Who are they and why are they here?
A User Analysis of the Digital Library

Documenting the American South: Beginnings to 1920

With the release of “Episode I - The Phantom Menace” on Wednesday [May 19, 1999], it’s tempting to look back at “Episode IV,” as the original was subtitled, and see an inevitable success, a movie that simply found a perfect moment. But “Star Wars” blasted off in part because of a carefully conceived marketing plan.

“It was the first film I’m aware of – and I started in the movie business in 1968 – that had a complete marketing manual: research flow charts and everything,” says David Weitzner, who ran the movie’s marketing as the Fox vice president for worldwide advertising, and now has his own company in Los Angeles.

Givens, R. (1999, May 14)

Marketing is a powerful tool that is all too often neglected by businesses of all types. The first movie of the Star Wars series is an example of a production that did not ignore marketing, and profited thereby. Its sequels and prequel followed in its footsteps and everyone is currently experiencing the power of well-planned marketing. Thus its example is a good reason for anyone engaged in the business of providing a product or service to take a look at what marketing can offer them. One of the groups that gives marketing the least amount of attention is non-profit organizations. In the minds of many, marketing has a disagreeable connotation; they think of it as crass materialism or pushy salesmanthropathy, with no altruistic motives behind it. But that can be a sad mistake, based on a misunderstanding of what marketing is and what it can do for one.

Kotler and Andreasen (1991) define marketing as the process of “creating, building, and maintaining exchanges...The ultimate objective of marketing is to influence behavior” (p. 38). With this definition in mind, one can understand how marketing applies to non-profit organizations as well as profit-oriented businesses. Whether one
wants to have a customer give money, donate blood, or simply use a proffered resource, the ultimate goal of a marketer is to trigger the desired behavior.

Kotler & Andreasen (1991) describe the four different marketing mindsets that have developed through the course of the Industrial Revolution. The product orientation came first, and is expressed by the aphorism, “If you build a better mousetrap, then the world will come to you.” There is a certain amount of truth in this, but it overlooks the need to convince the world that your mousetrap is better than what it already has. Many non-profit organizations still suffer from this malady. The workers within the organization believe so passionately in what they offer that they cannot understand those who do not. And it may be true that the customers also believe in what the organization offers, but that fact does not animate them to action.

The production orientation came next, with the focus on efficiency. From the moment the assembly line came into existence, and even before then, people have been trying to find out how to do something quicker, faster, better, assuming that if you can make it cheaper, then your profit will accordingly grow. And sometimes it does increase. But this ignores the fact that just because it is cheap does not mean that people will buy.

The sales orientation came along next, remedying the lack in the production orientation: once you make something, you still have to convince people that this product is just what they need and/or want. However, this can turn out to be an exercise in futility, since it often attempts to convince a person to change to suit the product, rather than the other way around. Forcing people to change to suit the product is a certain road to failure.

The customer orientation is the most recent marketing mindset. It is an honest attempt to find out what the customer wants and to satisfy that desire. This mindset starts with the customer, then leads to the product. When used properly, this approach teaches an organization to find what its customers truly desire or need, and seeks to satisfy them within the bounds of what the organization has to offer. Most people have been pleased
with companies that boast of customer service, such as the Saturn Corp.; behind that attitude is the customer orientation toward marketing.

As can be seen from these descriptions, many non-profit organizations are stuck in one or another of the older mindsets, such as the product orientation, believing that if they simply provide a quality or “worthy” product, tangible or intangible, then people will come to use it. But how many times have the most altruistic of organizations fallen by the wayside for lack of interest? It is not because people are not interested in doing good work or obtaining quality products. Rather, it is because the “sellers” did not bother to find out if the customers truly wanted what was offered, or if so, what circumstances would induce the customer to work or pay for their particular product.

All this is to say that marketing is not just valuable, but also necessary to any organization, when employed correctly. Marketing research is a tool that can help one to understand the needs and desires of the people one is attempting to serve. If those needs and desires are understood, then the customers are better satisfied. If the customers are better satisfied, then one’s future holds greater security. One must be careful not to go to the extreme in this, but properly applied marketing can make a noticeable difference in product success. To give this an easy-to-understand library application, that is what librarians attempt to do when they conduct a reference interview – find out what the patron really wants in order to truly provide it, so that the patron walks away satisfied. And a satisfied patron is more interested in supporting the library, thus enabling the library to continue serving its public.

This idea is not precisely new to libraries. In the SILS class, Resource Selection and Evaluation, it is taught that good librarianship requires one to regularly review the collection, the policies that govern selection, and the user community that one serves. Unfortunately, it is something that is often neglected, due to limited resources – an all too common problem – or a misunderstanding of the profound impact that proper marketing research can have. In this research, I intend to put into practice the idea that the first step
in developing an effective marketing strategy is to understand the characteristics of the user population of a library, in this case the digital library *Documenting the American South (DAS).*
Digital libraries in context

In recent years, the Internet has grown phenomenally. Although the Internet has existed since the early 1990’s, it really began to take off in 1996. “A recently released study…found that as of March 1996, 24 percent of the population had Internet access – an increase of 50 percent in just six months” (Frost, 1996, p. 28). More recently, in the Nielsen/Net Ratings’ regularly published statistics of the Internet, there was a total of 96.6 million Internet users, 41.9 million of which are active (1999, April 12). Plus, the Internet is no longer primarily an American phenomenon. There has been a high degree of growth in every continent except Africa.

Along with that growth in the number of users, there has also been a consequent growth in the number of websites. An enormous variety of websites exist on servers around the world. In January 1996, as the Internet was starting to really take off, there were just 9,472,000 domain names (each representing at least one website). By January 1997, there was significant growth, with 16,146,000 names; in 1998, it leapt again, with 29,670,000; and in 1999, 43,230,000. (Network Wizards, 1999) And this is not necessarily an accurate count: “We consider the numbers presented in the domain survey to be fairly good estimates of the minimum size of the Internet. We can not tell if there are hosts or domains we could not locate.” [emphasis in original] (Network Wizards, 1999)

Digital libraries are an integral part of that explosion in growth. Many governments and universities are making serious efforts to place on the Internet a wide variety of quality information in order to make it more easily accessible to the entire world. This should not be a surprise, since for many of them it is simply an extension of the underlying motivation of public service.

At this point, it becomes necessary to define a digital library. In truth the definition has not take its final form yet, but Marchionini (1998) offered a working
definition: “Digital libraries are the logical extensions and augmentations of physical libraries in the electronic information society. Extensions amplify existing resources and services and augmentations enable new kinds of human problem solving and expression.”

To use DAS as an example, the website utilizes an existing resource – the Southern Americana collection of the Triangle libraries – and offers them in a new format, thus extending the already considerable services of the library. The digital library also augments the resource by exercising the powerful capabilities of computers in order to search the collection with new and vigorous forms of analysis. For DAS, this definition works.

At a minimum, the digital library of today is a collection of primary and secondary texts, images, videos, sound clips, etc., available in electronic format. Usually there is a thematic cohesion that guides the selection process of digitization, although the theme can be fairly loose. And finally, the digital library attempts to take advantage of its electronic format to offer more and sometimes better ways to analyze the data so encoded. Some examples would be the American Memory site of the Library of Congress (http://memory.loc.gov/), to which DAS is to be linked eventually, or the Making of America (http://www.umdl.umich.edu/moa/), an example of a digital library that offers the texts through straight text files or image files, attempting to include not only the basic intellectual content, but also the visual data that accompanied it in its original printing. The Victorian Women Writers’ Project (http://www.indiana.edu/~letrs/vwwp/) is an example of a literature site that possesses quality standards and a thematic cohesion equivalent to DAS. Another similar site is the Brown University Women Writers Project (http://www.wwp.brown.edu/project/index.html), which offers the same degree of careful encoding according to TEI standards, as well as a serious attempt to cover a single topic as coherently and comprehensively as possible.

Some of the best known examples of digital libraries are the six projects funded
Indeed, these six might be considered groundbreakers. They have been studied
thoroughly from the moment of their birth, and the research has led to many innovative
techniques for handling a variety of different types of information. For example, the
Alexandria Digital Library (http://alexandria.sdc.ucsb.edu/) is creating ways to store,
manipulate, and search geo-spatial data. With the large variety of visual information that
is available on the Web, the mechanics of how to handle and how to describe it are
extremely debatable, but the ADL is demonstrating and testing the best methods for
doing precisely that. The Informedia Digital Video Library
(http://www.informedia.cs.cmu.edu/html/enter.html) at Carnegie Mellon University
explores the ways in which one can do the same thing for video data. The Stanford
University (http://diglib.stanford.edu/) project is not actually creating a digital library, per
se, but rather exploring the ways in which one can transparently merge many different
collections already in existence into one digital library. The University of Illinois at
Urbana-Champaign is working on collections of scientific literature
(http://dli.grainger.uiuc.edu/) and the University of Michigan is attempting to create a
digital library of earth sciences data (http://www.si.umich.edu/UMDL/), both of them
contending with the problems of encoding, metadata, and searchability. The last of the
six is the Environmental Planning and Geographic Information Systems
(http://elib.cs.berkeley.edu/) at University of California at Berkeley. Interestingly, of the
six, only the last named offers its data free of charge to anyone on the Web. All of the
others are limited in some way. For example, UIUC’s collection is based on scientific
periodical literature which is still under copyright, and therefore cannot be offered freely
without the consent of its owners. Its audience is limited to those connected to the
campus, which would automatically fall into its normal definition of patrons, and
therefore its accustomed users. Others are as yet limited by the software necessary to
handle the data. Thus, although the studies that these innovative projects have performed
are extremely valuable in developing many aspects of the formation of new digital libraries, none of them, save perhaps UC-Berkeley’s project, can approach their user studies from the same aspect as DAS. And they have not yet done so.

When digital libraries started, they were based on a perceived need and a realization that explosive new technology could provide a solution more expansive than anyone’s wildest dreams. The exponential growth of the Internet has more than justified that educated guess (or was it blind faith?), but in the years since the many electronic projects began, there has been no systematic effort to define their changing user population so. And the Internet changes more every day. Although its growth will eventually slow down from its initial exponential explosion, it does not seem to be happening very fast. And now we have millions of people using the Internet on a daily basis and a new generation of users who unthinkingly turn first to computers and the Internet to research those topics in which they have an interest. Knowing this, an effort must be made at every site of a quality digital library to comprehend the users so that one can both expand the user population and respond more adequately to their needs. Marketing research offers one the means to accomplish that goal.
Prior studies of digital library users

Very little marketing research is currently being done on digital libraries or other electronic text websites. The major concerns regarding users at this time seem to be user interface design rather than user population studies. Digital libraries are so new that their creators are just now beginning to concern themselves with a support base. However, there is some work in this area which points out the need and direction for this research in the near future.

One of the earlier efforts to examine the user population of a digital library was the study of the Chemistry Online Retrieval Experiment project at Cornell University. (Entlich, Garson, Lesk, Normore, Olsen, & Weibel, 1996) However, the focus of this study was not truly applicable to digital libraries that are available on the Internet, unfortunately. Although the format of the electronic resource was SGML-encoded text and TIFF images, the content was limited to recently published volumes of science journals, the user population was confined to the university campus (75 total) rather than open to worldwide access, and thus the study design defined a certain population rather than allowing the users to define themselves. The intent of this study was not user population definition but rather an effort to find out what format truly worked with digital information.

One of the more complete and current efforts to study users was conducted by the Alexandria Digital Library. Hill, et al (1997) discuss the various ways in which they attempted to elicit user feedback for the purpose of design improvement. This is a valuable – indeed an inescapable – purpose for user studies; their input is necessary in order to ensure that their needs are in fact being met through the user interface. The scientists were able to learn about users' problems by watching the interaction between the digital library interface and the patrons, and thus were able to make significant improvements in the next version. However, all of their methods were based on
assumptions about the potential user group, which directly contrasted with the eventual intent of making this database freely available to the full Internet population. This did not mean that their assumptions were necessarily wrong, simply unconfirmed. The most unlimited method employed was the beta tester survey, wherein everyone who registered to use the ADL was eventually considered a beta tester. Unfortunately, the response to that survey was limited and, anyway, the information would now be considerably out of date, assuming that the ADL user population has reflected the growth of the Internet population as a whole.

D’Alessandro, D’Alessandro, Galvin, and Erkonen (1998) approached user analysis from a different perspective, making an effort to evaluate usage over time. They conducted a website log analysis of one month (February) out of the year over four consecutive years (1995-1998). From this analysis, they were able to come to some valuable conclusions regarding what content should be included, what changes would increase usage, and how to ensure accessibility of the digital library by a growing community. These types of marketing research conclusions are some of the strongest justifications for regular and thorough user analysis, and this study is an excellent example to follow for any digital library that has the appropriate resources.

In a study of Internet use through a library, Tillotson, Cherry, and Clinton (1995) made some interesting observations about the University of Toronto user population and how and why they were utilizing the Internet as a whole. However, what these authors were studying encompassed the whole range of the Internet and its services, rather than one confined digital library site, and thus is not wholly comparable.

In 1995, Bowen made an early attempt to quantify the benefits of a web page for museum listings and links to individual museum web pages. He has continued to work with this page, and has posted more recent papers regarding his work with it. (Bowen, 1997, Bowen, 1999) These papers are not intended primarily as user analyses but rather as evidence of the benefits of web pages for museums. However, user analysis is a part
of the analysis that he offers, and therefore some access statistics are reported. Also, unlike some of the previous studies, this study is more closely related to humanities websites such as DAS, since most museums have a humanities emphasis. Nonetheless, when looking at this information, it is likewise important to bear in mind that this is a reference page rather than a digital library, thus the way it is used and to a certain extent who uses it are not truly comparable to a digital library such as DAS.

Bates (1996) focused on the computer search habits of humanities scholars visiting the Getty Institute. They discovered that there was a definite difference in how the humanities scholars approached the DRA database. Many search mechanisms are created by people in math or computer science, and thus presumably contain a bias towards the way these scholars think, therefore science scholars find database searching to be easy to learn. The Getty study clearly indicated that humanities scholars are different, not only in the manner in which they search, but also in the vocabulary on which they draw. And while search screens have improved greatly since the time of this study, that does not necessarily provide a solution to the different searching styles, nor solve the problems of the different type of vocabulary employed by these scholars.

Studies of the users of the University of Virginia’s Electronic Text Center (ETC) are most comparable with the current study. Established since 1992, the ETC site offers SGML-encoded text and images of many different humanities texts. Although their approach to selection of materials works very differently than DAS, they are nonetheless a comparable humanities site, and one of the few that offers some statistics on access and the email feedback that they have received over the years. (Seaman, 1997, 1998, 1999) In addition, Seamon, (1997) discusses “the integral place that our user community has in shaping the work of our library-based Etext Center.” Although this process could not work for DAS as it exists now, this kind of relationship does suggest intriguing possibilities for the future, as DAS comes to a better understanding of its user community and what they truly want. Once a relationship is established, then one can create a more
intimate link with the users that guides the future of the website, as the ETC has done.

All of this research looks at digital libraries in a variety of ways without necessarily discovering anything about their user populations. All but one of them approached their populations with certain assumptions. Yet marketing research in many areas has shown that assumptions are dangerous. Kotler and Andreasen offer several examples of managers who were certain they knew the reasons why their organization was having problems, and yet a small amount of research demonstrated that not only were they wrong, but that they had exactly the opposite idea of the true state of affairs. Anecdotal evidence and personal experience had led them down the garden path. Realizing this danger, it throws a shadow on much of the research that is performed in this area. How can the results be truly representative if one is not sure that the tested population is drawn from the actual population? With this in mind, regular analysis of a population of actual users is not only helpful, it is essential.
History & Mission of Documenting the American South

Currently Documenting the American South: Beginnings to 1920 exists at http://metalab.unc.edu/docsouth/. However, it actually began almost a decade ago in the mind of Dr. Patricia Buck Dominguez, the Humanities Bibliographer in Davis Library at University of North Carolina at Chapel Hill (P. Dominguez, personal communication, March 2, 1999). She is particularly concerned with improving the Southern Americana collection, both in selection and accessibility. At Davis Library as well as in the Manuscripts Collection, Southern Historical Collection, Rare Book Collection, and the North Carolina Collection, UNC-CH’s gathering of materials regarding the South is nothing short of magnificent, but large portions of it are available only through on-site visits. And even those items which can be checked out are limited by the rules that define who is a patron and is thereby permitted to borrow the library resources.

In 1989, UNC-CH in cooperation with Duke University and North Carolina State University received a Title IIC grant from the government which was focused on the development of the collection of works regarding the contemporary American South, and allowed the universities to significantly increase these specialized holdings. This grant concentrated Dominguez’s attention on the works available to users in that collection, as well as the problems that it experienced. One of those problems is, ironically enough, a high degree of circulation. Books that were over 100 years old were sometimes circulating six times a year! In her experience as a librarian, this was a remarkably high rate for any resource in an academic library, but especially so for items as old these books. This unarticulated but clearly expressed user need was a primary stimulus for the creation of DAS.

Aside from the libraries themselves, the UNC-CH faculty were a resource of considerable importance for the Dominguez. Accustomed to working collaboratively with them in her development of the monographic collection, she also regarded them as a
user group of primary importance. She was sure that they would welcome any resource, especially one with the potential of this one, that would aid them in their research. Thus she was also accustomed to counting on their aid in obtaining for them the appropriate tools. As proved by their later work and interest, this assumption has been well justified.

In addition to her knowledge of the collections and the local resources upon which she would be able to draw, Dr. Dominguez was also aware of the technical possibilities of digitization, as demonstrated by the pioneering work of the *Archivo de Indias*, the government records of the Spanish conquest and settlement of the New World. Thus, though she did not immediately apply it to her own work, she knew of some of the potential applications for the technology then becoming available to libraries.

All these factors contributed to Dominguez' conception of a digital library of Southern Americana. By 1994, she and Ms. Natasha Smith, currently the Digitization Librarian, had begun the actual work of encoding a text. From that point on, the project began to pick up momentum and has not slowed down since.

In 1996, the availability of DAS was announced (Grendler & Dominguez) along with some other related Web exhibits done by the Manuscripts Department on the Civil War. Since then, the collection has steadily grown. In the first few months of its appearance on the then SunSITE, it had 38 sections, which then jumped to 80 over the next six months, then 104 by the end of 1997, 170 by mid 1998, 274 by the end of 1998, and 458 by the end of May 1999. This is not an actual record of documents digitized, but rather directories and documents on the website. Currently there are 280 documents available, in SGML and HTML, as well as any related images.

The projects that have come from various collaborations have taken several different forms. In 1994, the project managers met and talked with Professor Robert Bain of the English Department, a specialist in Southern literature, who was extremely enthusiastic about the possibilities of this medium, and agreed to put together a bibliography of the 100 most important works in Southern literature, identified by him
and his expert colleagues. That work eventually became the beginning of the *Digitized Library of Southern Literature: Beginnings to 1920* (http://metalab.unc.edu/docsouth/southlit/southlit.html). In 1995, an Ameritech Grant was awarded by the Library of Congress for the *First-Person Narratives of the American South: Beginnings to 1920* (http://metalab.unc.edu/docsouth/fpn/fpn.html), which included both slave narratives and general biographies and autobiographies by other voices, especially minority voices not often heard. Currently the project is working under a grant from the National Endowment for the Humanities to digitize the *North American Slave Narratives: Beginnings to 1920* (http://metalab.unc.edu/docsouth/neh/neh.html) as completely as possible, with the assistance of Professor William Andrews. Also, they are working under a grant from the Institute for Museum and Library Services, for which they are encoding Confederate imprints, as well as another Ameritech grant that focuses on the black church community in the South. These will eventually be the sections called *Southern Homefront: 1861-1865* (http://metalab.unc.edu/docsouth/imls/index.html) and *The Church in the Southern Black Community: Beginnings to 1920* (http://metalab.unc.edu/docsouth/church/index.html).

Through these collaborations, the project has grown enormously, with great success, and clearly does not intend to stop anytime soon.

From its beginning, *Documenting the American South* was intended for educational use. The writers of the initial proposal clearly stated that “The goal of this project would be to provide *researchers and students* with remote access to a wide array of materials in electronic form that they could retrieve and manipulate …” [emphasis added] (Domínguez, Swindler, & Moltke-Hanson, 1992, p. 1). The project has never truly lost that focus. Even today, the purpose statement on the website (http://metalab.unc.edu/docsouth/aboutdas.html) lists students, teachers, and researchers as those for whom it was primarily created.

As a further population focus, in its initial proposal, the immediate user group was
actually identified as the people of the state of North Carolina (Dominguez, et al, 1992). UNC-CH has always had North Carolina's citizens as its primary focus. In fact, DAS was a bicentennial gift from the university library “to the university, the state, and the Hewitt” (Dominguez & Grendler, 1996, p. 9). The North Carolina state focus continues to be important, as can be seen in the next topic choice that Dr. Grendler hopes to put on the website: North Carolina culture and history (M. Grendler, personal communication, April 18, 1999). But in truth, considering the growth of the Internet over the past few years, there was actually no way that anyone could predict who the users would be. They could make educated guesses, but guesses are not facts.

The messages that have come in from email feedback have revealed a surprising variety. There are many who are students and teachers, but there are also many who are attempting to research family history, find out more about slavery, and a multitude of others topics, from all parts of the world. Once again, “customers” have confounded the expectations of those who offer a quality resource. Knowing this, more organized research becomes essential to guide the future of the website.
Research Question

In 1996, the website *Documenting the American South* came into existence on the Internet. It was a part of the Internet revolution that was changing the world. Specifically, it was and continues to be an example of a site that scans and encodes electronic texts in the humanities in order to make these texts available to the world in electronic format.

When *DAS* was first launched, it started small, but steadily grew. Its placement on the UNC SunSITE also helped to make it successful in attracting a rapidly growing number of users. I believe that its success reflected the rapid growth of the Internet, as well as its own efforts to provide a quality product in response to a perceived need. In looking at its access statistics (See Figure 1.), there is a steady rise in the number of hits to the main page from the date of its first appearance on the MetaLab site. However, that is not necessarily an assurance for the future. Other sites also show the same growth, which is not inevitably a result of website quality or value. To believe that is to fall into the fallacy of the production orientation to marketing. The access statistics (not shown here) obtained from the Virtual Library museums page (Bowen, 1999) and from the Electronic Text Center at UVA (Seamon, 1999, See Figure 2.) also show much the same steady rise in hits. As the Internet growth slows down, there must be more than simple
presence to ensure the continued expansion of the user community.

The mission statement of the website makes clear that the original and continuing focus of the project is service to students, teachers, and researchers. When this study first began, the original question started with an attempt to discover if DAS is actually fulfilling that mission. Is it being used by students, teachers, and researchers, and for what purposes are they using it?

However, as the survey was put into practice, it became clear that such a question was rather limited. A sampling of the many email responses sent to the site indicated that there were many more users than simply that small category. Thus, the investigation became an attempt to determine the full spectrum of patrons. In fact, when a certain point is reached in a product’s existence, and the product’s audience reaches a certain size, it becomes necessary to begin to study that population in order to better serve it. DAS has now reached that point. The users of this website must be defined and profiled in order to provide adequately for the future of the website and to ensure that the needs of the users are properly met. Armed with that knowledge, the project could not only justify its existence according to its own mission statement, but expand that mission to

![Figure 2: UVA Etext Center](http://etext.lib.virginia.edu/stats/)
encompass far more.

One must adapt to serve the user population. But that kind of adaptation to your users can only take place when you know who they are! We cannot just sit back and say – without the justification of proof – that a website is used more by young people than older, or that most people are downloading the SGML files for later use, etc. The Internet population has grown and changed so much that it is impossible to use its population as a certain comparison. We need to find out who the DAS population is.

By we I mean not only the project directors of DAS, who need to know for the future of their website development, but also the library world as a whole. Many libraries are involved in the effort to open their collections to the world through the possibilities of a digital library, thus they need to see and do a great deal of research in order to find out if it is worth the effort. Would people go to the website? If so, what subject areas would they go to see? Who are they and how much help do they need? What formats work best? All these questions need to be asked and answered, so that the growth and development of digital libraries in the humanities will not only grow fast, but in the right direction. It is useless to put a great deal of time and effort into something that nobody uses. Documenting the American South took advantage of the fact that at least one of their primary collections received a sufficiently heavy amount of usage that it would benefit from greater accessibility and intelligent preservation. This demonstrated empathic design, “the creation of product or service concepts based on a deep (empathic) understanding of unarticulated user needs” [emphasis in original] (Leonard-Barton 1995, p. 194). The aptness of this empathic design has been demonstrated by the constantly growing number of users who find the site and use it.

From the beginning, the users were offered an opportunity to comment, complain, or otherwise respond to the project managers. The email responses that were sent to the project managers were extremely varied, showing that the user population included young and old, student and retiree, researcher and browser, etc. But simply reading email
is an extremely limited form of assessing the user population. The responses only included information about those who chose to respond and the information these responses included could not begin to answer the questions needed to properly market or improve the product. In this research, therefore, I intend to conduct a survey of users on the website and also study the web transaction logs of the months during which the survey was available. Having collected this information, the survey will then be analyzed and compared to a reputable survey of the current general Internet population, in order to find out who is and is not using this website, and if there are any factors, for example age or experience on the Internet, which affect that level of usage. The analysis will be supplemented with and compared to the analysis of the web transaction logs. Through these methods, a better, more structured idea of the user population of DAS can be achieved. Having done all that, some recommendations can be made for future study and perhaps a few conclusions regarding the future course of the website.
Methodology

When attempting to perform user analysis, there are a number of options. The user evaluation performed by Alexandria Digital Library gave a good example of the variety of methods available for user analysis and the benefits that accrue from using several differing approaches. The current study combined two approaches: an optional questionnaire and web log analysis.

The questionnaire was linked to the website by means of the Feedback hyperlink. Whenever a user clicked on the Feedback link, a webpage came up asking if they would be willing to respond to a survey. At this point they were given three choices: continue on to the survey, give feedback by the usual email option, or return to the website’s main page. If they clicked on the link to the survey, then their consent to be used as a respondent was assumed. The link was in place from April 1- June 1, 1999 and there were 50 respondents during this time period. This was only a small percentage of the users who visited the site: 0.14%. Of those who clicked on the Feedback link, 9.6% responded to the questionnaire. Consequently, the survey data must be regarded as not necessarily representative of all users of the site.

The survey (see Appendix) consisted of seventeen questions on a variety of topics. The most important questions for general demographic identification came first: age, gender, race, and location, as well as educational level and profession. These were followed by several questions about Web use. First, by ascertaining the number of years a patron had been employing the Internet as a resource, one could discover if the users were experienced in the use of the Internet or just discovering its possibilities. The next question focused on what browser the patron utilized. Such a question is important due to the variability between different software packages. The two most popular browsers, Netscape Navigator and Internet Explorer, both have some proprietary HTML codes, which can limit what the DAS website employs if it is to be accessible to everyone.
The next two survey questions attempted to discover the purpose and subject interest of the users. By determining the percentage of visitors who employ the website for research purposes, it would be possible to justify the effort and expense that go into maintaining the quality of this website. The subject interests of the customers would also be important for the future of the website. If the visitors are more interested in certain topics, then the next grant proposals can try to take those interests into account as the project managers prepare them.

The next question attempted to find out how many of our visitors are repeat or first time visitors, and how often the repeaters frequent the site. Two other questions also addressed user satisfaction: one examined the navigability of the website, the other focused on the satisfaction rating of the search mechanism, if it was used.

The next question explored the diverse methods by which the users found this resource. Since publicity is often a valuable marketing tool, it was necessary to discern if it would be helpful, and if so, which method would be most effective.

The final question focused on the locations from which the users were accessing the website. This question would be used in conjunction with several of the other questions to clarify or support some of the study conclusions.

At the end of the survey, the respondents were given the chance to make any kind of comment they desired. This was given in part because the respondents were often answering the survey in lieu of sending an email, and it was therefore necessary to offer them a similar and convenient facility for writing whatever they wished. It was also to be hoped that at least some would offer some other kind of comment, criticism, or recommendation for improvement.

The data from the survey were compared to the surveys of the Internet population done by GVU in spring and fall of 1998 (Georgia Tech Research Corporation). By this comparison, one can discover what differences exist between the DAS population and the general Internet population.
To supplement the data from the questionnaire, a copy of the website logs during the time period that the survey was available (April 1-June 1) was obtained. Using the program WebTrends LogAnalyzer, an analysis on each month (April 1-30 and May 1-June 1) was performed. In order to gain a proper picture of the website activity, all of the background .gif images and several pages, such as the frames and navigation page, that are transparent to the user, were filtered out so that they could be excluded from the analysis.

The analysis of these logs complemented the survey data. The most important is the actual number of unique users (i.e., from unique IP addresses) who visited DAS could be calculated, as opposed to the general number of hits on various pages, which is freely available through the MetaLab site. The IP address and consequently the domain name of the majority of computers that access the site could also be determined. Thus, one could ascertain how many are from the .edu domain, the .com domain, .net, .mil, .gov, etc., including the various country domain endings. This will support the location information obtained in the questionnaire, and also elaborate on which of the Internet service providers are the most common connection suppliers. Other revelations from the log analysis included the number of users per number of visits, most active and least active days, average number of hits per day, summaries of hits by hour of the day, and a great deal of other data.
Results

The results from the survey are reported here, and compared with the pertinent findings from the 9th and/or 10th GVU surveys (Georgia Tech Research Corporation, Spring 1998, Fall 1998). The survey responses totaled 50, although not every single respondent answered every single question. They came at a fairly steady rate, 25 responses each in April and May. This number of responses, however, was extremely small in comparison to the number of actual users: 0.14%. However, it is important to note that this percentage may actually be a trifle higher, since the users identified by log analysis include the workers who regularly update the site and visit it for other reasons not having to do with actual research, both in the Academic Affairs Library and MetaLab. Plus, the total number of unique users was achieved by adding together the April and May totals of unique users, so monthly repeat users would have been counted twice. Additionally, since many of the Internet service providers reassign an IP address every time that a user dials up the service, there are probably some repeat users who are also counted as unique.

The age question was first, with an interesting distribution. In the DAS survey data, age data fell into a bimodal distribution, with 30% of the users falling into their teens and twenties, and 50% between 41 and 70. (See Table 1.) The 10th GVU survey data on general Internet users showed distinctly different data. Their numbers fell into a smooth curve, more symmetrical in shape, with over half in their twenties and thirties.

1 A unique user is identified by counting all log requests by the same IP address within a thirty minute period as one visit by one person.
The genders of DAS users (question #2) were fairly even, with 26 respondents being male (52%) and 21 female (42%). Three respondents did not state their gender. The 9th GVU survey showed that the Internet population was 61.3% and 38.7% female in Spring 1998. The 10th survey (Fall 1998) maintained that division, although with slightly less balance, with 66.4% male and 33.6% female.

The distribution of race (question #3) was as expected. (See Table 2.) In the DAS numbers, almost three quarters (74%) were in the White/Caucasian category, by far the largest amount. African-Americans and Multiracial users were also well represented. The 10th GVU survey showed similar patterns, with Whites dominating Internet use. African-Americans were not as well represented in the GVU results as they were among DAS users.

Table 1. Age distribution of DAS & Internet users

<table>
<thead>
<tr>
<th>Age</th>
<th>DAS Survey</th>
<th></th>
<th>GVU 10th Survey</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percent</td>
<td>Frequency</td>
<td>Percent</td>
</tr>
<tr>
<td>11-20</td>
<td>10</td>
<td>20.0</td>
<td>300</td>
<td>6.0</td>
</tr>
<tr>
<td>21-30</td>
<td>5</td>
<td>10.0</td>
<td>1435</td>
<td>28.6</td>
</tr>
<tr>
<td>31-40</td>
<td>4</td>
<td>8.0</td>
<td>1255</td>
<td>25.0</td>
</tr>
<tr>
<td>41-50</td>
<td>11</td>
<td>22.0</td>
<td>1108</td>
<td>22.1</td>
</tr>
<tr>
<td>51-60</td>
<td>10</td>
<td>20.0</td>
<td>608</td>
<td>12.1</td>
</tr>
<tr>
<td>61-70</td>
<td>5</td>
<td>10.0</td>
<td>182</td>
<td>3.7</td>
</tr>
<tr>
<td>71-80</td>
<td>2</td>
<td>4.0</td>
<td>44</td>
<td>.9</td>
</tr>
<tr>
<td>80+</td>
<td>1</td>
<td>2.0</td>
<td>6</td>
<td>.2</td>
</tr>
<tr>
<td>Not Say</td>
<td>0</td>
<td>0</td>
<td>83</td>
<td>1.7</td>
</tr>
<tr>
<td>Missing (DAS)</td>
<td>2</td>
<td>4.0</td>
<td>6</td>
<td>.2</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>100.0</td>
<td>5022</td>
<td>101.0*</td>
</tr>
</tbody>
</table>

* Total does not equal due to rounding error.

The distribution of race (question #3) was as expected. (See Table 2.) In the DAS numbers, almost three quarters (74%) were in the White/Caucasian category, by far the largest amount. African-Americans and Multiracial users were also well represented. The 10th GVU survey showed similar patterns, with Whites dominating Internet use. African-Americans were not as well represented in the GVU results as they were among DAS users.
The location information came next (question #4). By far the largest group, 44 respondents (88%), came from the United States. (See Table 3.) The 10th GVU Survey showed the same grouping – 84.4% of Internet users in the US; the 9th Survey possessed similar numbers. The log analysis demonstrates the same high percentage from America – 96.1% average of April and May – and a fairly close match between the order and percentages of users from other continents that matches the results from the 10th GVU survey.

Table 2. Race distribution of DAS & Internet users

<table>
<thead>
<tr>
<th>Race</th>
<th>DAS Survey</th>
<th></th>
<th>10th GVU Survey</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percent</td>
<td>Frequency</td>
<td>Percent</td>
</tr>
<tr>
<td>White</td>
<td>37</td>
<td>74.0</td>
<td>4380</td>
<td>87.2</td>
</tr>
<tr>
<td>African American</td>
<td>4</td>
<td>8.0</td>
<td>97</td>
<td>1.9</td>
</tr>
<tr>
<td>Indigenous</td>
<td>1</td>
<td>2.0</td>
<td>16</td>
<td>.3</td>
</tr>
<tr>
<td>Asian</td>
<td>1</td>
<td>2.0</td>
<td>145</td>
<td>2.9</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0</td>
<td>0</td>
<td>63</td>
<td>1.3</td>
</tr>
<tr>
<td>Latino</td>
<td>0</td>
<td>0</td>
<td>32</td>
<td>.6</td>
</tr>
<tr>
<td>Multiracial</td>
<td>6</td>
<td>12.0</td>
<td>82</td>
<td>1.6</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>0</td>
<td>81</td>
<td>1.6</td>
</tr>
<tr>
<td>Total</td>
<td>49</td>
<td>98.0</td>
<td>5021</td>
<td>100.0</td>
</tr>
<tr>
<td>Not Say</td>
<td>0</td>
<td>0</td>
<td>125</td>
<td>2.5</td>
</tr>
<tr>
<td>Missing</td>
<td>1</td>
<td>2.0</td>
<td>1</td>
<td>.0</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>100.0</td>
<td>5022</td>
<td>100.0</td>
</tr>
</tbody>
</table>

The location information came next (question #4). By far the largest group, 44 respondents (88%), came from the United States. (See Table 3.) The 10th GVU Survey showed the same grouping – 84.4% of Internet users in the US; the 9th Survey possessed similar numbers. The log analysis demonstrates the same high percentage from America – 96.1% average of April and May – and a fairly close match between the order and percentages of users from other continents that matches the results from the 10th GVU survey.

Table 3. Distribution of location data for DAS & Internet users

<table>
<thead>
<tr>
<th>Location</th>
<th>DAS Survey</th>
<th>Log Analysis</th>
<th>10th GVU Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percent</td>
<td>Frequency</td>
</tr>
<tr>
<td>Africa</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Antarctica</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Asia</td>
<td>1</td>
<td>2</td>
<td>56</td>
</tr>
<tr>
<td>Oceania</td>
<td>0</td>
<td>0</td>
<td>99</td>
</tr>
<tr>
<td>Europe</td>
<td>2</td>
<td>4</td>
<td>369</td>
</tr>
<tr>
<td>USA</td>
<td>44</td>
<td>88</td>
<td>4254</td>
</tr>
<tr>
<td>Canada</td>
<td>0</td>
<td>0</td>
<td>193</td>
</tr>
<tr>
<td>Mexico</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Central America</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>South America</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Middle East</td>
<td>1</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>West Indies</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Missing (DAS)</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>100</td>
<td>5022</td>
</tr>
</tbody>
</table>

* This percentage was only for April 1999. The rest of the log analysis percentages are an average of April and May 1999. The report did not list beyond the first ten countries, so other areas of the world may have visited the site but are not listed.
The responses to the number of years a person had spent on the Internet (question #5) also covered all the possibilities. (See Table 4.) Over two-thirds of the DAS users had 1-6 years of Internet experience. In the 10th GVU survey, the results were similar: 34.6% were in the 1-3 years group while 37.1% had 4-6 years experience.

<table>
<thead>
<tr>
<th>Internet years</th>
<th>DAS Survey</th>
<th>10th GVU Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percent</td>
</tr>
<tr>
<td>Less than 6 months</td>
<td>6</td>
<td>12.0</td>
</tr>
<tr>
<td>6-12 months</td>
<td>2</td>
<td>4.0</td>
</tr>
<tr>
<td>1-3 years</td>
<td>20</td>
<td>40.0</td>
</tr>
<tr>
<td>4-6 years</td>
<td>14</td>
<td>28.0</td>
</tr>
<tr>
<td>7+ years</td>
<td>4</td>
<td>8.0</td>
</tr>
<tr>
<td>No say</td>
<td>1</td>
<td>2.0</td>
</tr>
<tr>
<td>Missing</td>
<td>3</td>
<td>6.0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>50</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>

Question #6 addressed the browser used. Over half the DAS survey respondents were users of Microsoft’s Internet Explorer: 27 (54%). Next came Netscape Communicator, with 15 users (30%), and another four (8%) for Netscape Navigator. The 10th GVU survey showed the reverse: a 53.7% dominance of Netscape over 31.3% Microsoft’s share of the population.

Educational attainment (question #7) included responses in every category. (See Table 5.) Three respondents had completed elementary school (6%); six, middle school; and four, high school. It is important to note that not all of these are actually in these school levels, but merely that that is the highest degree of education that they have attained, since of the thirteen who gave these grade levels, four gave ages greater than twenty. In the higher education ranges, there was a clear dominance of college or master’s degrees. The 10th GVU Survey ranges showed greater concentration of users in slightly lower levels of education.
The job description list (question #8) offered many choices, and the results were distributed fairly evenly. (See Table 6.) The most common occupations were student (9), writer/journalist (5), and secondary school teachers (4). When occupations are grouped, the education-related professions accounted for 32% of the total. In the 10th GVU Survey, respondents working in the Education industry came to 16% of the total, a percentage even greater than those connected to the computer industry. The log analysis indicates that 26.83% of the April 1999 visitors came from the .edu domain, while there were 37.31% in the .com domain and 31.56% in the .net domain. (17.18% unknown domains were included in the total.) The May totals showed 17.35% in the .edu domain (a decline probably due to the end of the semester in most colleges and universities), 48.19% in the .com domain, 30.82% in the .net domain, and 37.24% in unknown domains.

Table 5. Educational attainment distribution of DAS & Internet users

<table>
<thead>
<tr>
<th>Education</th>
<th>DAS Survey</th>
<th></th>
<th>10th GVU Survey</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percent</td>
<td>Frequency</td>
<td>Percent</td>
</tr>
<tr>
<td>Grammar school</td>
<td>9</td>
<td>18.0</td>
<td>64</td>
<td>1.3</td>
</tr>
<tr>
<td>High school</td>
<td>4</td>
<td>8.0</td>
<td>347</td>
<td>6.9</td>
</tr>
<tr>
<td>Vocational/technical</td>
<td>0</td>
<td>0</td>
<td>157</td>
<td>3.1</td>
</tr>
<tr>
<td>Some college</td>
<td>4</td>
<td>8.0</td>
<td>1430</td>
<td>28.5</td>
</tr>
<tr>
<td>College degree</td>
<td>13</td>
<td>26.0</td>
<td>1701</td>
<td>33.9</td>
</tr>
<tr>
<td>Master's</td>
<td>12</td>
<td>24.0</td>
<td>914</td>
<td>18.2</td>
</tr>
<tr>
<td>Doctorate</td>
<td>3</td>
<td>6.0</td>
<td>189</td>
<td>3.8</td>
</tr>
<tr>
<td>Professional</td>
<td>4</td>
<td>8.0</td>
<td>169</td>
<td>3.4</td>
</tr>
<tr>
<td>Total</td>
<td>49</td>
<td>98.0</td>
<td>51</td>
<td>1.0</td>
</tr>
<tr>
<td>Missing (DAS)</td>
<td>1</td>
<td>2.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>100.0</td>
<td>5022</td>
<td>100.0</td>
</tr>
</tbody>
</table>
The purposes for use of DAS (question #9) came in all possible varieties. The largest group was personal research, with 27 replies (54%) in this category. Nine (18%) named an educational purpose and seven (14%) named academic research. Another five (10%) stated they were visiting for curiosity/browsing purposes, and two did not offer a reason for coming to the site.

All but one of the choices for subject interests was selected more than once (question #10), as well as some extra responses. (See Table 7.) The most common were the Civil War and Confederacy with 12 responses (24%) and literature with 9 (10%) responses. The “Other” topics given included genealogy or family history, plantation life, and slave clothing.

Table 6. Distribution of job description in DAS Survey

<table>
<thead>
<tr>
<th>Job title/description</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Civil servant</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Military</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Admin/secretary</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Health care</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Counselor</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Religious</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Student</td>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td>Writer/journalist</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Service (food)</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Primary teacher</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Secondary teacher</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>College+ teacher</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Retired</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Other</td>
<td>14</td>
<td>28</td>
</tr>
<tr>
<td>Missing</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>50</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

The purposes for use of DAS (question #9) came in all possible varieties. The largest group was personal research, with 27 replies (54%) in this category. Nine (18%) named an educational purpose and seven (14%) named academic research. Another five (10%) stated they were visiting for curiosity/browsing purposes, and two did not offer a reason for coming to the site.

All but one of the choices for subject interests was selected more than once (question #10), as well as some extra responses. (See Table 7.) The most common were the Civil War and Confederacy with 12 responses (24%) and literature with 9 (10%) responses. The “Other” topics given included genealogy or family history, plantation life, and slave clothing.
The ways in which people found the site (question #12) were varied. (See Table 8.) Over one third arrived via a search engine. There was a certain amount of overlap in the choices of actively searched and used a search engine. These respondents used a variety of search engines: Altavista (4 times), Yahoo (3), Netscape (2), Infoseek (2), Hotbot (1), Google (1), Excite (1), CEO Express (1), and two unknowns. No one search engine dominated. Nine users came to DAS from other websites. The referenced websites given included www.ipl.org and www.cnn.com, etc., and several webpages that were not given or could not be remembered.

Table 7. Distribution of subject interest in DAS Survey

<table>
<thead>
<tr>
<th>Subject interests</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>South</td>
<td>4</td>
<td>8.0</td>
</tr>
<tr>
<td>Civil War/Confederacy</td>
<td>12</td>
<td>24.0</td>
</tr>
<tr>
<td>Slavery</td>
<td>5</td>
<td>10.0</td>
</tr>
<tr>
<td>Literature</td>
<td>9</td>
<td>18.0</td>
</tr>
<tr>
<td>Women's studies</td>
<td>4</td>
<td>8.0</td>
</tr>
<tr>
<td>Autobiographies, etc.</td>
<td>5</td>
<td>10.0</td>
</tr>
<tr>
<td>History</td>
<td>4</td>
<td>8.0</td>
</tr>
<tr>
<td>Other</td>
<td>4</td>
<td>8.0</td>
</tr>
<tr>
<td>Missing</td>
<td>3</td>
<td>6.0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>50</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>

The frequency of use (question #11) showed a surprising ascendancy by the newcomers: twenty-five – literally half – named themselves as first-time users of the website. (See Table 9.) The web log analysis extends this data, making it clear that the proportions in the entire population of DAS users are somewhat different. In April, 86.61% were one-time visitors and in May 86.5%. The rest were repeat visitors, who

Table 8. Distribution of finding method for DAS users

<table>
<thead>
<tr>
<th>Finding method</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass media</td>
<td>4</td>
<td>8.0</td>
</tr>
<tr>
<td>Word of mouth</td>
<td>8</td>
<td>16.0</td>
</tr>
<tr>
<td>Actively searched</td>
<td>9</td>
<td>18.0</td>
</tr>
<tr>
<td>Search engine</td>
<td>17</td>
<td>34.0</td>
</tr>
<tr>
<td>Another website</td>
<td>9</td>
<td>18.0</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>2.0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>48</strong></td>
<td><strong>96.0</strong></td>
</tr>
<tr>
<td><strong>Missing</strong></td>
<td><strong>2</strong></td>
<td><strong>4.0</strong></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>50</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>
came back anywhere from twice that month to more than ten times, a total of only 13.28% in April and 13.5% in May of the website patrons. However, these numbers may be skewed by the dynamic assignment of IP addresses to the customers of Internet service providers like AOL or Mindspring. Thus, the proportion of recent users may be higher than the web log analysis implies.

**Table 9. Distribution of frequency of use DAS Survey**

<table>
<thead>
<tr>
<th>Frequency of use</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st time</td>
<td>25</td>
<td>50.0</td>
</tr>
<tr>
<td>1 a month</td>
<td>7</td>
<td>14.0</td>
</tr>
<tr>
<td>1 a week</td>
<td>3</td>
<td>6.0</td>
</tr>
<tr>
<td>2+ a week</td>
<td>10</td>
<td>20.0</td>
</tr>
<tr>
<td>Daily</td>
<td>2</td>
<td>4.0</td>
</tr>
<tr>
<td>Total</td>
<td>47</td>
<td>94.0</td>
</tr>
<tr>
<td>Missing</td>
<td>3</td>
<td>6.0</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>100.0</td>
</tr>
</tbody>
</table>

The access points (question #13) included all the possibilities. (See Table 10.)

The largest number access the Internet from home rather than work or school. This makes it more difficult to identify numbers of unique users and domains, due to dynamic IP addresses.

**Table 10. Distribution of where people access the Internet from DAS & GVU Surveys**

<table>
<thead>
<tr>
<th>Where accessed from</th>
<th>DAS Survey</th>
<th></th>
<th></th>
<th>9th GVU Survey</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percent</td>
<td>Frequency</td>
<td>Percent</td>
<td></td>
</tr>
<tr>
<td>Work*</td>
<td>10</td>
<td>20</td>
<td>7046</td>
<td>55.9</td>
<td></td>
</tr>
<tr>
<td>Home†</td>
<td>24</td>
<td>48</td>
<td>4964</td>
<td>39.4</td>
<td></td>
</tr>
<tr>
<td>School‡</td>
<td>9</td>
<td>18</td>
<td>361</td>
<td>2.9</td>
<td></td>
</tr>
<tr>
<td>Distributed/mobile work place (GVU)</td>
<td></td>
<td></td>
<td>51</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>Mostly from a friend’s home(GVU)</td>
<td></td>
<td></td>
<td>28</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>Mostly from a public terminal</td>
<td>2</td>
<td>4</td>
<td>66</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td>75</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>Missing (DAS)</td>
<td>5</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>50</strong></td>
<td><strong>100</strong></td>
<td><strong>12591</strong></td>
<td><strong>100</strong></td>
<td></td>
</tr>
</tbody>
</table>

* GVU data includes responses to two items: “Work only” and “Primarily work/school but also home”. GVU data includes responses to two items: “Home only” and “Primarily home but also at School only” on GVU Survey.

Navigation ease of the website (question #14) was relatively easy for the majority
of respondents. Twenty-five stated that it was “quick and easy”; 15 said they only
needed “a little bit of exploration to find their way”; one found it “confusing”; and six
“couldn’t find what they wanted”. Three did not respond to the question.

The site’s search mechanism (question #16) was evaluated by 22 of the
respondents (i.e., those who had used it). Of those, eleven said that it was “Easy”; seven
said it was of “Medium” difficulty; one said it was “Difficult”; and three said they found

The comments were varied, some of them containing a reference question, most
praising the website, and a few criticisms, mostly related to the search mechanism. A
few of them, intriguingly, did not just ask a question, but were interested in starting a
dialogue on a research topic related to the website subject matter.

To further analyze the data, Spearman correlations were calculated to discover if
there were any statistically significant relationships among the variables. The results of
these analyses are reported in Table 11. There is a moderate relationship between age
and level of educational attainment. However, the strongest relationships are between
Internet years and Frequency of use and Internet years and ease of navigation of DAS.

\textit{Table 11. Most significant Spearman correlations of DAS survey}

<table>
<thead>
<tr>
<th></th>
<th>Spearman correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>correlation coefficient</td>
</tr>
<tr>
<td>Age * Education level</td>
<td>0.286</td>
</tr>
<tr>
<td>Age * Internet years</td>
<td>0.202</td>
</tr>
<tr>
<td>Internet years * Frequency of use</td>
<td>-0.423</td>
</tr>
<tr>
<td>Internet years * Navigation ease</td>
<td>-0.467</td>
</tr>
</tbody>
</table>
Discussion

In the comparison of the survey results to the GVU survey, it is necessary to note that the GVU survey was not a perfect comparison. Its own recommendations preclude that, since they make it clear that the GVU survey “has a bias towards the experienced and more frequent users” of the Internet rather than the newcomers (http://www.gvu.gatech.edu/user_surveys/survey-1998-10/). This makes the comparison imperfect. However, in searching the Internet for survey research on its users, I was not able to find any other that was so comprehensive in scope, having as its intention a study of the general Internet population. All the others that I located attempted to focus on one segment of that population rather than the population as a whole. At a later date, a narrower focus may become desirable as the DAS users become explicitly defined, but at this point, it is too early for such comparisons.

The first question of interest is the gender distribution of Internet users. In the GVU commentary on the 9th Survey results, the researchers noted that the figures reflected a recent, noticeable rise in the number of females on the Internet that is a result of the newcomers to the Internet. “For the first time, we see a category of users which has more females than males – users who have been online for less than a year (51.7% female, 48.3% male).” (http://www.gvu.gatech.edu/user_surveys/survey-1998-04/#exec) Considering the growing pervasiveness of the Internet in society, this is not all that surprising. The newcomers responding to the 10th survey had an almost even split between males and females, but the males return to being the majority. The DAS survey respondents also showed more men than women, but to a more balanced degree than the GVU survey – 52% male and 42% female.

These trends in the gender of Internet/DAS users may also reflect the disciplinary background of the users. “During the [Internet’s] ‘Bronze Age’, which ended with the invention of the World Wide Web in 1992, the typical ‘netizen’ was easy to describe. He was a highly educated, comfortably-off, English-speaking white male around thirty years
old, living in a town or city.” (Boukhari, 1998, p. 43) The majority of the Internet population was then male, reflecting the science orientation of its creators and users. But nowadays that population is showing a greater balance of sexes, as more and more people from all walks of life become familiar with it, not just those with a science background. Thus, this greater gender balance in the DAS users may demonstrate that the new wave of users includes a higher percentage of users with a humanities focus. This could imply several things in regards to growth of the DAS website: that these patrons were less aware of the power of computers, that they might need more direction in how to use the website, and that there were many of them that were as yet unaware of its existence. As the remaining data are scrutinized, further information may clarify the truth or not of this alternative possibility.

We will next deal with age. The general population on the Internet shows a fairly
smooth bell curve, with the largest section of the population in the 21-30 years range. (See Figure 3.) As the previous quote described, this was the group which first entered the Internet and thus it is only natural that they should still be its largest section of users. Interestingly, the DAS survey responses were distinctly different. (See Figure 4.) Instead of a smooth curve, there was a dip in the middle. The second largest group was in the 11-20 range with a dip in the 21-30 range that continued in the 31-40 range. The percentages then increase, once again dropping off in the above sixty groups. The low number of DAS users in the 21-40 age range is the most intriguing discrepancy between the two profiles. It may be possible to explain it by evaluating the age groups according to educational attainment. Those who research the DAS website tend to possess a college degree or higher. The majority of those who do not are students in primary or secondary school. Thus, there is a higher number in the younger and older age brackets. The

![Figure 4. Age distribution of DAS Survey](image-url)
statistical analysis supports this theory somewhat, since there is a statistically significant relationship between age and educational attainment. However, this relationship is not sufficient to explain the dip in the 21-40 age brackets, since many in that range do possess a college degree or higher. Perhaps the differences between the age graphs reflect the fact that the website is attracting more Internet beginners than experts. If the most common and experienced Internet group is in the 21-40 age category, and DAS is attracting a smaller amount from this category but more from the younger and older categories, then perhaps there is justification for supposing that DAS has more newcomers than experts. This hypothesis is also bolstered by the fact that the average age of the “newcomer” GVU respondents (i.e., those with less than 1 year of Internet experience) is 41.4.

Both surveys were in agreement that the huge majority of users are in the White or Caucasian category. This undoubtedly reflects the majority population of White/Caucasian Americans, since most Internet users are from the US. The most recent US census figures show that 82.1% of the population is White, while only 12.9% is Black. However, there was a greater percentage of African-Americans looking at this website when compared to the small Internet population percentage, as well as a higher percentage of multi-racial individuals that name African-American as part of their heritage. Bearing in mind the focus of the website, which includes North American Slave Narratives and is now adding documents from the existence of the Confederacy, this higher percentage is to be expected and confirms the idea that more African-Americans are likely to be interested in a subject area that so closely concerns their history.

Geographic location of the users was most often from the US. This is fully in agreement with the GVU surveys. The variations in the percentages from other areas of the world simply reflect the small size of the survey. The log analysis proportions also are consistent with the DAS survey, though the log does show an even higher proportion from the United States. Since the subject matter focuses on the American South, this is
not surprising, for though the emails have shown that many people across the world have an interest in the literature of this region, there is no doubt that most of those who are interested are Americans.

The level of prior Internet experience among Internet users appears to be shifting. The 9th and 10th GVU surveys show a difference in their graphs, with a switch taking place between the largest group on the Internet. The 1-3 years category has given way to the 4-6 years group. “This shift toward more experienced users (on average) is consistent with other results showing that the growth of the Web (in terms of percentages) is slowing down.” (http://www.gvu.gatech.edu/user_surveys/survey-1998-10/tenthreport.html#ex) The Internet population is maturing. Interestingly, the DAS survey showed a strong similarity to the 9th survey, the greater number of users falling into the 1-3 years group, with the 4-6 years next. Nonetheless, there was one significant difference, in that there were quite a few DAS users in the brand-new category, with under six months on the Internet. This is, once again, in accord with the suggestion made earlier, that this reflects the growing Internet population that is interested in humanities topics, as well as the growth of sites devoted to such research in the liberal arts.

One of the few technical questions on the survey asked about which browser the patrons utilized. Reflecting the current market prevalence of Microsoft and Netscape, their browsers were the dominant programs for the DAS users. Nevertheless, there was an interesting divergence from the general Internet population. The DAS patrons predominantly used Internet Explorer over the two primary Netscape products – 54% for IE to 38% for Netscape – which does not echo the findings of the GVU survey. Both the 9th and 10th GVU surveys showed a significant dominance of Netscape over Microsoft’s share of the population. In searching for a reason for this discrepancy, a closer examination of the DAS data demonstrates that it is the new users that cause the imbalance. Of those that are in the under 6 months and the 6-12 months groupings, all but one use IE, whereas the Netscape users are all in the more experienced Internet users
category. This is confirmed by a recent survey that shows Microsoft’s IE to be the top browser in the new users section of the Internet, while Navigator is used more by the older inhabitants of the ’Net (Edupage, 1999, June 7). Since the GVU surveys state that their own biases tend to favor the older users of the web, the higher proportions of Netscape users is made clear. At this point in time, it seems appropriate that the DAS website continue to be designed to be readable by either of these leading browsers.

The educational attainment levels found in the GVU survey and the DAS survey were fairly close, although some disparity did raise its head. For example, a noticeably higher proportion of the respondents to the DAS survey had a Master’s degree than those in the Internet population. One possible conclusion is that scholars and researchers are in fact employing the website for its intended purpose, since such people would be more likely to possess such degrees and to pursue such research. Also, there was a higher fraction of users who had only completed grammar school, as opposed to the GVU survey, which indicated there are more high-schoolers on the ’Net. A variety of factors may account for this; one suggestion is that in secondary school there is a more controlled approach to the study of literature. High-schoolers are more likely to be assigned “great literature”, easily obtainable in hard copy, than the literature of minority voices, which is found in such great abundance on this website. Thus, a secondary school student might not need to use this site as much as those in grammar school, who are less restricted in the literature they read for book reports and other assignments. However, much more focused research is needed before any such conclusion could be drawn.

While both DAS users and Internet users at large represent diverse occupations, it is interesting to note the prevalence of educators and students. The DAS log analysis also indicates that a large percentage of .edu visitors came to the site during both months, although larger amounts came from both the .com and .net groups. On the other hand, the survey responses indicate that 48% of the respondents accessed the DAS site from home, which would indicate that some educators are likely included in the .com and .net
groups. With the evidence of these numbers, it is safe to say that the DAS is fulfilling its mission of service to educators and students.

Unfortunately, the 10\textsuperscript{th} GVU survey changed the options on the job description question enough that it is not truly comparable. For example, there is no student option in the response list, thus losing the possibility of comparing those numbers, because they are now included in other categories. However, there was a significant rise in the percentage of educators in comparison to the previous survey of users on the Internet. Considering the high amount of recent Internet users coming to the DAS website, perhaps this more similar proportion in the 10\textsuperscript{th} survey supports the idea that a significant amount of these new visitors to the Internet are humanities educators, in line with the previous idea that many of the newcomers are not necessarily science-oriented. Unfortunately, the GVU surveys did not make a distinction between science teachers and humanities teachers, so there is no way to be sure of this. Also, the significant number of 11-20 year olds in the university educators group leads one to speculate that many college students as well as teachers chose this as their primary occupation.

DAS users’ purposes for visiting the website were varied. Unfortunately, there is some difficulty in interpreting the responses to this question because the differing purposes overlap somewhat. For example, if a patron is studying a historical personage, like Frederick Douglass, he or she may regard it as either personal research or education, depending on the viewpoint. Although 18\% are using it for education, and another 14\% for academic research, more than half of the visitors – 54\% – are performing personal research. The various comments that the respondents made indicated interests in genealogy, literature, and history. The implication that people regard the Internet as an all-purpose reference desk is reinforced by the large number who actively conduct research at sites like this. Also, there is a consistent theme in several of the email responses that come to DAS: they ask for help. It is clear that the users regard the site as the extension of a library that continues to offer the electronic version of traditional
services, and not just the texts themselves.

At this point in time, academic researchers are apparently the smallest proportion of serious patrons. To continue to attract academics, and to increase their numbers, it is necessary to offer a site that produces and maintains high quality texts. Although there are quite a few sites on the Web that offer electronic texts, very few of them that I explored made the kind of detailed effort at quality control that DAS does. If the users are in fact using the website for scholarly research, as was intended from the beginning, then the expense of quality control is extremely necessary. Since the mission of DAS is to offer a resource for reference, studying, teaching, and research, it is extremely important to know if people are using it for those purposes.

The topics in which the patrons expressed interest were also varied. Although there are many factors that go into the content decisions for the site, nonetheless an unarticulated but clearly demonstrated user interest in books of a certain subject area is what initially stimulated Dominguez’s proposal to create a digital library. Clearly stated user interests can be of even greater importance for the website’s future selections. The largest number of respondents were studying the Civil War/Confederacy. However, all but one of the topics listed made an appearance in the responses. This is important, since the list of options was obtained from the metatags on the main page of DAS. These are the keywords which will bring up this URL from a search engine that catalogues DAS. Thus, this spread of subject interests demonstrates the fact that these searches are functioning correctly, at least in a small sample. However, there were a few responses that fell into the Other subject category. Also, some respondents listed more than one topic. With this in mind, it might be a good idea to experiment by adding some new words to the metatag list, reflecting those other interests (e.g., plantation life, genealogy), in order to see if visitor numbers increase as a result of better indexing by search engines. There are many subjects that could be studied at the DAS website that are not obvious at first glance. The metatag list, like any other part of a webpage, needs regular review for
maintenance and improvement.

The accuracy and currency of the metatag list is particularly important in view of the information regarding how the visitors found the website. More than a third – 34% – found the website through the use of a search engine. Another 18% stated that they actively searched for a website such as this one. The answer list on the survey did not clarify these two responses, thus it is quite possible that there was an overlap between the two, and many in the actively searched category would also have fallen into the search engine category. Thus, it is abundantly clear that the findability of the website through search engines is vital.

As for the other responses, there was an even spread between word of mouth and another website as the information source, with the mass media coming in a poor last. However, once again there may be some overlap, since one respondent stated another website as the source, but the website in question was www.cnn.com, which could also be classed as mass media.

This question of how frequently the users come to the site is of considerable importance, since repeat users are usually satisfied users. Thus a high percentage of repeat visitors is an encouraging sign of patron satisfaction. Remembering the previous percentage of newcomers to the Internet in the DAS survey, it is not surprising that 50% of the respondents were first time visitors. However, the other 44% (6% did not respond) are repeat users of varying frequencies. From this, it is clear that DAS is collecting a good number of consistent patrons that will sustain it. The web log analysis extends this data, casting some doubt on the high proportion of repeat users. Based on the web log analysis only about 13% of the users could be identified as repeat users. These repeat users came back anywhere from twice a month to more than ten times a month. The most frequent repeat users most probably include a large amount of the MetaLab and DAS workers. The only clear conclusion to be made is that a large majority of visitors only come once. The question to be investigated is why. The Spearman correlation between
years on the Internet and frequency of use is a possible indicator, but the reason for the relationship is unclear. Evidently additional research is needed.

Data about the location from which people accessed the website did not offer much information by itself, but rather supplemented replies to prior questions. However the fact that almost half the of the DAS users accessed the site from home does point to one concern which was not addressed in the survey, namely how long it took to download the various pages and texts of DAS. In the web log analysis, there is an unmistakable preference for the HTML over the SGML pages. It seems likely that accessing the site from home is limiting the functionality available to users.

The two questions on navigability and searchability of the site are extremely significant, since all the work that has gone into this website goes to waste if the patrons cannot find what they are looking for. In general, the users seem satisfied, with 50% saying that they were able to navigate the site quickly and easily, and another 30% finding their way with only a little exploration. However, the site is not perfect, since another 14% found it confusing or impossible to find their way. Those users that had lesser experience on the Internet were the ones that expressed difficulty in navigating the site, as indicated by the Spearman correlation between the two variables. Nevertheless, even that cannot be taken as the only factor of importance, since several who were newcomers were able to navigate easily through the site and at least two of those with several years’ experience did find the site extremely confusing to navigate.

More than half of the respondents – 56% – did not evaluate the site’s searchability, most probably indicating that they did not use the search mechanism. However, of the other 44%, there was a clear trend: 18 of the 22 stated that the search mechanism was Easy or of Medium difficulty. This small number of responses indicates that the search mechanism is not used a great deal, presumably because people are able to find what they need without resorting to the search mechanism. The high number of Easy or Medium ratings make it clear that of those who do use it, the majority are satisfied.
On the other hand, the actual mechanism is fairly limited in the kind of overall analysis that it can do online, thus limiting the users. As yet, this is not a problem; however, as the user population grows more experienced and sophisticated and the collections expand, the search mechanism’s ease of use will undoubtedly grow in importance.
Recommendations for DAS development

Although a survey was the only direct contact with DAS users that was possible at this point, it is only a starting point. More thorough, regular, and detailed surveys that concentrate on specific areas of interest, as well as personal, in-depth interviews, and other types of data collection will be necessary in the future to pursue the general observations that came out of this questionnaire. Many of the questions used in this study are general in nature, and could be usefully augmented with more specific questions.

This research was undertaken with the idea that decisions to improve the site and increase the customer base could only be made with a knowledge of the users, for whom the site was created, and their needs. So, having looked at the data, what conclusions can be reached about marketing to these customers?

First, there are certain populations, specifically linked to the targeted user group of researchers, teachers, and students that are coming to the site. However, that group is as yet small. It is among university professors of the humanities, and teachers of both primary and secondary school, that immediate publication of the availability of this site needs to be pursued. “The importance of outreach in particular has been confirmed by a number of user studies. Librarians will need to play a strong role in promoting awareness of new DL services since lack of awareness was a major reason for lack of use” (Bishop, 1996, p. 353) Since North Carolina is the first target population of importance to this university, it would be natural, appropriate, and feasible to spread the news among North Carolina educators first and foremost. They could then be made the site’s ambassadors to other educators in all areas of the country and the world.

The project managers have not ignored publicity in their efforts to support the digital library. Unfortunately, their efforts have not been focused on specific populations or specific methods. It has been a general, dispersed effort that has garnered good results, as evidenced by email responses that cited the various newspaper articles that announced the appearance and benefits of this electronic collection. But their efforts have not been
focused on target populations, based on their mission statement or any kind of marketing research.

This publicity could be done in many ways with no cost to the program. For example, if DAS worked with the College of Education here at UNC-CH, the teachers who are now graduating could easily be sent out already knowing about this resource. Any of the educators’ organizations and associations that exist in the state would doubtless be happy to include information about so potentially rich a resource in their newsletters and workshops, wherever appropriate. Any professional development workshops that the university puts on for North Carolina teachers could include this website among their list of resources for the humanities. For example, one resource for North Carolina teachers that already includes DAS in its list of links is the LEARN NC site (http://www.learnnc.org/). But that is an extremely passive form of publicity. More active forms are required. The academic departments that study the relevant topics could easily include this website both in their orientation sessions, webpages, and classrooms, not only here at UNC-CH, but at all the state universities of North Carolina. As the survey responses made clear, there are many who are just now learning about the site, and others who probably still do not know, and therefore are not yet using it. How many more visitors would there be if the news was spread throughout the state? How many more people would they tell?

This type of publicity does not have to be limited to the state of North Carolina. One of the publication methods that was of use, as indicated by the survey replies, was other website referrals. Plus, there is the fact that one of the primary advantages of the Internet is ease of communication. Although I would not suggest adding to the spam already cluttering up email boxes, there are other ways to spread the word, such as discussion list announcements in humanities listservs or bulletin boards and searching out referral websites to recommend that they add DAS to their list of links. The possibilities of spreading the news by means of the Internet are extensive, and all it takes is some time
to search them out. Since the resources are not widely available, simply asking the DAS student assistants to spread the word in their normal Internet wanderings would probably reach a surprisingly large number of people. Marketing publicity does not have to cost a great deal.

Another area of the population that the survey responses point to is the African-Americans category. A larger percentage of African-Americans visited this site, when compared to their percentage of the general Internet population, which is only natural considering its topical focus. With this in mind, here is another section of the Internet population that could be informed of this website, and thereby increase the customer base. With some help from people such as the UNC-CH Department of African and Afro-American studies, it would be easy to publicize the existence of DAS to a section of the population who would doubtless be very interested in this previously muted part of their history.

The other reason mentioned for marketing research is the improvement of the site. Much of the research that is going on in digital libraries is concerned with this side of the question, thus an active pursuit of the literature and comparison of other sites would doubtless offer a great many ideas for improvement. However, such changes must be made with the customer satisfaction in mind, thus this type of research, individualized to the website, must balance out the proffered ideas. The respondents, in general, indicated that they were extremely happy with this website and what it had to offer them. However, there were some, a small percentage, but nonetheless important, who stated that the search mechanism and the ease of navigation were unsatisfactory. This clearly points out the direction for future research and improvement. In fact, this is already a concern for the project managers, who have recently overhauled the search mechanism and are looking toward future study. Smith, the Digitization Librarian, has also stated that an overhaul of the site’s design is in the near future of the website, since with the addition of the two most recent grants, it is in the process of growing to the point that its
current incarnation will soon be unwieldy (personal communication, April 4, 1999). A simple addition to the website – a site map – might help some of those who find the navigation difficult at this stage, and at least work as a stopgap measure until the site design has been overhauled for even greater ease of use. A narrative explaining the site’s current design on the site map might also help those who work better with words than pictures. To help with the problems of the search mechanism, again, an explanation of how to use it, and its current limitations, might help the users to achieve more than they currently can. This is especially necessary in consideration of how that page has changed to take advantage of the SGML encoding, with which many people are not truly familiar.

The only other improvement that might be added is somewhat problematical. Some of the email messages and the comments included in the survey responses indicate that the users often come with questions that the site alone does not satisfy. Sometimes they email those questions through the Feedback mechanism, clearly expecting the website managers to respond. In fact, the librarians do respond when they can, forwarding the messages to the appropriate reference librarians. At this point in time, this is not an unsatisfactory method, but it could easily become one as the users’ numbers increase, thus swelling the number of questions that come in. Forwarding or answering one or two questions per week is not too much to handle, but one or two questions per day can easily become wearisome. It would be advisable to anticipate this problem and design a process to handle it, both in its current form and in the future as the website population grows. In this way, the patrons never find a time when they need to complain.

This potential problem can be turned into an advantage. Several comments and email messages have asked for general responses, not necessarily asking specific reference questions but rather showing a desire to start a dialogue regarding what interests them. The Internet is used by many people to reach out to others with shared interests, as can be seen in the many listservs and bulletin boards and chat rooms based on shared interests. The future of digital libraries may be not only the providing of old
resources in new ways, as described by the current definition, but a new concept of a library. Marchionini (1999) suggested that, as these new shapes take form, there will need to be a new definition, a *sharium*, which

is all the things that a library is today, but also adds strong sharing components. First, creating facilities and tools to support ephemeral sharing of time and expertise during the collaborative problem solving that takes place in physical libraries is a grand challenge for extension and amplification in digital libraries. The second type of sharing that must be supported is not typical in physical libraries and allows individuals to contribute physical objects and materials to the collection. This physical sharing suggests one possible augmentation effect of DLs. (Marchionini, 1999)

With a vision for the future, *DAS* may be a part of that revolution.
Future Research

There are quite a few areas in this research which would bear further, more detailed investigation. One area of importance is the ease of navigation and study of the habits of users in the site. Web logs could track user sessions to find out what are the most consistent entry and exit pages and frequent user pathways through the site, and use this data to improve the site design. Another subject that could use detailed analysis is the searching mechanism and its user satisfaction rating. In recent weeks, the search mechanism has been greatly improved to take advantage of the SGML encoding. It could still benefit from continued upgrading, but the process has begun, and its benefits need to be studied. Most importantly, DAS needs to know if the users are satisfied with the search mechanism, because their dissatisfaction rating may stimulate the process of improvement.

Another question that needs to be studied is the subject interests and purpose of those who come to the site. As mentioned before, these two things will contribute to the decisions about what items are digitized and how detailed are the quality controls in the process. To enable scholars to use the site in their work, they need to feel confident that it is a superior product. But that level of work is expensive and must be justified to those who provide the money. Evidence of consistent use by scholars, researchers, and students will go a long way toward providing that justification. Related to this subject also is the study of how frequently the patrons use the SGML files as opposed to the HTML versions. Web log analysis could provide a considerable amount of information here, but would need to be supplemented with more personal data, such as what types of analysis the advanced users need or why they might prefer the HTML document.

Other areas that would benefit from further research and affect future decisions about the website are the findability of the site through the search engines and what method of publication worked best to bring more people to the site. Both of these topics were touched on in the current questionnaire, but the responses were extremely
inconclusive. More detailed surveys, log analysis, direct interviews, etc., should be used to elaborate on the minimal conclusions from this research.

Since the search engines were a primary source of users, it would be very interesting to search on all the top search engines and find out what search words actually produce the DAS URL, and how high up in the hit list it shows. If the findability is improved with a better metatag list and regularly maintained registration with all relevant search engines, who knows how many more people might come this way?

Another interesting area for further inquiry would be to look into the Internet media source in more detail. The CNN site has already been mentioned; others that the respondents mentioned included a discussion list announcement and several reference page websites. The reference websites could be discovered through a regular analysis of the referrer logs (not analyzed in this study). The main limitation is that not everyone would go directly to DAS from these websites. However, the large majority of users probably would jump directly, using the hyperlink, thus the referrer log analysis would undoubtedly find a majority of these sites and one could discover exactly how valuable they are to DAS.

Another area that web logs would illuminate is browser usage. The web logs can record not only what browser is used, but also which version, which provider, and which operating system a patron employs. With website analysis, this information can be easily discovered and tracked on a regular plan. Armed with this information, DAS can make plans for the future regarding such things as the use of XML and more advanced searching mechanisms.

A final area that would benefit from web log analysis and other forms of investigation is a study of download speeds, especially for patrons that access from home. This study could be linked to research into the preferences for HTML over SGML files. If the SGML files offer more flexibility in analysis, why are so many more turning to the HTML files? Do they not yet require the flexibility of SGML or is it because the SGML
files are more unwieldy? There are many possibilities that need to be explored.

The benefits of regular, systematic user analysis are legion. This research shows what a very limited analysis can do. With regular marketing research as a guide, the project can widen its patron base, establish a groundwork of consistent and regular use by a broad spectrum of people, and ensure customer satisfaction. The foundation of happy patrons will lead to continued growth, enabling the website to justify its existence and thus continue to serve the public, which is its primary reason for being.
Acknowledgements

This research has been long and exhausting and certain people need to be thanked for providing me with what I need to accomplish it.

- Dr. Barbara Wildemuth, who recommended the GVU surveys to me when I mentioned my desire for a comparison with the most complete and scholarly survey done on that topic. I would like to thank her for that, since it gave me very much what I wanted.

- Natasha Smith, the Digitization Librarian who is in charge of the technical aspects DAS. Once the appropriate authorizations were given, it was she who kindly made the changes necessary to link up my survey to the website and then later take it down. Without her agreement and help, this would not have been possible.

- Dr. Patricia Dominguez, for her recommendations in the creation of the survey questionnaire.

- Bruce Sigmun, a friend who offered encouragement and intelligent comments.

- Janie Polutta, my mother, who has supported me throughout this degree.

- And everyone who prayed for me. There were quite a few!
Bibliography


Bibliography

*Edupage.* (1999, June 7) URL http://listserv.educause.edu/


Frost, M. (1996). That was the year that was. *HRMagazine, 41*, 28.


Bibliography


Bibliography


Appendix

Demographic info:

1. How old are you?
   - 6-10
   - 11-20
   - 21-30
   - 31-40
   - 41-50
   - 51-60
   - 61-70
   - 71-80
   - 80 and above
   - Rather not say

2. Male or female?
   - Male
   - Female

3. What is your race? (Note: if more than one is applicable, please check them both.)
   - White/Caucasian
   - African-American
   - Hispanic
   - Indigenous/Aboriginal Person
   - Asian/Pacific Islander
   - Latino
   - Rather not say
   - Other:

4. Where are you located?
   - Africa
   - Antartica
   - Asia
   - Oceania (New Zealand, Australia, etc.)
   - Europe
   - USA
   - Canada
   - Mexico
   - Central America
   - South America
   - Middle East
   - West Indies
5. How long have you been using the Internet (including using email, gopher, ftp, etc.)?

- Less than 6 months
- 6 to 12 months
- 1 to 3 years
- 4 to 6 years
- 7 years or more
- Rather not say

6. What type of browser are you most likely to be using?

- Lotus Notes (Lotus/IBM)
- Hot Java (Sun)
- Internet Explorer (Microsoft)
- Navigator, stand-alone version (Netscape)
- Navigator, as part of Communicator (Netscape)
- Lynx
- Other:

**Educational info:**

7. What is the highest level of education you have attained? (Note: This does not necessarily mean completed. For example, if you are currently in high school, for example, put high school, not middle school.)

- Elementary
- Middle school
- High School
- Vocational/Technical School
- Community/Some College
- College
- Master's
- Doctoral
- Professional
- Other:
8. If you are not in school, what is your current job description?

Civil Servant
Homemaker
Military
Administrator/Secretary
Health Care Worker
Counselor
Religious Occupation
Student
Writer/Journalist
Artist/Musician
Performer
Designer
Service Industry Occupation (including food service)
Teacher (primary)
Teacher (secondary)
Teacher (college and above)
Temporarily Laid Off
Unemployed. Looking for Work
Unemployed, Not Looking for Work
Retired
Other Position
Other Position (if not mentioned above, please give a brief job title or description):
open text reply

Website info:

9. What is your purpose in using this website?

Educational
Academic research
Personal research (of any sort)
Curiosity/browsing
Other:
Other:

10. What subject are you researching? (Check or fill in as many as are applicable.)

the South
Civil War/Confederacy
Slavery
Literature
Women's studies
Autobiographies, diaries, memoirs
(Southern) History
Popular culture
Other:
Other:
Other:
11. How often do you use this website?

This is my first time
Once a month
Once a week
Several times a week
Daily

**Searching info:**

12. How did you find this website?

Word of mouth
From another website
Actively searched for a site like this
Mass media (newspaper, radio, tv, etc.)
Through the search engine:
Other:

13. From where are you accessing it?

From home
From school
From work
From a public terminal (e.g. a library)
Other:

**Website commentary:**

14. How easily did you navigate this website?

I found what I needed quickly and easily.
I found what I needed with a little bit of exploration.
I found what I needed eventually, but the site was confusing.
I don't understand how this site is set up and still cannot find what I need.

15. Did you use the search mechanism on this website?

yes  no

16. If you did use the search mechanism, how would you rate it?

Easy
Medium
Difficult
Impossible
Not applicable

17. Do you have any recommendations for changes and/or improvements in the website? (Feel free to elaborate.)
Who are they and why are they here?
A User Analysis of the Digital Library
Documenting the American South: Beginnings to 1920

by
Melanie Polutta

A Master’s paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in Library Science.

Chapel Hill, North Carolina
July, 1999

Approved by:

____________________________________
Advisor