SunSite project leader at the University of North
Carolina is Judson Knott. "Due to the nature of the education and research markets, accessibility
to free or low-cost software is extremely important." Regarding this joint project, Knott added,
"The SunSite project at the University of North Carolina will allow academics to access public
domain software easily and in one central source and will also provide them the vehicle to post
research papers or software packages of their own."

Key to the success of any software in the public domain is its quality. Significant quality assurance will be built directly into the SunSite program to ensure that its software is as bug-free as possible. Data and software packages must be submitted to SunSite. After a prompt evaluation, the packages will be accessible to SunSite users.

The Technology

The SunSite server will use an advanced ftp interface and WAIS protocol, which will make it easy to use. Anonymous ftp is a process that permits retrieval of files from a properly configured machine, typically over the Internet, a global wide-area network. An example of the types of files that can be retrieved via anonymous ftp include archived public-domain software, shareware, and technical documentation, as well as other types of electronic information. Worldwide, more than 1000 anonymous ftp servers provide a wealth of information across the Internet.

Although anonymous ftp servers have been available for some time, the original anonymous ftp interface is difficult to use and has changed very little since the mid 1970s.**NB: Knott thinks anonymous ftp may have been around only since 1984--Judy attempting to verify** Because this interface has been a barrier for average users, anonymous ftp has been almost exclusively the domain of advanced users. SunSite has attempted to overcome this barrier by incorporating Wide Area Information Server (WAIS) a new public technology, developed by Brewster Kahle of Cambridge, Massachusetts-based Thinking Machines. WAIS is an emerging open-systems standard protocol for querying and retrieval of information. It provides an effortless interface by accepting queries in plain English and returning programs that match the descriptions in the queries. WAIS will make anonymous ftp servers much more accessible to less experienced users.

Knott and his colleagues at UNC expect to see an explosion of electronic information in the coming decade. "We must act now to develop the tools to manage this expected information avalanche," says Knott."The development of standard wide-area network protocols such as WAIS and the integration of these standard protocols into new and existing services, such as ftp, are absolutely critical to our success in accessing, publishing and managing electronic information. The SunSite project's integration of WAIS into the standard ftp environment will allow people with little or no previous ftp experience to quickly and easily locate and retrieve relevant electronic information, in a wide variety of formats, from SunSite." Building A Relationship

SMCC has extensive ties to the educational and research communities and supports efforts such as the SunSite project through collaborative research, conferences/seminars, equipment grants, hardware and software purchase programs, software development programs, success stories, third-party solutions and, the higher education council**generic term or specific title?**. The relationship between technology and academia is a longstanding, mutually beneficial one. Today, the academic environment has more than 6,500 SPARCstations and SPARCservers, a figure that represents nearly 12 percent of SMCC's overall global revenue.

"Sun's philosophy within the academic community is one of partnership. The SunSite project is a direct result of the integral relationship between education and technology. By working together we can provide better solutions and technology to the general academic population," said Michael Majdalany, SMCC's Market Segment Manager for Education and Research.

The University of North Carolina at Chapel serves all the people of the state and the nation as a center for creative scholarship and creative endeavor. The SunSite project is located in the Office of Information Technology/Computing Systems.

Sun Microsystems Computer Corporation, a subsidiary of Sun Microsystems, Inc., is the world's leading supplier of client-server computing solutions, which feature networked workstations and servers that store, process and distribute information. Used for many demanding commercial and technical applications, SMCC's products command the largest share of the computer industry's fastest-growing market segment: workstations and servers. Sun Microsystems, Inc., founded in 1982 and headquartered in Mountain View, Calif., is a multibillion-dollar corporation doing business worldwide.

      ###

Sun, Sun Microsystems, the Sun logo, SMCC and Solaris are registered trademarks are trademarks or registered trademarks of Sun Microsystems, Inc.

==
      Paul Jones
     "Gort! Klaatu barada nikto!"
http://www.ibiblio.org/pjones/ at the Site Formerly Known As MetaLab.unc.edu
 pjones@ibiblio.org   voice: (919) 962-7600   fax: (919) 962-8071
==

Appendix 4-A

Subject: Automated INFO response for SunSITE

Thank you for your interest in UNC and Sun Microsystems' software and information archive, SunSITE.unc.edu, operated by the Office of Information Technology at the University of North Carolina at Chapel Hill. This letter includes three different documents (other than this heading) explaining (1) how to use SunSITE's anonymous ftp service, (2) how to use the Wide-Area Information Server, and (3) how to submit things that you think would be of interest to the Internet community. Please direct comments to ftpkeeper@SunSITE.unc.edu.

You may now access SunSITE via Mosaic and WWW. Use this URL:
http://sunsite.unc.edu/unchome.html

You may also access SunSITE by gopher. Simply point your gopher client to SunSITE.unc.edu. Gopher clients for various computers can be found via anonymous ftp in pub/packages/infosystems/gopher.

SunSITE supports several special guest logins. These login ids do not require passwords for their use.

gopher   - a simple terminal window gopher client
politics - a simple WAIS client with about 10 databases related to politics
swais   - a simple WAIS client with over 400 databases of all kinds

----

A Short Guide to Anonymous ftp
on SunSITE.unc.edu
152.2.22.81

Introduction

A great deal of useful information is stored in files at computers throughout the world. Many of these file are freely available to you as a member of the Internet. A simple method for transferring such files from a remote computer to a user's computer is anonymous ftp. Anonymous ftp allows a user to transfer files without having an account on SunSITE (i.e. the user is anonymous.) SunSITE uses a special version of the ftp program from Washington University - St. Louis that displays README files that describe each directory level of the archives as you move through the archives' file directory. This will help you find the programs, documents, and data items that you want to retrieve fairly easily.

How Do I Access the Anonymous ftp files at SunSITE.unc.edu?

An example of connecting to the SunSITE.unc.edu anonymous ftp site follows:

===

EXAMPLE:

% ftp sunsite.unc.edu (you type this at your local computer)

Connected to sunsite.unc.edu.
220 calypso FTP server (Version 6.9 Tue Aug 25 13:46:21 EDT 1992) ready.
SunSITE.unc.edu is a service of Univ of North Carolina & SUN Microsystems.
Read DISCLAIMER.readme for a statement of legal responsibility.
Contributions, Complaints and questions should be sent to
<ftpkeeper@SunSITE.unc.edu>. For Name: use anonymous Name
(sunsite.unc.edu:pjones): anonymous (you type anonymous)

331 Guest login ok, send e-mail address as password. Begin your
password entry with a - (dash) to suppress the helpful messages.
Password: pjones@lambada.oit.unc.edu (please use your own mail address)

230-   WELCOME to UNC and SUN's anonymous ftp server
230-   University of North Carolina
230-   Office of Information Technology
230-   SunSITE.unc.edu
< a few introductory lines deleted>
230- An OpenLook FTPtool can be found in
230-   /pub/X11/Openlook/Ftptool4.3.unc.1.tar.Z
230- This is the top directory!! cd pub will take you to the good stuff.
230-
230-Please read the file README
230- it was last modified on Thu Aug 20 23:14:18 1992 - 7 days ago
230 Guest login ok, access restrictions apply.

ftp>

==

Once you have gained access to SunSITE, the ftp> prompt returns and acknowledges that the
system is ready to use.

Listing the Directories and Changing Directories

Once you have connected to Sunsite, to transfer a file, you may have to change directories to
reach the directory that includes the files in which you are interested. In the example below the
user has typed in the commands that follow the ftp> prompt in order to retrieve a file named
Ftptool4.3.unc.1.tar.Z from SunSITE. Ftptool4.3.unc.1.tar.Z is an OpenLook Graphical User
Interface for FTP. You may use Ftptool instead of the line mode ftp command described here
once you retrieve it.

==

EXAMPLE:

ftp> cd pub
250- UNC Sun Archive - /pub - top level directory

250-
250-X11/ interesting packages that run under X-Windows
250-academic/ software written by researchers in different disciplines
250-archives/ archives of mailing lists and Usenet newsgroups
250-bbs/ the source to utilities running on the internet bbs
250-docs/ a large selection of documents, books, speeches, papers, etc.
250-gnu/ Unix Software written by the Free Software Foundation.
250-books/ entire books for download
<lots of other stuff left out>
250-sun-fixes/ patches to Suns from the Computer Emergency Response Team
250-sas/ programs and drivers for SAS (particularly those for SUNs)
250-sun-managers/ archives of the sun-managers newsgroup
250-sun-sounds/ sounds that you can play on SUNs
250-sunflash/ archives of SunFlash, a Florida-based SUN newsletter
250-sunspots/ archives of the SUNSPOTS mailing list
250-wais/ wais clients for accessing archived databases on this host
250-
250 CWD command successful.

ftp> ls

200 PORT command successful.
150 Opening ASCII mode data connection for /bin/ls.
README
sun-sounds
docs
DISCLAIMER.readme
micro
sas
X11
terminal-emulators
<many many lines left out>
226 Transfer complete.
231 bytes received in 0.051 seconds (4.4 Kbytes/s)

ftp> cd X11

250-Openlook clients that run under the Openlook window manager
250-4va-1.21.tar.Z program to view and manipulate 4 dimensional objects
250-xfract106.shar.Z xfractint. program to calculate and view fractals
250 CWD command successful.

ftp> cd Openlook
200 PORT command successful.

ftp> ls   (this lists the contents of the directory)

200 PORT command successful.
150 Opening ASCII mode data connection for file list.
xvnews.tar.Z
olvwm3.tar.Z

workman1.0b.tar.Z
Ftptool4.3.tar.Z
Ftptool4.3.unc.1.tar.Z
28507 bytes received in 1 seconds (28 Kbytes/s)

ftp> binary   (this lets you retrieve binary information -- see notes)
200 Type set to I.

ftp> get Ftptool4.3.unc.1.tar.Z (this grabs the file for you)
200 PORT command successful.
150 Opening BINARY mode data connection for Ftptool4.3.unc.1.tar.Z (609253 bytes).
226 Transfer complete.
local: Ftptool4.3.unc.1.tar.Z remote: Ftptool4.3.unc.1.tar.Z
609253 bytes received in 2.7 seconds (2.2e+02 Kbytes/s)
ftp> quit
221 Goodbye.

Now you have the Ftptool. Use uncompress and tar -xvf to uncompress and unarchive the file.
You will have created a directory called Ftptool4.3. cd Ftptool4.3 and type make. You will have
made a program called ftptool. By running ftptool, you may easily access SunSITE and other ftp
reposities all over the world.

==

Some Useful ftp Commands for SunSITE

Unix directories are hierarchal. In the example above some simple Unix
commands were used to change directories and list possible directories or
files (i.e. move up or down the hierarchy.)

the ls command lists the contents of the active directory

the cd <directory name> command enables the user to change directories
       (move up or down the hierarchy)

the cd .. command allows the user to return to the previous directory,
       using this command it is possible to return to the initial directory

the get command retrieves files from SunSITE

the put command places files on SunSITE. Note that this will ONLY work
       within the uploads directory.

the binary command sets the transfer type for binary files. You should
       always issue the binary command before retrieving compiled programs or
       files with endings of tar or Z.

A Note About Special File Types in the SunSITE Archives

You will notice that many of the files on SunSITE have special endings such as .Z .tar .zip .hqx and .sit Such files have been compressed and/or specially encoded to allow them to be more quickly and easily transfer in working order via ftp. It is a good idea to use the binary command before attempting to retrieve any of these file types.

.Z files have been compressed using the UNIX compress command. Once you have retrieved a file with the .Z suffix, you should use the UNIX uncompress command which will create a new file (uncompressed and larger than the original) without the .Z ending.

.tar files are usually a collection of several related files and directories combined as one file for transfer or archiving by the UNIX tar (tape archive) command. Once you have retrieved such a file, you should use the tar -xvf <filename.tar> command to extract the files and directories from tar format. The <filename.tar> file will remain on your local computer even after the extraction has successfully completed. It is a good idea to remove the tar file after extraction.

.zip files are usually files for pcs running DOS or Windows that have been processed by the zip program. Use the unzip program (or pkunzip) to make such files usable.

.hqx files are usually Macintosh files that have been converted to binHex format for files transfer or archiving. Use the binHex program's download option to convert .hqx files to a usable format.

.sit files are usually Macintosh files and folders that have been combined and compressed for transfer and archiving (much like tar files in UNIX). Use the StuffIt program to unpack such files and folders.

WAIS - The Wide Area Information Server

WAIS is pronounced "ways", and was developed at Thinking Machines Corporation.

WAIS is a database system that exploits two recently popularized computer science concepts: the client-server model, and full-text databases. It gives the ability for users to search existing databases of articles, books, references, abstracts and specialist information (such as genome databases, usenet group archives, ftp-site listings, etc), and for people with information to publish it at little expense and effort over the Internet.

The client-server model is a commonly used method of providing services over a network. The end-user uses a client program to access information by communicating with a server program. Typically the server and client are running on different machines and communicate over the network. You need not have a server of your own to be able to use the client.

The full-text database is a model designed explicitly to search documents on any word that appears in those documents. The old approach was to have a set of keywords which you could search on, and these words represented a small subset of all the words in the documents.

It is the WAIS client that lets you search databases. Databases are provided over networks by WAIS servers. Servers are available to run under System V and BSD releases of Unix, VMS, the IBM RS/6000, and the NeXT. Client software is available for dumb terminal Unix, curses on

Unix, GNU Emacs, Macintosh (with MacTCP), the NeXT, MS-DOS machines with and without Windows, as well as VMS, RS/6000, X Windows, OpenLook, Motif and Sunview. You can get these clients and servers from SunSITE.unc.edu via anonymous ftp. Look in the pub/wais directory.

The fundamental concepts in WAIS are the database, the document, the source and the hit. A source is a short text file that describes how a client can access a database that is provided by a server. It typically lists the database name, the machine the server program is running on, a brief description of the database, the name of the maintainer of the database, and the cost (if any).

A document is the basic unit - when you perform a search and look at results, you will be looking at documents. Databases hold lots of documents, and the server will search all the documents in the database. When the server finishes the search, it sends the client a list of hits - the names of documents that looked like they matched what you were searching for. A hit is one document name.

To try out WAIS, without compiling anything, you have to be on the Internet. The following is a brief guide to the curses-based Unix client. Telnet to SunSITE.unc.edu and log in as "swais" (do not type the quotation marks). You will not need a password. Then enter your terminal type when prompted for it (most will be vt100). You will then be connected to a simple terminal window client program, and will be presented with a list of names of databases to search (each name of a database corresponds to a source).

You can move through the list of sources with your arrow keys or the j and k keys. The question mark (?) key gives you help. To search for a word in the database names, use the / key. For example, /sun will move you quickly to the databases relating to Sun computers. A database fo selection will be highlighted. (For instance, SunSITE.unc.edu offers the "sun-fixes" database. Use the arrows or j keys to highlight that database name if you wish to search it). Once your cursor is on the source you wish to search press Return and you will be asked for keywords to search on. (Notice that an * has been placed beside the selected source). Enter the keywords, separated by spaces (for instance, "mail sendmail", without quotation marks, would search for the words mail and sendmail).

The client then gives you a list of hits - the names of documents that match your request. You can move through the list with the same commands as before. To view a document, press Return when the name of the document is highlighted. Press H for help while viewing the document. Press 'q' to leave from the list of documents.

Nat (gnat@kauri.vuw.ac.nz -- Nathan Torkington -- is the electronic text and MS-DOS archivist for the Victoria University of Wellington, New Zealand)

Modified and extended for Sunsite by Paul_Jones@unc.edu

SUBMISSION PROCEDURES FOR SUNSITE

Any large archive requires a clearly defined set of policies regarding submissions. These guidelines are designed to allow the maintainers to quickly and efficiently provide access to

materials submitted to the archive, and to ensure that only the most recent versions of all submissions are provided. If you have any questions about these guidelines, send them to ftpkeeper@SunSITE.unc.edu.

Submission procedures for the SunSITE archive are quite simple, but must be followed by all submitters. We have three broad "package" areas; software, documents, and data.

SOFTWARE PACKAGE SUBMISSION POLICY

a) The software must be provided as a single archive file with a clear name containing the version number of the package. If the software consists of more than one file, a standard archival program must be used to create a single file. All archives must be "system friendly", i.e. they must create a subdirectory for the software they contain.

b) The archive must contain sufficient documentation to allow an inexperienced user to compile and operate the software. This documentation must contain at least a users manual and an installation guide.

c) A software package description document must be provided and is described below. This document must conform to the standards set forth by the Internet Engineering Task Force Internet FTP Archives Working Group (IAFA).

d) The archive file, complete with the IAFA document and all user documentation must be placed in the "uploads" directory at sunsite.unc.edu and a mail message containing the IAFA document must be sent to ftpkeeper@sunsite.unc.edu. This mail message must also contain a telephone number where the author may be reached. This phone number will not be distributed.

SAMPLE ARCHIVE FILE

Consider a hypothetical statistical analysis package called "analysis".

        analysis-1.0.tar

This package must create a directory and unpack itself into it.

        tar -xf analysis-1.0.tar

        analysis/IAFA-PACKAGE`
        analysis/README
        analysis/doc/manual.txt
        analysis/doc/manual.ps
        analysis/src/main.c
        analysis/src/parse.c
        analysis/src/graph.c
        analysis/src/Makefile
        analysis/src/README

> analysis/src/INSTALL

The IAFA document must be called IAFA-PACKAGE and contain the following information:

Title: Title of the package
Version: A version number or string associated with the package
Description: A description of the function of the program(s) contained
              in the package.
Author: Name and e-mail address of the author. A telephone number
              is optional
Maintained-by: This field should be included when the current maintainer
              of the package is known. In most cases, it will be the author.
Maintained-at: sunsite.unc.edu. Other sites may also be specified by the
              submitter.
Platforms: Any requirements or restrictions that the package may have
              in terms of hardware or software (OS) platforms. The
              programming language the package is written in must be
              included.
Copying-Policy: The status of the package for copying purposes. This field
              must contain one of the following entries:
                    "Public Domain"
                    "Freely Redistributable"
                    "Voluntary Payment" (shareware)

Keywords: Keywords appropriate for users trying to locate this package.

DOCUMENT SUBMISSION POLICIES

a) Single documents are acceptable, but we prefer that documents be submitted as a package containing:

   1) The document in plain ASCII text format.
   2) The document in PostScript format.
   3) The document in the author(s) preferred word processing
      format.
   4) An IAFA document description abstract called IAFA-ABSTRACT. The
      description of this IAFA document follows.

The filenames of the documents must be descriptive, with the file extension defining the document format. Examples:

   1) ASCII Text - .TXT
   2) PostScript - .PS
   3) SGML - .SGML
   4) HTML - .HTML
   5) TEX - .TEX

These examples are not meant to limit the acceptable formats. SunSITE will maintain a list of formats that may be found in the archive.

b) Each document must have the name of the author, and while we prefer that the author(s) provide an e-mail or postal address in the document, it is not required.

c) Document packages should be created using standard archival software such as TAR. The document packages must be "system friendly", as discussed in the software submission policies.

d) The document package, complete with the IAFA document and all user documentation must be placed in the "uploads" directory at sunsite.unc.edu and a mail message containing the IAFA document must be sent to ftpkeeper@sunsite.unc.edu. This mail message must also contain a telephone number where the author may be reached. This phone number will not be distributed.

Format of IAFA-ABSTRACTS

Document-Name: Filename containing the document, without extension
Title: Title of the document
Authors: Names of authors, with e-mail and/or postal addresses.
Authors: Names of authors, with e-mail and/or postal addresses.
                 Please separate these name/address entries with a
                 semicolon.
                 Example:

                 Jim Fullton <Jim_Fullton@unc.edu> 310 Wilson Library,
                 UNC Chapel Hill, NC 27599-3460, USA; Paul Jones
                 <Paul_Jones@unc.edu>
Revision-Date: Date of last document revision
Category: A "author defined" category into which the document
                 might fall.
                 Example: "User Guide", "Conference Proceedings",
                 "Technical Report", etc.
Abstract: Summary of the document, in plain text.
Format: A list of formats in which the document is available.
Citation: The official bibliographic entry for the document.
Publication-Status: Current status of the document (draft, published, etc).
Keywords: Keywords relevant to the document
Size: Size of document in pages.

DATASET SUBMISSION POLICIES

Datasets may be submitted to the SunSITE archive. No restrictions are placed on dataset format.

a) Datasets must be submitted as an archived package (see above) containing the following:

    1) The dataset itself, with a descriptive name.
    2) An IAFA document (IAFA-DATASETS) describing
       the dataset. The description of this document
       follows.

b) The dataset package, complete with the IAFA document and all user documentation must be placed in the "uploads" directory at sunsite.unc.edu and a mail message containing the IAFA document must be sent to ftpkeeper@sunsite.unc.edu. This mail message must also contain a telephone number where the author/creator may be reached. This phone number will not be distributed.

Format of IAFA-DATASETS

Dataset-Name: Name of the file containing the dataset.
Title: Title of the dataset
Version: Any version number associated with the dataset
Revision-Date: Date of last dataset revision
Source: The group or organization providing the source
                 for the dataset. Email or postal addresses must
                 be provided.
Compiled-by: The group or organization responsible for compiling
                 the datasets into the format for which this description
                 applies. Email or postal addresses must be provided.
Size: The size of the dataset in octets (bytes) or records.
                 If records are specified, a record size must be
                 provided as well as other relevant record information.
Format: The format in which the dataset is distributed.
Software: A list of programs used to manipulate the dataset.
                 Contact names and postal/email addresses should be
                 provided if available.
                 The SunSITE administrators will attempt to locate
                 and retrieve non-commercial software if it meets
                 SunSITE standards. Otherwise, pointers to appropriate
                 sites will be provided, if available.

Appendix 5-A

From: jeffk@cyberspace.org (Jeff Kopmanis)
Newsgroups: comp.os.linux.announce
Subject: LSM: Submission Specification (4/1/93)
Date: 2 Apr 1993 18:59:38 GMT
Organization: GREX Public Access Unix +1 313 761 3000
Lines: 174
Approved: linux-announce@tc.cornell.edu (Matt Welsh)
Message-ID: <1pi2eq$d8o@fitz.TC.Cornell.EDU>
NNTP-Posting-Host: theory.tc.cornell.edu
Keywords: Linux Software Map
Originator: mdw@theory.TC.Cornell.EDU

LSM Submission Specification (April 1, 1993)

As promised, I've delivered on getting a standard format up and going for the LSM. It is
basically an enhanced version of the IAFA format currently being debated by the Internet folks.
An example package is given toward the end of the document (thanks to Adam Thornton for a
pretty complete, early-submission). For anyone who submitted before this document came out, I
will convert your entries into the format.

Additionally, at the end of the document, I provide a template for easier submission to the LSM.
Please use the template when possible, as I'll be using its layout features for other purposes.

------------------------------- cut here ---------------------------------

Linux Software Map -- Package submission format

Basically, the LSM submission format is an extension of the IAFA-PACKAGES format
documented in /pub/docs/iafa/draft.part.II on sunsite.unc.edu The extension amounts to the
addition of 3 fields: "Approx-Size", "Last-3-Releases", and "Comments" They follow the
"Keywords" field in the IAFA-PACKAGES format. Also, the "Platforms" field content
definition has been amended to include information for the LSM's purposes. In the format
specification, all LSM-specific enhancements are marked with (LSM).

Amended from IAFA-PACKAGES:
==
====

Software Packages Information

Filename: IAFA-PACKAGES

This file contains records with the following fields. The record is started and delimited from other
records by the "Package-Name" field.

Any number of these files may exist in the archive.

Package-Name: Name of the file or directory containing the package

Title: Title of the package

Version: This field can be used if a version number or string is
                associated with the package

Description: Description of the function of the programs in the package

Author: Name and Email address of authors if available

Maintained-by: This field should be included when the current maintainer
                of the package is known. Contact information should be
                included

Maintained-at: Host name of the "home" of the package if known. This is
                the site at which the most uptodate version of the package
                would be expected to be found

Platforms: Any requirements or restrictions that the package may
                have in terms of hardware or software (OS) platforms.
                The programming language the package is written in should
                be included.

                (LSM) In the above description, "requirements" refers to
                items such as the Linux version required or any utilities,
                libraries or other packages needed for this package to
                run. Version numbers should be included for any non-OS
                requirements to ensure that the package will run.

Copying-Policy: The status of the package for copying purposes.

                The most common entries for this field would be
                "Public Domain" or "Freely Redistributable" or
                "Voluntary Payment" (shareware). However since a record
                may exist for software packages not resident on the AFA,
                this field may be entered as "Proprietary" or some other
                form of restricted access.

Keywords: Keywords appropriate for users trying to locate the
                package

Approx-Size: (LSM) Size of archive to the nearest K. Size in bytes is
                helpful, but not required.

Last-3-Release: (LSM) The version numbers of the last 3 releases of this

package.

Comments: (LSM) Any other pertenant information. PLEASE BE BRIEF.

```
------------------------------------------------------------------------
Example record for the LSM:
------------------------------------------------------------------------
```

Package-Name: linux.tads2exe.tar.Z
Title: Text Adventure Development System (TADS)
Version: 2.0.13 (first release, ported)
Description: C-like object-oriented programming language designed to
        make the design and implementation of classic text
        adventures (a la Infocom) a fairly straightforward
        process
Author: Unknown, High Energy Software
Maintained-by: Adam Justin Thornton (adam@rice.edu)
Maintained-at: ftp.gmd.de: /if-archive/programmings/tads
        wuarchive.wustl.edu: /mirrors/if-archive
Platforms: Linux 0.99 (early March, 1993 SLS release); GCC 2.3.3;
        lib4.3
Copying-policy: Freely Redistributable. Copyright held by author.
Keywords: tads, linux, text adventure, High Energy Software
Approx-Size: 631K (631245 bytes)
Last-3-Release: n/a
Comments: Many TADS games are available at ftp.gmd.de in
        /if-archive .

```
------------------------------------------------------------------------
```

```
------------------------------------------------------------------------
For Multi-Filed Packages:
------------------------------------------------------------------------
```

Because Linux has such entities as the SLS distribution and the /procps series of programs, special considerations must be given to them. An SLS disk should be addressed as a single entity. For example, the SLS disk c2 should have an LSM entry that looks something like this:

Package-Name: disk c2
Title: SLS Distribution Disk C2
Version: 0.99 (January, 1993)
Description: Second disk in the C series of disks. Contains the
        following packages: f2c, gdb, lib42, make, p2c
Author: SLS
Maintained-by: SLS, (insert.Peter's.address.here)
Maintained-at: tsx-11.mit.edu: /pub/linux/SLS/*
        sunsite.unc.edu: /pub/Linux/SLS/*
Platforms: Any ISA-bus Intel 386/486 PC with >2MB RAM and >10MB disk
Copying-policy: Freely Redistributable [please correct me if I'm wrong]
Keywords: SLS, Linux, disk c2, c series

Approx-Size: 1MB (1039605 bytes total)
Last-3-Release: 0.99, 0.98pl5, 0.98 [Not sure]
Comments: All files should be written to an MSDOS diskette and used
           with the SLS install system.

Template file for easy submission to LSM: (start typing ON the _)

Package-Name: _
Title: _
Version: _
Description: _
Author: _
Maintained-by: _
Maintained-at: _
Platforms: _
Copying-policy: _
Keywords: _
Approx-Size: _
Last-3-Release: _
Comments: _

------------------------------ cut here --------------------------------

----
Jeff Kopmanis   jeffk@cyberspace.org
(313) 769-6911 (home)   (313) 393-4784 (work)

--
Send submissions for comp.os.linux.announce to: linux-announce@tc.cornell.edu

Bibliography

1987. Hypertext'87 Proceedings. Paper read at Hypertext'87, at Chapel Hill, NC.

1989. Hypertext'89 Proceedings. Paper read at Conference'89.

2002. *What's New, October 1993* ncsa.uiuc.edu, 1993 [cited March 12 2002]. Available from http://archive.ncsa.uiuc.edu/SDG/Software/Mosaic/Docs/old-whats-new/whats-new-0693.html.

2002. *Welcome to SunSITE!* sunsite.unc.edu, 1993-94 [cited March 23 2002]. Available from http://www.ibiblio.org/sunsite/unchome.html.

2002. *FAQ on the Internet Multicasting Service* info@radio.com (Automatic Response), 1994a [cited March 12 2002]. Available from http://www.ibiblio.org/pub/talk-radio/faq_itr.readme.

2002. *The University of North Carolina at Chapel Hill Teams Up With Sun Microsystems to Launch Public Information Server on the Internet* SUN Microsystems, 1994b [cited March 24 2002]. Available from http://www.sun.com/smi/Success/IndustrySpecific/Education/UnivNCarolina.html.

1995a. Intuition Proves successful at sunsite.unc. *The Sun Observer* Volume 8 (No 5):10, 49-50.

2002. *New at CEARCH Archives* CEARCH, 1995b [cited March 23 2002]. Available from http://www.ibiblio.org/cisco/archives/1995/july1995/whatsnew.july23.html.

2002. *Thank you for choosing SunSITE Classic* sunsite.unc.edu, 1995c [cited March 23 2002]. Available from http://www.ibiblio.org/sunsite/sunsitelocal.html.

1997. newsbrief . . . The biweekly bulletin of computing, instructional and information technology for UNC-CH. , 30 May 1997.

Agre, Philip E. 1998. Designing Genres for New Media: Social, Economic, and Political Contexts. In *CyberSociety 2.0: Revisiting CMC and Community*, edited by S. Jones. Thousand Oaks, Calif: Sage Publications.

Beckett, David. 1995. IAFA Templates in use as Internet Metadata. Paper read at World Wide Web Journal - Proceedings of the Fourth International World-Wide Web Conference.

Berman, Andrea. 2002. [webgroup] 10 years of ibiblio/metalab/sunsite: webgroup@metalab.unc.edu.

Berners-Lee, (attrib) Tim. 2002. *Conclusions* w3.org, 1990a [cited March 8 2002]. Available from http://www.w3.org/Conferences/ECHT90/Conclusions.html.

Berners-Lee, (attrib) Tim. 2002. *Points* w3.org, 1990b [cited March 8 2002]. Available from http://www.w3.org/Conferences/ECHT90/Points.html.

Berners-Lee, Tim. *The World Wide Web: A very short personal history* [cited . Available from http://www.w3.org/People/Berners-Lee/ShortHistory.html.

Berners-Lee, Tim. 2002. *WWW-WAIS Gateway* w3.org, 1991 [cited Feb 10 2002]. Available from http://lists.w3.org/Archives/Public/www-talk/1991NovDec/0002.html.

Berners-Lee, Tim. 2002. *How does www keep track of the available servers?* [web page]. w3.org, 1992 [cited March 6 2002]. Available from http://www.w3.org/History/1992/WWW/FAQ/KeepingTrack.html.

Botafogo, R., and B. Shneiderman. 1991. Identifying aggregates in hypertext structures. Paper read at Hypertext `91, at San Antonio, TX,.

Bush, Vannevar. 2002. *As We May Think* The Atlantic Online, 1945 [cited March 1 2002]. Available from http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm.

Christensen, Steven M. 2002. *Steven M. Christensen and Associates, Inc.* Steven M. Christensen and Associates, Inc.

Scientific Computing Software and Consulting Services, March 1, 2002 2002 [cited March 3 2002]. Available from http://smc.vnet.net/Christensen.html.

Ciolek, T. Matthew. 2002. *Global Networking: a Timeline 1960-1989* ciolek.com, 10 Dec 2001 1999 [cited Feb 25 2002]. Available from http://www.ciolek.com/PAPERS/GLOBAL/1900mid.html.

Cobb, Robert. 2002. *The Mbone (or Virtual Internet Backbone for Multicast IP)* School of Computer Science, Otto-von-Guericke-University of Magdeburg, 1996 [cited March 5 2002]. Available from http://www.cs.uni-magdeburg.de/Mbone/.

Daly, James. 1995. Music by Modem. *Rolling Stone*.

Dempsey, Paul. 1994. Developing a Campus-Wide Information System. Paper read at Middle States Association of Collegiate Registrars and Officers of Admissions (MSACROA) Conference, at Shippensburg University of Pennsylvania.

DeRose, Steve. 2002. *Brown University and Electronic Books* Brown's Scholarly Technology Group (STG), [cited Feb 9 2002]. Available from http://www.stg.brown.edu/%7Esjd/brownChron.html.

Deutsch, P., A. Emtage, M. Koster, and M.Stumpf. 2002. *Publishing Information on the Internet with Anonymous FTP* (Internet-Draft) Internet Engineering Task Force (IETF) and Loughborough University department of Computer Studies, Mon Dec 18 12:56:01 GMT 1995 1995 [cited Feb 12 2002]. Available from http://www.roads.lut.ac.uk/Reports/iafa-draft/node4.html.

Dodge, Martin. 2002. *An atlas of cyberspace; historical maps of computer networks* Centre for Advanced Spatial Analysis, 2002 [cited Feb 18 2002]. Available from http://www.cybergeography.org/atlas/historical.html.

Easterly, Greg. 1994. Busy computer junction at UNC-CH. *The News & Observer*, June 25, 1994, B1.

Engelbart., Douglas C. 1962. Augmenting Human Intellect: A Conceptual Framework. Menlo Park, Ca.: Stanford Research Institute.

Farley, David. 2002. [webgroup] 10 years of ibiblio/metalab/sunsite: webgroup@metalab.unc.edu.

Fenton, Serena. 2002. *How is Ibiblio organized?* ibiblio, 2000 [cited March 10 2002]. Available from http://www.ibiblio.org/collection/indexing.

Filo, David, and Jerry Yang. 2002 March 27. *Yahoo homepage* Bilkent University, 1994 [cited 2002 March 27]. Available from http://web.bilkent.edu.tr/History/yahoo/.

Flor, Micz. 2002. *Bandwidth and the Deceiving Beauty of the Interface* <nettime>, 16 Jul 1997 1997 [cited Feburary 28 2002]. Available from http://amsterdam.nettime.org/Lists-Archives/nettime-l-9707/msg00053.html.

Fullton, Jim. 1990a. Future of BBS'ing: alt.bbs.

Fullton, Jim. 1990b. Opinions Needed: alt.bbs.

Fullton, Jim. 2002. *IAFA-SITEINFO* sunsite.unc.edu, 1992 [cited Feb 10 2002]. Available from SunSITE server: /public/home3/old-ftp.

Furuta, Richard, and John Leggett. 1994. From the Conference Chairs. Paper read at First Annual Conference on the Theory and Practice of Digital Libraries, at College Station, Texas.

Garner, Kellyjo. 2002. *Welcome to UNC-CH's Heliocentric Information Map* ibiblio.org, 1996 [cited March 12 2002]. Available from http://www.ibiblio.org/heliocentric.html.

Garrett, Jesse James. 2002. *A visual vocabulary for describing information architecture and interaction design;version 1.1b* Jesse James Garrett, 6 March 2002 2002 [cited MArch 10 2002]. Available from http://www.jjg.net/ia/visvocab/.

Gates, Dominic. 2002. *The Library of Babel; the Dream of Cyberspace as a Universal Library* pretext magazine, 1997 [cited March 8 2002]. Available from http://www.pretext.com/oct97/features/story1.htm.

Gates, Rick. 1992a. First Internet Hunt !!!: bit.listserv.libref-l 'Discussion of Library Reference Issues'.

Gates, Rick. 1992b. Results: First Internet Hunt, Pt. 2 of 2: bit.listserv.pacs-l.

Gill, Jock. 1994. Electronic Budget Press Release: <jgill@eop.gov>.

glen. 2002. [webgroup] 10 years of ibiblio/metalab/sunsite: webgroup@metalab.unc.edu.

Graef, Gerald. 1997. Book Review: The UNIX Web Server Book. *Linux Journal*, November 01, 1997.

Gray, Matthew. 2002. *Web Growth Summary* M.I.T, 1996 [cited March 24 2002]. Available from http://www.mit.edu/people/mkgray/net/web-growth-summary.html.

Griffiths, Richard T. 2002. *History of the Internet, Internet for Historians (and just about everyone else)* Leiden University, October 4, 2001 2001 [cited February 25 2002]. Available from http://www.let.leidenuniv.nl/history/ivh/chap4.htm.

Grobe, Michael. 2002. *An Early History of Lynx: Multidimensional Collaboration* University of Kansas Academic Computing Services, April 10, 1997 1997 [cited Feb 27 2002]. Available from http://www.ku.edu/~grobe/early-lynx.html.

Halasz, Frank G. 1987. Reflections on Notecards: Seven Issues for the Next Generation of Hypermedia Systems. Paper read at Hypertext87, 1989, at Chapel Hill, NC.

Hallman, Judy. 2002. *Campus-Wide Information Systems* sunSITE

University of North Carolina at Chapel Hill, 1992a [cited Feb 27 2002]. Available from ftp://sunsite.unc.edu/pub/docs/about-the-net/cwis/hallman.txt.

Hallman, Judy. 1992b. Using INFO: comp.os.vms.

Harri.Salminen. 2002. *http://www.nic.funet.fi/index/FUNET/history/internet/en/* nic.funet.fi, 1998 [cited Feb 17 2002]. Available from http://www.nic.funet.fi/index/FUNET/history/internet/en/etusivu-en.html.

IAFA-WG. 2002. *iafa-doc.package* IAFA-WG and sunsite.unc.edu, 16-Aug-1992 1992a [cited March 6 2002]. Available from http://www.ibiblio.org/pub/docs/iafa/iafa-doc.package.

IAFA-WG. 2002. *Publishing Information on the Internet with Anonymous FTP; (IAFA DOC II)* (DRAFT 92.06.10) IAFA-WG, 1992b [cited February 15 2002]. Available from http://www.ibiblio.org/pub/docs/iafa/draft.part.II.

Johnson, Steven. 1997. *Interface Culture; How New Technology Transforms the Way We Create and Communicate*. New York: Harper Edge.

Jones, Paul. 1992a. Re: Soviet archives online: bit.listserv.pacs-l.

Jones, Paul. 1992b. Re: status?: paul jones <pjones%lambada@lambada.oit.unc.edu>.

Jones, Paul. 1992c. Sun archive proposal guidelines: <pjones@mento.oit.unc.edu>.

Jones, Paul. 1992d. Sun GOD needed: support@uncvx1.bitnet.

Jones, Paul. 1992e. sunsite pr announcement: MetaGroup <sungroup@metalab.unc.edu>.

Jones, Paul. 1992f. UNC's bbs moves (samba now lambada): alt.bbs.

Jones, Paul. 1992g. what I'm sending to Graves and Groves: Message-Id: <9202111647.AA15082@mento.oit.unc.edu>.

Jones, Paul. 1993. success story: nprdocs@unc.edu.

Jones, Paul. 1998. End of SunSITE.unc.edu - don't panic: Web Group <webgroup@sunsite.unc.edu>.

Jones, Paul. 2002. *About ibiblio* ibibli.org, October 19, 2001 2000a [cited March 10 2002]. Available from http://www.ibiblio.org/about.html.

Jones, Paul. 2002. *Open[Source]ing the Doors for Contributor-Run Digital Libraries* ibiblio.org, 2000b [cited March 10 2002]. Available from http://www.cni.org/tfms/2000b.fall/.

Kahle, Brewster. 2002. *WAIS Corporate Paper version 3*

An Information System for Corporate Users: Wide Area Information Servers Thinking Machines Corporation, 1992 [cited Feb 22 2002]. Available from http://www.cs.washington.edu/research/projects/ai/590i/wais/wais-corporate-paper.txt.

Kapor, Mitchell. 1990. A Software Design Manifesto. In *Bringing Design to Software*, edited by T. Winograd. Boston, MA: Addison-Wesley.

Kapor, Mitchell, and John Perry Barlow. 2002. *Across The Electronic Frontier* Electronic Frontier Foundation, July 10, 1990 1990 [cited Febuary 28 2002]. Available from http://www.eff.org/Publications/John_Perry_Barlow/HTML/eff.html.

Kinneer, Patty. 1999. UNC-CH's MetaLab and Digital Bitcasting forge multi-media research alliance. *UNC News Service*, March 30, 1999 -- No. 223.

Kleinrock, Leonard. 1961. Information Flow in Large Communication Nets, Electrical Engineering, MIT, Cambridge.

Kling, Rob, and Margaret Elliott. 1994. Digital Library Design for Usability. Paper read at First Annual Conference on the Theory and Practice of Digital Libraries, at College Station, Texas.

Knott, Judson, and Paul Jones. 1996. SunSITE Serving Your Internet Needs Since 1992. *D-Lib Magazine* February 1996.

Kopmanis, Jeff. 1993a. Announcement: Linux Software Map (preliminary): comp.os.linux.announce.

Kopmanis, Jeff. 1993b. First release of the LSM to sunsite.unc.edu: comp.os.linux.

Kopmanis, Jeff. 1993c. Linux Software Map (LSM): Release 1.0: comp.os.linux.announce.

Kopmanis, Jeff. 1993d. LSM Volunteers linux software Map: comp.os.linux.

Kopmanis, Jeff. 1993e. LSM: Submission Specification (4/1/93): comp.os.linux.announce.

Kopmanis, Jeff. 1993f. Re: Suggestion: Current Version List: Kansas City Linux Users Group.

Kress, Henriette. 2002. Re: [webgroup] 10 years of ibiblio/metalab/sunsite: Henriette Kress <hetta@saunalahti.fi>.

LaQuey, Tracy. 1994. *The Internet Companion; A Beginner's Guide to Global Networking*. 2nd edition ed. Rockport, Mass.: Open Book Systems.

Levy, David M. 2002. *Cataloging in the Digital Order* Systems and Practices Laboratory; Xerox Palo Alto Research Center, 1995? [cited March 10 2002]. Available from http://www.csdl.tamu.edu/DL95/papers/levy/levy.html.

Magid, Jonathan. 1992a. Re: ftp sites + demise of banjo.concert.net: Newsgroups: comp.os.linux.

Magid, Jonathan. 1992b. Re: Help: can I get files from sunsite using ftpmail?: comp.os.linux.

Magid, Jonathan. 2002. *wais-on-sunsite.html* sunsite.unc.edu, 1994 [cited Feb. 3 2002]. Available from http://www.ibiblio.org/jem/wais-on-sunsite.html.

Manovich, Lev. 2002. *Digital Constructivism: HCI and Contemporary Style* Venice Biennale, Dec. 15, 1998 1998 [cited MArch 1 2002]. Available from http://absoluteone.ljudmila.org/5.php.

Miller, Robin 'roblimo'. 2002. *iBiblio Takes MetaLab Concept To A New Level* slashdot, 2000 [cited March 23 2002]. Available from http://slashdot.org/features/00/09/17/155240.shtml.

Mohney, Doug. 1990. Future of BBS'ing: alt.bbs.

Molich, Rolf, and Jakob Nielsen. 1990. Improving a human-computer dialogue. *Communications of the ACM* Volume 33 (3 (March 1990)):338 - 348.

Morville, Peter. 1996. Revenge of the Librarians. *WebReview*.

MouseSite, curators of. 2002. *Mouse Site - the Demo* Stanford University, [cited Feb 10 2002]. Available from http://sloan.stanford.edu/MouseSite/1968Demo.html.

Myers, Brad A. 1998. A Brief History of Human Computer Interaction Technology. *ACM interactions.* Vol. 5 (no. 2, March, 1998):44-54.

Nelson, Ted. 1987. One for All and All For One. Paper read at Hypertext87, 1989, at Chapel Hill, NC.

Nielsen, Jakob. 1988. Hypertext'87 Trip Report. *SIGCHI Bulletin* 19 (4):27-35.

Nielsen, Jakob. *Hypertext'89 Trip Report* Jakob Nielsen, 1989 [cited . Available from http://www.useit.com/papers/tripreports/ht89.html.

Nielsen, Jakob. 1994. Enhancing the explanatory power of usability heuristics. Paper read at CHI'94 Conference.

Nielsen, Jakob. 2002. *Report From a 1994 Web Usability Study* useit.com, 1997 [cited March11 2002]. Available from http://www.useit.com/papers/1994_web_usability_report.html.

Nielsen, Jakob, and Rolf Molich. 1990. Heuristic evaluation of user interfaces. Paper read at Conference on Human Factors and Computing Systems

Conference proceedings on Empowering people : Human factors in computing system, at Seattle, Washington, United States.

O'Kelley, Robert. 1999. Technology grant sparks new teaching tools for film. *Chapel Hill News*, 05/01/98.

Pinkerton, Brian. 2002. *WebCrawler Timeline* thinkpink.com, [cited March 12 2002]. Available from http://www.thinkpink.com/bp/WebCrawler/History.html.

Pinkerton, Brian. 2002. *Finding What People Want: Experiences with the WebCrawler* thnikpink.com, 1994 [cited March 9 2002]. Available from http://www.thinkpink.com/bp/WebCrawler/WWW94.html.

Rheingold, Howard. 2002. *Tools for Thought* Howard Rheingold, 1985 [cited March 1 2002]. Available from http://www.rheingold.com/texts/tft/9.html.

Rosenfeld, Louis. 1996. What is Web Architecture and Why Should I Care? *webreview.com* (September 20, 1996).

Ross, Kirk. 2000. UNC's online library to be built by users. *The Chapel Hill News*, September 13, A1.

Savetz, Kevin, Neil Randall, and Yves Lepage. 1996. *MBONE: Multicasting Tomorrow's Internet*. San Francisco: IDG - Hungry Minds, Inc.

Simun, April. 1999. MetaLab, WUNC transmit live broadcasts of Clinton's impeachment, Senate trial. *UNC-CH News Services*, Jan. 29 , 1999 -- No. 73.

staff. 2002. *WAIS databases on sunsite* sunsite.unc.edu, 1994 [cited March 24 2002]. Available from http://www.ibiblio.org/jem/wais-on-sunsite.html.

staff, ibiblio. 2002. *ibiblio.org Collection Criteria* ibiblio.org, last updated: July 3, 2001 2001 [cited March 10 2002]. Available from http://www.ibiblio.org/collection.html.

Troan, Erik. 1993. Linux related WAIS databases available: linux-announce@tc.cornell.edu (Lars Wirzenius).

Tuttle, Marcia. 1991. Newsletter On Serials Pricing Issues. N.S. Number 3 -- June 5, 1991.

Vecchia, Gioacchino La. 2002. [webgroup] 10 years of ibiblio/metalab/sunsite: webgroup@metalab.unc.edu.

Wacker, Stephen. 2001. Art and Technology—notes on creativity in a digital age (thoughts and words on ones and zeroes). *Spark-Online Version* (22.0).

webmaster. 2002. *Automated INFO response for SunSITE* sunsite, Jun 24 1994 undated [cited February 20 2002]. Available from http://www.ibiblio.org/sunsite/sunsiteinfo.

Wedeles, Lauren. 2002. *Prof. Nelson Talk Analyzes P.R.I.D.E.* VASSAR MISCELLANY NEWS February 3, 1965, 1965 [cited Feb 17 2002]. Available from http://iberia.vassar.edu/%7emijoyce/MiscNews_Feb65.html.

Wiggins, Richard W. 1993. Re: WWW vs. Gopher: bit.listserv.cwis-l.

Wilson, Brian. 2002. *Browser Timelines (Releases important to HTML and CSS development)* blooberry.com, 1996-2002 [cited March 9 2002]. Available from http://www.blooberry.com/indexdot/history/browsers6.htm.

Wilson, Trish. 1993. President on-line at UNC-CH. *The News & Observer*, May 30, 1993, A1.

Zakon, Robert H'obbes'. 2002. *Hobbes' Internet Timeline v5.5* zakon.org, January 3, 2002 2002 [cited Feb. 17 2002]. Available from http://www.zakon.org/robert/internet/timeline/.