Editor's Note: After editing this publication for more than four years, David Frederickson has moved on to pursue new endeavors. David’s fine editorial and design sensibilities brought this magazine to a new level, providing a new focus, a new layout, and even a new name: Connect. We will miss David’s hard work, his editorial insight, and his puckish sense of humor.

I am honored and excited to step into his shoes in editing this issue of Connect. We have gathered some interesting articles for you to mull over during the summer months, including a guide to affordable digital editing systems, a look at using VRML in the sciences, and a discussion of new search tools for NYU Web. You will find an update on Computer Store offerings and a schedule of summer classes offered by ACF. Perhaps most exciting to us here at ACF is an introductory article from our new Assistant Director for Humanities Computing, Lorna Hughes.

We wish David the best in his new ventures, and hope he finds this issue lives up to the high standard he set for the magazine.

—Melissa Whitney
The above messages probably look very familiar to you. Every day we all receive unsolicited paper mail offering goods or services unrelated to our interests or needs. Many of us find that most to almost all of our mail is "junk mail." In my own case, probably ninety-five percent of my mail falls into this category. I generally throw it out without opening it.

What is unusual about the text above is that it was not delivered by the Postal Service, but rather was delivered to me over the Internet. This is a small sample; the amount of such junk e-mail that I receive has been growing steadily. Some of these messages are sent directly to my e-mail address; others come through one of the mailing lists in which my name appears. To add insult to injury, a portion of these unsolicited mailings are for programs and services that would allow me to send my sales pitch (if I had one to send) to thousands of recipients. Some go even further, promising that my e-mailed advertisement will be untraceable, sometimes making the assurance that hiding the origin of the message is completely legal.

Perhaps it was inevitable that junk mail would make its way to the Internet in this fashion. After all, the protocols underlying the Internet and related computing systems have an implicit level of trust that reflected the small group environment and cooperative culture within which the Internet was initially nurtured. And as participation in the services of the Net has increased over time, many known aspects of behavior have migrated to the Net, some exciting, some discouraging.

When networking was in its initial stages of development, it offered exciting possibilities to those who wanted to use computers more efficiently and effectively. As network technology and deployment progressed, the benefits of networking were increasingly apparent first in research and then in instruction. Researchers join newsgroups to share ideas with col-
leagues they may never meet in person. Professors and students supplement class hours and office hours with the informal twenty-four hour contact of e-mail.

Those of us involved in this discovery and implementation process felt that we had a precious tool with which to increase the effectiveness of the academic processes. That is why what we see happening over the Net is so distressing. We see a medium that has the capacity to unite many diverse communities in the world and deliver valuable services to the developing nations starting to become glutted with unsolicited commercial services.

More than mere annoyance
The cost of paper junk mail turning up in your mailbox is greater than the annoyance of the trash can filling up more quickly. The costs of producing the paper itself are borne by the sender, but the price of mailing is borne by us all. Third class mail is heavily subsidized by the U.S. Postal Service. Advertisers are mailing at a discount subsidized by other users of the postal service. Additional sorting costs are absorbed in my case by the Courant Institute, in whose building I work.

Internet junk mail has similar costs which may not at first be obvious. The current “all you can eat” pricing structures characterizing much of the Internet favor users who send very large amounts of information. We need to move rapidly toward a pricing model for Internet services that prices, \textit{inter alia}, by volume, so that users will bear the full costs of their network utilization directly. While there are some conceptual problems in allocating full costs that remain to be worked out, an imperfect solution may be better than none. A solution sooner rather than later may be preferred so that current expectations do not have time to spread and behavioral changes are induced before the current behaviors take too strong a hold. In addition, such a pricing policy should result in a flow of funds toward those areas of the Net where additional investment is needed to cope with demand.

The increase in electronic junk mail damages the core usefulness of the Internet for exchanging ideas. If the present trend continues and junk becomes a significant part of traffic on the Net, then the network equivalent of Gresham’s law will take place: the bad traffic will drive out the good. People will see the Net as clogged with commercialism. We will have in large measure crippled the potential and usefulness of a tool that has enormous promise, but that is being cheapened by the uses to which it is put.

The Net is now highly useful as a communications tool in part because it is relatively unpolluted. The ratio of unwanted mail is still low in comparison to the overall volume, and the delete key is still easy to press. But when the ratio of unwanted mail becomes high, then even the ease of hitting the delete key will not cancel the frustration of having to open and view so much material in which we have no interest.

We need to update our laws and institutional policies so that they can cope in appropriate ways with the effects of various behaviors on the Net. Even though the need for additional “cyberlaw” is probably small, we need to come to a common understanding of how our current legal and policy structures apply to cyberspace. We need to apply these rules in such a way as to preserve and enhance the socially valuable uses of cyberspace and not allow them to be whittled away through frivolous or malicious use until the Net becomes a useless resource to the community as a whole.

We need to understand as a global community the potential destructive effects of the diminution of responsible behavior on the Net. Any society, electronic or otherwise, depends for its continued usefulness on its members abiding by common conventions, and most people in most societies do just that. It is the tiny minority at the extremes who threaten the whole. At one edge, there are marauders from the outside, attempting to profit at the expense of the community, and at another, there are potential despots and conscienceless (or merely thoughtless) functionaries within, stifling the contributions members of the community might make.

As the popularity of the Internet has taken off in recent years, we have seen most forms of human behavior, good and bad, migrate to the Net. In retrospect this is not unexpected, since the active character in the transaction, the human being, has not changed. Just as we need to be vigilant in real life, establishing structures of law and custom that allow us the freedom to live effectively, we need to be equally vigilant in the new world of cyberspace if we are to preserve and enhance the promise it offers to so many people and institutions throughout the world.
Inexpensive Digital Video Systems: Basics For Buyers

Bill Horn
william.horn@nyu.edu

Editing video is traditionally a complex and expensive undertaking, frequently out of the reach of individuals and small organizations. This is changing with the growing popularity and affordability of digital non-linear editing (NLE), a technology which is bringing powerful editing tools to an ever-widening group. Many at the university who have never before considered purchasing a digital video editing system are now giving it serious thought.

Today it is difficult to find a professional level film or video production that does not pass through a digital editing phase at some point in its creation. This dominance of the professional post-production market and the explosion of CD-ROM and Web-based multimedia uses for video have dramatically expanded the market for these products. The capabilities of systems are continually improving, and manufacturers are making scaled down products available at consumer level prices (if you call $700-$1,000 for a digitizing card a “consumer level” price).

Unfortunately, the world of digital video can be confusing, even for experienced computer users. The purpose of this article is to give an overview of the issues involved in buying a desktop digital video system, and to report briefly on the ACF’s Arts Technology Group’s (ATG) research into inexpensive digital video systems.

Bill Horn is a Senior Arts Technologist in the ACF’s Arts Technology Group. He received his MFA from the School of Visual Arts in 1994. Bill has worked in post-production on projects for PBS, HBO, Bravo and other commercial venues and has taught at Brooklyn College, Rutgers University and the School of Visual Arts.

Why Invest in a Digital Editing Solution?

In the simplest terms, digital, non-linear systems have done for video editing what the word processor did for typing. In traditional video tape editing one does not have the option of simply splicing a shot into the middle of an existing tape. In order to insert the shot, you have to re-edit every shot after the new one, not unlike having to retype an entire paper in order to squeeze a couple of extra paragraphs into the middle. With an NLE system, you can add a shot anywhere and have the rest of your program shift over automatically to make room, not unlike inserting text with a modern word processor.

NLE systems offer much more than simple editing functions. Tools for producing professional titles and making fancy transitions between shots are now standard fare even on modest systems. If you are willing to pay for extra features, you can purchase additional software that allows you to do advanced compositing and create 3D effects. All these tools were once limited to the professional “on-line” (i.e., expensive) editing suite. Having a professional editor’s toolkit at your disposal is appealing to the small-budget producer, but before rushing to buy an NLE solution, there are certain questions you should consider.

Do I really need a digital editing solution?

Despite the appealing features cited above, there are still many people for whom a digital, non-linear video editing system is not the best solution. The first question to consider is: What do you want to accomplish with the system? If your needs are limited to cutting a few shots together into a simple video document of...
an experiment or compiling a set of scenes to show to a class, a simple deck-to-deck editing system may be the best solution, particularly if your video material is lengthy (more than thirty minutes). Such traditional analog editing setups can be significantly less expensive at the lower end of the scale and are considerably easier for the novice editor to learn than a comparable digital system.

If you are convinced that an NLE system is what you need, the question of what you want to accomplish still applies. There are four basic parts to this question. Considered together, they will help you identify the right system for your needs.

• How and where will the final work be viewed: on video tape, as a Web movie, on CD-ROM? Who does the video quality have to satisfy: a network engineer, a thesis committee, or just some friends and family?
• How long will the videos be, and how much material must be available from which to edit?
• Who will use the system? Will there be many users of a single system? Will they have computer or editing experience?
• Do you envision these answers changing in the next two to three years?

Each of these questions, in turn, raises certain technical issues in putting together a system. Keeping them in mind as we look at the basic components of an NLE system will help you focus on the most appropriate hardware combinations for your needs. First, however, we will examine the issues which most often plague, if not paralyze, potential NLE buyers.

Buy now or wait for the technology to improve?
With the rapid development of new technologies in the computer industry it is tempting to postpone your purchase until the next new thing comes to market—and there is always something new on the horizon. The image quality of today’s video cards and the speed of current hard drives, for example, far exceeds what was available only a couple of years ago, and at the same, or even lower, prices.

The more important point to consider, however, is that if you wait six months or a year to buy your system, that is six months or a year that you will not be making video. If you are ready to make video now, then now is the time to buy. As with the rest of the computer market, you will have to accept that no matter when you buy your system, it will be super-seded, or may become obsolete, within a few years. This applies to everyone—even the most farsighted buyer.

Macintosh or Windows PC?
The Macintosh was once the platform of choice for the digital video market. This is no longer true. The Windows/PC combination now provides a viable alternative, though Macintosh still has some clear advantages in certain instances. At the middle and the high end of the market, Windows NT provides a serious challenge to the Macintosh. NT has received positive reviews for many media applications, including video. As for Windows95, the professional consensus seems to be that it is a substandard operating system for video work, though there are low-end products available for Windows 95 that work reasonably well.

The important factors in choosing a platform and operating system are software access, familiarity of use, compatibility, and ongoing development of products.

Some software packages are still only available for the Macintosh, such as Adobe’s After Effects, the professional standard for video compositing. If access to this software is essential, then the Macintosh is still the standard.

If you already have experience trouble-shooting, installing, and working on one particular operating system, you may want to stay with that platform. Consider how long it took you to acquire the platform-specific skills which you have. Learning a second operating system may be easier than learning the first, but it can still represent a significant investment of time.

You must also consider which other systems will need to share your files and which systems store previous work. The integration of the QuickTime standard for media files in the Macintosh system software is an advantage for this type of media sharing. QuickTime is also available under Windows, but you will have to obtain it separately.

A year ago, it seemed that Windows NT was the hot operating system for new development, yet NT still has fewer video products available, and the most aggressive development of new products is still being done for the Macintosh platform. All this makes the Macintosh more attractive in general, though there are good reasons to work off of WindowsNT, or even Windows 95, not the least of which may be that you already own a suitable PC.
How much will it cost?
This, of course, is the big question, and it is often where people start the process. The short answer is that it will cost as much as you are willing to spend. The systems discussed below start at less than $6,000 for a complete low end setup; there is virtually no limit on the high end. The best solution may be for you consider your budget, calculate your needs based on the four questions listed above, and then decide if NLE is a worthwhile investment. If it is, set yourself a dollar limit and get as much system as you can for the money, tailoring it to your particular needs.

Turnkey or Mix-and-Match?
Some products come bundled together as turnkey systems: storage, capture card, computer, and software, all sold as a single unit. This has distinct advantages in terms of setup, compatibility, and service. With a turnkey system, you know from the outset that there will be no unexpected conflicts between products by different manufacturers, or difficulties in getting the software to talk efficiently to the hardware. Theoretically, all these things are worked out in advance by the manufacturer. All you need to do is plug it in and turn it on. Calls to technical support are also easier because the person on the other end is familiar with how the system should be configured.

These advantages come at a premium. There are relatively few turnkey systems available at the low end of the price range. There may be fewer options for designing the system to meet your specific needs. Buying your system in stages, by getting a couple of pieces now and adding more later, may not be an option. You may also run into problems using the system for applications other than video editing without upsetting the turnkey system’s potentially delicate configuration, though this can also be an issue with mix-and-match systems.

Visit the Arts Technology Group Web Site
http://www.nyu.edu/acf/atg

The ACF Arts Technology Group is a discipline centered unit which focuses on digital technologies used by artists. The ATG Web site provides up-to-date information on facilities and projects, including NYU’s participation as a charter member of New Media Centers.

What makes up a system?
The main tasks of an NLE system are to move video information into and out of the computer realm and, once inside, to move that information around rapidly within the computer. This is a difficult task and requires the help of some specialized hardware. At the minimum, a non-linear editing system consists of the following elements:

- video capture card (and possibly an audio card)
- high capacity, high speed storage (and possibly a SCSI card)
- off-the-shelf computer (CPU)
- editing software
- computer monitor
- video monitor
- video source deck(s)

These elements, or some variation of them, are part of almost any NLE setup, from the most humble desktop system to the most expensive commercial suite.

Many people, even those with considerable computer experience, are baffled by the new vocabulary involved with NLE. It is important to have an idea about what these components do and about the terminology that you will encounter. Once you are familiar with the “dialect,” you will be able to ask better questions and avoid being misled by a technospeak sales pitch.

The Capture Card
Bringing video information into the computer realm is often referred to as digitizing or capturing video. A specialized digital video card, called a capture card, is installed in the computer to handle this conversion. Capture cards often do more than capture, and are sometimes more than a card. Some systems use more than a single card; others use an external box which connects to the card. Most cards also do the work of getting your video back out of the computer so it can be recorded to tape, and some cards can be harnessed to reduce rendering times for certain effects. All capture cards, however, share the basic job of taking an analog video signal, like that produced by your VCR, and converting it into digital information.
Compression
The capture process involves more than grabbing a sequence of video frames. The capture card also compresses the digitized video into smaller pieces so that the volume of data does not overwhelm the rest of the system. Except at the very highest professional levels, digital video is always compressed. In an uncompressed state, video files are huge, on the order of 27MB for each second of video. That is equivalent to twenty floppies per second of video, or a full Zip disk of information for four seconds of video. The most highly compressed movies are those delivered on CD-ROM. A CD-ROM movie typically contains about 150K (0.15MB) of information for each second of video—about 1/200th of the uncompressed rate. Compression rates can fall anywhere between these two extremes.

To accomplish the necessary compression, the capture card uses a special algorithm known as a "codec" (for COmpression/DECompression algorithm). Some codecs work with your system software, such as QuickTime, while others use their own unique systems for organizing video and audio information. Almost all codecs require that the capture card be present in the system in order to play back your files. This is particularly significant if you intend to move your video files among locations or systems.

Video Quality
Video quality depends, in part, on the data rate your system can handle; the higher the data rate, the higher the video quality. Typical maximum data rates produced by capture cards range from 3MB-5MB/second for inexpensive cards up to 7MB-9MB/second for more expensive cards. Cards that consistently deliver 3MB-5MB/second are referred to as VHS/S-VHS quality cards; S-VHS being a step above the quality of a home VCR. Values above 5MB/second are called Betacam quality. Betacam is the de facto standard for professional production, though this is changing as new digital tape formats take over the professional market.

The data rate of your system is also dependent on the disk's ability to keep up with the flow of information produced by the capture card (see High Capacity High Speed Storage below). If your capture card can not deliver as much information as you are requesting, or if your drives can not keep up with the speed of the capture card, you may "drop frames." The card basically skips a particular frame and goes on to the next one, leaving a stutter in your digitized video which may also throw off audio synchronization.

You can compensate for dropped frames by capturing fewer frames per second, capturing a smaller video window, or increasing the amount of compression. These changes decrease the flow of information through your system (sometimes called throughput), but they limit your options for how the video will be used in the end.

The VHS/S-VHS and Betacam quality designations serve as decent generalizations, but they can be misleading. All else being equal, the higher the data rate, the lower the amount of information discarded, and the better the picture. However, each system's codec and signal processing hardware also affect image quality, so a 3.5MB/second clip captured from one system may look as good as a 4.5MB/second clip captured by another. It is a good idea to look at the test images in trade magazines or, better yet, go see a system in action at a friend's studio, a production facility, or a showroom.

Multimedia and Off-line Work
With analog video signals, image size is not a consideration: an analog signal will fill the screen on any size video monitor. With digital video, however, you must specify the size of the clips when you capture them. For tape output, full-size frames are considered to be 640x480 pixels. High-end systems will reliably capture full-size frames at 30 frames/second. Many low-end systems are approaching this stand-

---

1 Each second of analog video is made up of 30 frames (the standard in North America) and each frame, in turn, is divided into two fields. This is why you will often see advertisements claiming 30 frames/second or 60 fields/second capture rates.

2 Most of the cards mentioned at the end of this article use some form of motion-JPEG compression, also known as M-JPEG compression. With newer versions of QuickTime system software for the Macintosh, and soon for the PC, there are standard codecs called Motion JPEG-A and Motion JPEG-B. You should be aware that these codecs are not directly supported by most capture cards. You have to recompile your files into one of these codecs in order to take advantage of the portability they offer, and even then you will likely have difficulty getting quality playback. This may change as capture cards are redesigned to work directly with these codecs.

3 Pixels are the colored squares that make up most digital images. Just to confuse things, some new video systems use a special non-square pixel resolution of 720x486 to get more detail in the image. It leads to a lot confusion for first time users, but does deliver higher quality.
standard, but not all of them are there yet. If you are putting your final videos back out to tape, your system should supply full-size capture at 30 frames/second, and as much quality as you can afford. If you are producing for the Web or for CD-ROM, a card that works best at quarter-screen size (320x240 pixels) or one that provides lower image quality can be appropriate. Professional multimedia producers often capture full screen video at the highest quality and then down-sample it to a smaller size. The quality gained with this method is relatively small, however, and may not be worth the extra expense for the card and the drive space.

Finally, if you are editing large amounts of material, you may want to consider using your system as an “off-line” editing tool. Off-line is a term from traditional video editing that refers to a lower quality edit used to make most or all of the editing decisions. The final version of the video is then re-edited on a high quality, “on-line” system. The simplest way to accomplish this is to combine an NLE system with professional video deck and deck control software. This combination allows you to incorporate the time-code stamp of professional tape formats into the editing process. Once you have worked out your decisions with low resolution material, the timecode lets you directly recapture the material at a higher quality, or to output an edit decision list (EDL) for assembling the final version on a traditional linear tape editing system. These techniques are dependent on your system’s ability to reliably handle audio synchronization and to avoid dropped frames.

Audio Capture

Some capture cards handle audio capture as well as video capture. By capturing both audio and video with the same card, the digitizing hardware can make sure that the video and audio signals are properly synchronized as they come in. Audio information brought in through a different card or through a built-in audio port on the computer will need to be synchronized by the CPU and your digitizing software. This process is prone to error—the CPU may have a difficult time keeping track of what is going on with the capture card and may interrupt the card in order to find out, causing dropped frames or loss of synchronization.

The synchronization issue is particularly important when working with video clips longer than a minute or two. For longer work, it is worth the additional expense to get a capture card with high quality built-in audio. Most audio capture solutions now allow 44kHz/16bit (CD quality) sampling rates. Lesser rates are acceptable for CD-ROM and Web production, but even in these cases, it is preferable to capture audio at high quality and then down-sample (reduce the amount of information) using specialized software.

Upgrading and Service

Buyers should also be aware of the upgrade options available with particular capture cards. This is especially important for those who envision their needs expanding over the next two to three years. There are two basic ways to upgrade. The first is to simply trash your existing card and buy a newer card with the features you need. This may sound a little flip, but with entry level cards selling at well under $1,000 you may be able to consider it an investment in your education. A couple years down the road you will know a lot more about NLE systems and will be able to get the card that best meets your needs—plus benefit from the most current technology at the point when you are ready to expand to a more professional setup. In some cases this will be the only upgrade strategy, since few low-priced capture cards offer significant upgrade possibilities.

With higher priced cards there are upgrade options that allow you to protect your investment. The most basic upgrade is usually the addition of professional audio and video connections (balanced audio and component video). Other options include higher data rates, more sophisticated editing software, real-time transitions, and specialized 3D effects cards and software. The prices of these options vary greatly among manufacturers, but generally raise the price of the capture card well above $5,000. If you suspect that you will need some of these options in the future, make sure to investigate the expected cost and the reputation of the manufacturer for offering upgrade packages.

The reputation of the manufacturer or reseller is also important when it comes to service. Many resellers offer fifteen or thirty day return policies, which give you a chance to work out the bugs before committing to the purchase. The quality of phone support is important and fairly easy to check. You can verify the wait time by placing a call yourself, and since many problems occur during the initial setup of a system, you are likely to get plenty of real-world phone support experience while still within the return period.
High Speed High Capacity Storage

Even after compression, video files are very large. High-end systems can deliver upwards of 12MB/second of video information, while mid- to low-range systems deliver between 3MB and 9MB/second. At these rates you still need a lot of drive space to store your movie files and a fast drive to keep up with the flow of information. For example, a typical mid-range system might put out 4.5MB/second of video when capturing at 30 frames/second, full-size, best quality. At this rate, 2 gigabytes of drive space will hold only about 7.5 minutes of video. At the other end of the spectrum, a data rate of 600K/second allows the same 2 gigabytes of storage space to hold almost an hour of material, albeit at low quality. In general, the higher the quality and the longer the project, the larger and more expensive your drive system will have to be. Those who do not mind lower quality images, smaller windows, or reduced frame rates may be able to get by with less complicated and less expensive storage options.

Hard drives for digital video systems should be A/V rated SCSI drives. Many hard drives have incorporated the technology that used to set the A/V drives apart, but it is still worth confirming that a drive is A/V rated. SCSI (pronounced “scuzzy”) stands for Small Computer Systems Interface. This is the standard for connecting high performance hard drives to personal computers. There are currently three basic types of SCSI connections: fast, wide, and ultra-wide. Ultra-wide is the fastest of the three.

SCSI Cards

Macintosh computers have built-in fast SCSI connections; PCs come standard with IDE/EIDE drive connections, which are not suitable for NLE drives. PCs must be upgraded to SCSI, and Macs may be upgraded to faster SCSI. In selecting a SCSI card for your computer, be sure that your card and your drive have the same SCSI type. For best performance, an ultra-wide SCSI card should be paired with an ultra-wide SCSI drive. Mismatched drives and cards can be used, but they will perform at the speed of the slowest component.

SCSI cards can coordinate the flow of information between the drives and the capture card in order to get the maximum performance out of your system. If your maximum rate is 2-3MB/second, this coordination may not be crucial, and a single high speed drive on a fast SCSI connection will suit your needs. For higher rates and better image quality, you will want to invest in a wide or ultra-wide accelerator card.

RAID Systems

The other way to increase the performance of your drives is to get them to work together. This is called a RAID system (Redundant Array of Inexpensive Disks/Redundant Array of Independent Disks) or sometimes simply an “array.” The type of RAID setup used in NLE is called a “striped array.” With this configuration, the drives take turns writing portions of your video file. While one drive is receiving instructions, waiting for data, or looking for the right spot to put a piece of your file, the other drive is busy actually writing. This means much higher throughput to your drives. Wide arrays are currently the most common, though they are quickly being replaced by ultra-wide systems. Such arrays are capable of the 9MB/second data rates produced by higher-end capture cards.

An array is a must for those requiring excellent quality output to tape. Those doing simple multimedia work or off-line editing can get by without a wide SCSI card and without using an array. The good news is that PCI SCSI cards (the kind used in many PCs and in newer Macs) are relatively inexpensive ($200-$300) and the wide and ultra-wide drives that go with them are only five to ten percent more than the “narrow” versions. Preconfigured RAID systems, combining SCSI card, drives, and controller software, are more expensive, but like other turnkey systems, they often give better performance, better service, and less overall hassle.

The bottom line on high speed storage is that you need enough to hold the amount of video at the desired quality for getting your work done. Those wanting to use 30 minutes of video at 5MB/second quality will need at least a pair of 4GB wide drives and an accelerator card ($2,500-$5,000). Those needing 20 minutes of medium quality video (3MB/second) could probably get by with a single 4GB drive and no card, as could those doing off-line work with several hours of highly compressed, low resolution source material. Those doing simple multimedia production may even be able to get by with a gigabyte of free space on a decent A/V drive by capturing at very low data rates and small sizes.

4 A/V rating assures that the drive's self-adjustment operation compensating for changes in temperature (T-Cal recalibration) does not happen right in the middle of your video capture. Recalibration used to cause dropped frames while the drive paused to do its housekeeping;
Removable Storage

One other option to consider is removable media for storing your video files. Unlike traditional hard drives which you install once and then leave connected to your system, removable storage drives allow you to switch storage media in and out of your system. Media may be popped out to take to another location, or to be replaced with data for a different project. Two types of removable systems allow you to work this way and, as always, there are trade-offs.

At $100 per 1GB disk, plus $500 for the drive, Jaz cartridges are inexpensive compared to traditional hard drives. Unfortunately they are relatively slow. Even when configured in arrays, Jaz drives are only suitable for 2-3MB/second data rates. As single drives, their data rates are much slower and are often unpredictable. This makes them a possible solution for low resolution off-line work or for backing up small projects, but not for most other NLE uses.

Removable hard drives are just what the name implies. Instead of attaching a single hard drive to your computer, you attach a case or bay with several slots into which you can insert a hard drive. When you want to change drives, you shut down the system, pull out the old drive, pop in the new one and you are off again. You can even do this with the system running if you get what is called a “hot-swappable” system. These systems are a great convenience: the drives function at true hard drive speeds, can be configured as arrays, and offer an excellent way to have multiple projects available on the same system. The downside is price. The drive modules cost anywhere from slightly more to twice that of comparably sized conventional drives. Docking bay prices start at about $700 and go up rapidly with the number of docking ports available. Removable hard drives are available from numerous resellers, including Seagate, Quantum, IBM, and Micropolis.

The CPU

The computer itself (the CPU) has been conspicuously absent from our discussion so far. The CPU is a difficult place to start the purchasing process, because the most important consideration in selecting one is the capture card manufacturer’s recommendation. Card makers usually have a list of compatible CPUs which they have tested. In addition to the manufacturer’s recommendation, there are a couple of other considerations.

The first relates to anticipated changes in your needs. Expandability is one benefit of a more expensive CPU. Most often this comes in the form of space—empty slots in which to install additional cards and memory, or empty internal drive bays in which to install additional hard drives. Recently, the processor chip itself has become upgradable in some models. This allows you to buy a newer, faster processor at some point in the future, or, in some cases, to add multiple processors. This is a useful feature, especially if you do a lot of effects rendering where speed is at a premium. For video capture and playback, however, the CPU speed will have little impact in most systems.

The second factor in selecting a CPU is service. While not specific to video, good service will be important if your system goes down. You will need easy access to those who can help you solve the problem—either over the phone or from a service center. This is equally important for all the components you purchase.

Editing Software

The final major component in an NLE system is the software that you will use to edit. If your hardware choices primarily affect the quality and quantity of the video images with which you work, then the software affects how you conceive and create that work. For those who know what they want when they sit down to edit, the software matters less. Most editing packages have the same basic set of capabilities, and experienced editors can get the functions they need out of any of them. Just how easy it is, or how well that process meshes with the editor’s own thought process, is a different matter.

Some packages are better suited to the compression and recompression processes necessary to multimedia production; others emulate traditional tape editing systems. Still others will be familiar to those who have done optical printing of film. In selecting software, take into consideration the kind of experience you have with video editing, and the backgrounds of those who will be using your system. These choices are most relevant at the high end of the editing spectrum. With the Media 100 systems, for example, the upgraded path includes both better hardware capabilities and a switch from third party editing software to Media 100’s proprietary software. If you are prepared to spend a lot on software (about $5000), you can also buy professional packages by Avid and D-Vision, among others, which will run on your hardware.
These, however, are esoteric solutions as far as most buyers are concerned. Most capture cards come bundled with some software package, of which Adobe’s Premiere is the most common for both Macintosh and PC. Premiere is the de facto standard for inexpensive NLE interfaces and offers excellent integration with Adobe’s other industry leading production tools. As an editing tool it is imperfect, doing a lot of things well but few of them exceptionally. Nonetheless, it is a robust tool that has set the standard for all other applications. Ulead’s Media Studio software is another commonly bundled item on the PC side. Some capture cards, such as those from Radius, come with proprietary editing software (Radius Edit) that is designed to work especially well with the hardware, rather than relying on the more general software design of a multi-platform giant like Premiere.

**Additional Hardware**

There are a few additional pieces you will need in order to get your system up and running. The most obvious are the standard computer components like a monitor and additional RAM. Video editing software usually demands a significant amount of RAM. You will also want as large a monitor as you can afford. Editing timelines can get long and complicated, and many software packages, including Premiere, have multiple windows that can quickly clutter your screen space.

A regular video monitor is also helpful to preview your material before you digitize or to judge your final product before putting it back out to tape. Seeing video on the computer screen is always different than seeing it on a video monitor. If your final work will be seen on a video screen, then you will need to see it that way from time to time as you edit. The monitor does not have to be anything fancy—a simple TV with video inputs will do in most cases. For professional production, you may want to invest in a studio monitor. These can cost from $500 to $1500 depending on size and features. On the other end of the spectrum, those doing only simple multimedia work may be able to get by without any video monitor at all.

A video source of some kind is also a necessity. This can be as simple as a camcorder or home VCR, or as complex and expensive as professional video deck. If you routinely use professional video formats to acquire your video material then the investment in a professional deck may be worthwhile. Professional decks begin around $3,000 and go to well above $20,000, depending on format and features. For reference purposes, a simple Betacam player/recorder deck is around $7,000. Professional decks are also timecode capable, which allows you to redigitize your material automatically with the appropriate software (see Off-line Work above). Those just starting out often digitize directly from a camera. This works well in many cases, but can become cumbersome if you need to digitize a lot of material.

**What's on the Horizon**

**DV format**

Currently entering the scene are a number of new technologies that will impact future non-linear editing systems. Chief among them is the increasingly popular digital video (DV) format. The big question is how soon these innovations will be available at affordable prices. Cameras are now being sold which record video to tape in a compressed digital format instead of as an analog signal. The image quality of these cameras is excellent, rivaling that produced by professional Betacam equipment—at a significantly lower price. Consumer level DV cameras now begin at around $1,000 but go up rapidly as you add professional controls and features.

With DV format there is no need to convert from analog to digital. With the proper configuration, it is possible to transfer video information directly from your DV videotape to your hard drive. The configuration that allows this direct transfer is the IEEE 1394, or Fire Wire standard. Fire Wire is also a general protocol for getting computer devices to talk to one another, like the SCSI protocols discussed above.

Another promising aspect of FireWire is the possibility of transferring DV information from tape to hard drive at 4x speeds. This means that getting four minutes of video material from tape to hard drive would take only one minute. Unfortunately it’s not quite as simple as it sounds, at least not without spending a lot of money.

You need several components to take full advantage of these features of DV video. The first is a specialized FireWire card capable of bringing the DV information into the computer. These cards are just starting to become available from companies that produce existing capture systems. Initial pricing is in the same range as an inexpensive capture card.

Like the files captured from traditional analog
video, DV information is compressed. The compression takes place in the camera before the signal is recorded to tape. So while the FireWire card will allow you to get your files to disk, you still need something to decompress them when you want to play them back on your computer screen. Right now this decompression is only available in the form of a "software codec" which relies on software and your CPU to decompress the file. In other words, there is no specialized hardware yet available for your desktop computer which is dedicated to decompressing DV files. Until this specialized hardware is released (probably this summer or fall) you will not be able to get smooth playback on your computer monitor. These forthcoming DV hardware codec boards will cost about as much as the higher-end boards in our survey, around $4,000.

Your DV files can be recompressed into a codec supported by your existing capture card and played back smoothly to your computer or video monitor, but this is a very lengthy process, and obviates many of the advantages of editing with DV files. You can, however, still see your DV movie without recompressing it by sending the information back out the FireWire connection to your DV camera and viewing the analog output from there.

In order to take advantage of the promised 4x transfer speeds you will need a professional level DV deck capable of playing your tape at the 4x speed. These decks are also just becoming available and are likely to cost more than $10,000. You will also need drives capable of keeping up with the extremely high data rates associated with 4x transfer. The basic DV signal provides about 3.5MB/second of information. At 4x speeds that will quadruple to around 14MB/second. These are very high data rates and drive systems capable of these speeds are expensive.

DV files are fairly drive-space intensive. Three and a half minutes of DV video uses about a gigabyte of drive space. Enough space to hold thirty minutes of video at that quality currently costs between $2,500 and $4,000. These additional costs and the limitations to playback make DV editing less appealing at the entry level. This will undoubtedly change as the technology evolves, but that evolution is probably another year or two away. For the time being, analog to digital capture cards are still the most viable solution for low- to mid-range NLE systems.

**Editing Software**

On the software side, the dominance of Adobe's Premiere may soon be challenged by Macromedia's new editing product, Final Cut. Final Cut will incorporate all the essential features of Premiere and several new ones that will speed the editing process—particularly the ability to play back your cut without building previews, better audio control, and some compositing features now found only in After Effects. You can be sure that Macromedia will leverage the dominance of its Director product together with Final Cut to establish a suite of products to rival those of Adobe. Final Cut is apparently still in development, but was scheduled to be shown at this spring's National Association of Broadcasters (NAB) convention. Macromedia's Web site lists the first half of 1998 as the delivery date for Final Cut.

---

**National Association of Broadcasters '97**

As this issue of Connect was going to press, the annual National Association of Broadcasters (NAB) conference took place in Las Vegas. We plan to provide a more detailed report on the ACF Arts Technology Group Web site at [http://www.nyu.edu/acf/atg](http://www.nyu.edu/acf/atg). Here we will just touch on some significant trends.

While the DV standard and its related professional flavors from Sony (DVCam) and Panasonic (DVCPro) continue development as a low cost "prosumer" acquisition format, both companies also introduced more advanced digital formats. Sony introduced a 4:2:2 MPEG-2 format Digital Betacam variant called BetaCam SX, while Panasonic introduced a 4:2:2 non-MPEG-2 DV variant called DVCPro 50. As has often been said, the nice thing about standards is that there are so many to choose from.

A number of desktop DVD (Digital Video Disk) production systems are now available priced in the low six figures. Sometimes thought of as the next generation CD-ROM, DVD is also viewed as the digital replacement for the consumer VHS video cassette. Prices are expected to drop as they did for CD production hardware, and for the time being educators creating DVDs are advised to hire external services.

Finally, and perhaps most importantly, whether its called HDTV or just DTV, the next generation of broadcast television will enter the New York City marketplace over the next twelve months. The transition from the traditional 4x3 television screen to the higher definition 16x9 'cinema-like' screen will take place over the next ten years. This trend, along with the impending death of 16mm motion picture film stock, will someday associate the 4:3 aspect ratio with the same kind of nostalgic feel that black and white footage suggests today.

— Philip Galanter

*Associate Director, Arts Technology Group, ACF*
System Performance Expectations: What Can You Get for Your Money?

This page details some possible system configurations at three distinct price points. These are not intended as recommendations for any specific hardware or system. They are meant to give you an idea of the kind of functionality you can expect for different amounts of money. Putting these figures together with an outline of your projected needs will give you a good first estimate of what you will have to spend to get your system. The prices for these components are intended to reflect street pricing as of April 1, but are only estimates. These prices tend to change rapidly, so make sure to check before making any decisions.

### High Quality Video System
**Approx. $15,000**

<table>
<thead>
<tr>
<th>Component</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Media100kgx or</td>
<td>4,000</td>
</tr>
<tr>
<td>Radius Video Vision Studio PCI or</td>
<td>3,300</td>
</tr>
<tr>
<td>Truevision Targa 2000 or</td>
<td>3,300</td>
</tr>
<tr>
<td>DPS Perception</td>
<td>2,400</td>
</tr>
<tr>
<td>(DPS: add audio board, +700-1,300)</td>
<td></td>
</tr>
<tr>
<td><strong>Capture Card Range</strong></td>
<td>3,100-4,000</td>
</tr>
<tr>
<td>2x 9G ultra-wide SCSI array w/accelerator</td>
<td>4,200-7,500</td>
</tr>
<tr>
<td>Mac 9600/200MP or</td>
<td>4,300-4,700</td>
</tr>
<tr>
<td>Micron Millenia Pro2 360</td>
<td>4,300-4,700</td>
</tr>
<tr>
<td>Premiere and After Effects (included w/Media 100 card)</td>
<td></td>
</tr>
<tr>
<td>Radius Edit (included w/Radius card)</td>
<td></td>
</tr>
<tr>
<td>Premiere or SpeedRazor (Targa &amp; DPS cards: +700-1,300)</td>
<td></td>
</tr>
<tr>
<td><strong>Software Range</strong></td>
<td>0-1,300</td>
</tr>
<tr>
<td>17&quot; monitor</td>
<td>600-900</td>
</tr>
<tr>
<td>S-VHS/VHS Deck</td>
<td>700</td>
</tr>
<tr>
<td>additional RAM</td>
<td>200</td>
</tr>
<tr>
<td><strong>TOTAL:</strong></td>
<td>$13,800-19,000</td>
</tr>
</tbody>
</table>

For redigitizing from timecode add:
- pro video deck (Beta, DV, Hi-8) +6,500
- (or player only deck) +4,500
- (or pro S-VHS deck) +3,000

### Multimedia and Low Resolution Video System
**Approx. $6,000**

<table>
<thead>
<tr>
<th>Component</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>miroVIDEO DC30 or</td>
<td>1,000</td>
</tr>
<tr>
<td>Targa Bravado</td>
<td>700</td>
</tr>
<tr>
<td><strong>Capture Card Range</strong></td>
<td>700-1,000</td>
</tr>
<tr>
<td>4G fast SCSI A/V drive</td>
<td>1,200</td>
</tr>
<tr>
<td>Mac 7300/180 or</td>
<td></td>
</tr>
<tr>
<td>Micron Millenia 180 w/ fast SCSI card</td>
<td>2,300-2,500</td>
</tr>
<tr>
<td>Adobe Premiere or Ulead Media Studio (included w/cards)</td>
<td></td>
</tr>
<tr>
<td>17&quot; monitor</td>
<td>600-900</td>
</tr>
<tr>
<td>VHS Deck</td>
<td>300</td>
</tr>
<tr>
<td>additional RAM</td>
<td>200</td>
</tr>
<tr>
<td><strong>TOTAL:</strong></td>
<td>$5,300-6,100</td>
</tr>
</tbody>
</table>

For off-line editing w/ timecode add:
- pro video deck (Beta, DV, Hi-8) +6,500
- (or player only deck) +4,500
- (or pro S-VHS deck) +3,000

### Basic Video System
**Approx. $10,000**

<table>
<thead>
<tr>
<th>Component</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>miroVIDEO DC30 or</td>
<td>1,000</td>
</tr>
<tr>
<td>Targa 1000</td>
<td>2,100</td>
</tr>
<tr>
<td><strong>Capture Card Range</strong></td>
<td>1,000-2,100</td>
</tr>
<tr>
<td>2x 4G Ultra-wide SCSI array w/accelerator</td>
<td>3,000-4,500</td>
</tr>
<tr>
<td>Mac 8600/200 or</td>
<td></td>
</tr>
<tr>
<td>Micron Millenia P200 Plus</td>
<td>2,700-3,200</td>
</tr>
<tr>
<td>Adobe Premiere (included w/cards)</td>
<td></td>
</tr>
<tr>
<td>17&quot; monitor</td>
<td>600-900</td>
</tr>
<tr>
<td>S-VHS/VHS Deck</td>
<td>700</td>
</tr>
<tr>
<td>additional RAM</td>
<td>300</td>
</tr>
<tr>
<td><strong>TOTAL:</strong></td>
<td>$8,300-11,700</td>
</tr>
</tbody>
</table>

For redigitizing from timecode add:
- pro video deck (Beta, DV, Hi-8) +6,500
- (or player only deck) +4,500
- (or pro S-VHS deck) +3,000

**Comments:**
- VHS/S-VHS quality video
- Data rates up to 5MB/second (approximately 2 hours of storage at 1MB/sec or 25 minutes of storage at 5MB/sec)
- Good quality video for output to tape, excellent capabilities for professional level multimedia production

---

**Comments:**
- High quality video, suitable for finishing, probably for broadcast (with professional source/record deck and depending on network specifications)
Computing in the Humanities: An Overview of Tools and Techniques

Lorna Hughes
lorna.hughes@nyu.edu

The use of computers for scholarly research is a relatively new but fast expanding section of academic computing. Many humanists think of the computer as a word processor or a simple information storage device, but it can also provide more sophisticated access to large amounts of information, and it can act as a powerful tool to enhance teaching and research. As computer technologies become increasingly sophisticated, more and more humanities scholars have begun to devise innovative ways to use the computer as a tool for manipulating, storing, and analyzing information.

The earliest humanities projects in which computers made possible the impossible were literary and textual studies, where the introduction of computer technology in the late 1940s made it possible to prepare exhaustive concordances. Most humanities computing specialists agree that the most notable of these pioneering concordancing projects is Father Roberto Busa’s more than 60 volume complete concordance and word index of the eight million words in the works of St. Thomas Aquinas. The *Index Thomisticus* was begun in 1949 and was not completed until 1980. It would not have been possible at all without the assistance of a computer, but still took a lifetime of manual data entry using punch cards. By the mid-1960s many similar computer-assisted concordances were being prepared, with the goal of a multi-volume printed edition.

Today’s technology has advanced to the point where computers can generate concordances more or less automatically, allowing humanities researchers to carry out more complex projects in computer assisted textual analysis, by retrieving text interactively rather than exactly as it is stored. Thus, even as early as the mid-1960s, computers could be used to supply a quantitative basis for stylistic and authorship analysis, such as Andrew Morton’s work on the style of St. Paul’s Epistles based on sentence length and word distribution, or Frederick Mosteller and David Wallace’s study of the authorship of disputed texts in the Federalist Papers. Other possibilities for textual analysis now include sophisticated vocabulary studies of collocations (co-occurrences of words), the collation of manuscript variants for critical editions, and the metrical analysis of verse.

The preparation of machine-readable texts for concordancing has led to the creation of major electronic text archives such as the *Thesaurus Linguae Graecae*, the *Oxford Text Archive*, and the *Trésor de la Langue Française*, and to the development of flexible hypertext systems, perhaps the most well known application of humanities computing, through multimedia databases such as the classical Greek hypertext archive, *Perseus*, or via the World-Wide Web.

Technological advances have led to computing projects in other disciplines as well. In the 1970s, for example, historians began to use the computer to apply statistical techniques to historical analysis. The economic historian Robert Fogel was one pioneer in this area, notably in the history of slavery he co-authored with Stanley Engerman, *Time on the Cross*. It gradually became accepted practice for historians to design databases to store information which could be broken down into a quantitative format. The complexity of historical data, however, soon presented more problems than could easily be solved within...
the framework of simple data-processing software. To produce more methodologically relevant databases, systems were devised which could account for the vast differences among historical source materials and establish links between important names in a variety of different documents and document types, such as ephemera, baptism records, and images. The most flexible of these, KLEIO, still remains a remarkable example of this kind of system. Other database systems have further extended the reach of humanities computing to the fields of art history and archaeology, notably ICONCLASS and ORACLE, which have assisted the cataloguing and analysis of numerous collections of works and artifacts. Cutting edge computing in these fields is now focusing on the digitization of images, computer-aided reconstruction of sites, and the possibility of automated pattern recognition.

Future Developments in Humanities Computing
International humanities computing associations, such as the Association for Computers and the Humanities (ACH) and the Association for Literary and Linguistic Computing (ALLC), have collaborated with more specialized organizations and with humanities computing specialists from a number of institutions to address what may be the single most important problem posed by the development of humanities computing over the past forty years—the need for a common encoding (or “mark-up”) format for scholarly machine-readable texts. In the past, scholars have developed encoding schemes more or less at their own discretion in accordance with their particular scholarly goals. The traditions of centuries of print culture incline us to take phenomena such as non-standard characters, footnotes, marginalia, markers for logical divisions (e.g., chapter, verse, stanza) and illustrations at face value. But the means for encoding such features electronically are often incompatible; they remain restricted to the features of a given text or corpus, and risk being completely dependent on one kind of hardware or software. In response to these issues, members of the various humanities computing associations developed the Text Encoding Initiative (TEI). The TEI guidelines specify a common interchange format for all types of machine-readable texts (such as prose, poetry, corpora, manuscripts, etc.), provide a set of recommendations for the encoding and representation of all possible features in the preparation of new textual materials, and document major extant encoding schemes, developing a metalanguage which allows the encoding schemes themselves to be encoded and described in a machine-readable form.

Future developments in humanities computing will certainly incorporate the TEI initiative in the development of new applications—in particular the development of new text analysis software. Also significant is the role the Internet has to play in the dissemination and storage of data. In this respect, the Next Generation Internet and Internet 2 initiatives will be a key focus for those involved in humanities computing research and development. What is clear is that in the humanities, the computer is now an acknowledged part of the collection of resources available to scholars.

WYLBUR To Be Phased Out in May 1997
The Academic Computing Facility is ending its support of academic use of WYLBUR on May 31, 1997. Licenses for statistical software will not be renewed for the 1997-98 academic year. No new academic research accounts will be created for WYLBUR IBM/MVS. Most researchers will be moving to PC-compatible computers, using SPSS for Windows or SAS for Windows.

Faculty and graduate student researchers should download files they wish to keep by the end of April. All files must be downloaded in a text format independent of the mainframe. Files downloaded in non-text format will not be usable on other computers.

SPSS and SAS system files should be converted to "portable" files. Creating these portable files in SPSS or SAS involves writing "export" programs in those languages specifically for the IBM mainframe. Additionally, files in WYLBUR libraries must be reformatted to individual data sets. FTP is available on WYLBUR to move the converted files to your PC.

ACF's Statistics Group is available to assist you in migrating off WYLBUR (contact Frank Lopresti at 998-3398). Plan out the move early, as the statistical package licenses will not be renewed by ACF after this academic year.

— Frank LoPresti
Finding Your Way on NYU Web: New Search Tools for Visitors and Providers

Randy Wright and Kristina Abeson
randy.wright@nyu.edu, kristina.abeson@nyu.edu

NYU Web serves out thousands of Web pages regarding schools, departments, classes, club activities, and other valuable information for prospective and current students, and for faculty and staff. The main NYU page at http://www.nyu.edu serves as the starting point for NYU Web, and although a user can navigate anywhere from the main page, the sheer size of NYU Web makes some items difficult or time-consuming to locate.

We have taken a significant step in alleviating this problem by instituting a search engine on NYU Web. The search engine is a modified version of Netscape Communication Corporation’s Catalog Server. It currently involves programs running on 3 different computers. Every day the search engine re-indexes every word in over 30,000 different NYU Web documents.

Querying the Search Engine
A full NYU Web search page is available for use at http://www.nyu.edu/search.shtml. The NYU search engine can be driven by entering a single word into the search form. For example, one might search for documents containing the word art by entering the term “art.” In addition, a higher level of searching precision is available to NYU users. Searches can be augmented by using and, or, and not in the search terms. You can look for documents that contain both the word art and the word dutch by entering the terms “art and dutch.” It is also possible to find documents that contain one or both of these words by entering “art or dutch” as the search term. You can find documents that contain the word art but do not contain the word dutch by entering “art not dutch.”

In fact, the query language available on the NYU search engine is very rich. It was supplied by Verity Corporation and is completely documented at http://www.verity.com/. The NYU search engine will also offer help pages to support users at http://www.nyu.edu/search/help.shtml. In addition to the Verity language, the Netscape Catalog Server includes technology from the Harvest Project, RSA Data Security Inc., and Berkeley DBM (See Multiple Technologies in the NYU Search Engine).

How It Works
Three computers work together to index University servers within the “nyu.edu” domain. There is one catalog server and two resource description servers. The resource description servers house programs called robots. Each night at midnight, the robots begin sending out probes to NYU Web servers. The robots retrieve hyperlinked documents from the Web sites and use the links in each document to obtain additional documents. The documents are indexed as they are retrieved. This process continues until every document on NYU Web that can be found is retrieved and examined. The information obtained about each document is fed into the catalog server running on www.nyu.edu. This server imports and combines the information from the robots, responds to users’ search requests, and outputs results based on the most recent index.

There are currently more than 100 Web servers running at the University. The robots do not probe...
A sample NYU Web search results page.

Multiple Technologies in the NYU Search Engine

The NYU search engine uses Netscape Catalog Server, which has roots in the Harvest project. Harvest was part of a project funded by the Advanced Research Projects Agency, U.S. Air Force Office of Scientific Research, National Science Foundation, Hughes Aircraft, and Sun Microsystems, to provide an integrated set of tools to gather, extract, organize, search, and replicate information across the Internet.

In particular, the Netscape Catalog Server adopts Summary Object Interchange Format (SOIF). Created by Harvest, the SOIF standard allows servers to communicate with each other in a non-proprietary way. The Catalog Server can exchange information with any SOIF-compliant server.

The Harvest software itself was among the applications evaluated by ACF. Netscape’s Catalog Server was selected because it more reliably found all URLs in the NYU target area.

The NYU search engine system also includes security functions provided by RSA Data Security Inc.; interfaces in Sun Microsystems’s Java language; and database functionality from the Berkeley DBM package from the Regents of the University of California.

Multiple Technologies in the NYU Search Engine

The NYU search engine uses Netscape Catalog Server, which has roots in the Harvest project. Harvest was part of a project funded by the Advanced Research Projects Agency, U.S. Air Force Office of Scientific Research, National Science Foundation, Hughes Aircraft, and Sun Microsystems, to provide an integrated set of tools to gather, extract, organize, search, and replicate information across the Internet.

In particular, the Netscape Catalog Server adopts Summary Object Interchange Format (SOIF). Created by Harvest, the SOIF standard allows servers to communicate with each other in a non-proprietary way. The Catalog Server can exchange information with any SOIF-compliant server.

The Harvest software itself was among the applications evaluated by ACF. Netscape’s Catalog Server was selected because it more reliably found all URLs in the NYU target area.

The NYU search engine system also includes security functions provided by RSA Data Security Inc.; interfaces in Sun Microsystems’s Java language; and database functionality from the Berkeley DBM package from the Regents of the University of California.

The search engine indexes only those pages that are linked, however distantly, to NYU Web pages. You can keep a page from being indexed by making it a stand-alone page—one that is not linked to or from anywhere else on NYU Web. Alternatively, you can use a “robots.txt” file to prevent pages or sections from being indexed. For more information on controlling how the robots index your pages, see http://www.nyu.edu/search/robots.txt.html.

Meta tags provide another method for controlling the indexing of your pages. The NYU search engine recognizes Meta tags, allowing you to specify information about a Web page. For example, documents have an “expiry” time associated with them in the search engine index. A document’s index entry is removed when its expiry time has passed. By default, each document is assigned an expiry time of one day. In the header of an HTML document, the author can insert a specific time for the document to expire, such as: <META NAME="RD-Expires" CONTENT="Thu, 11 September 1997 12:00:00 GMT">. This date overrides the default expiry time. The page will be maintained in the index until September 11, 1997. After that date, when the page is retrieved by a robot the default expiry time of one day will again be applied.

New Tool for Webmasters

In addition to the full campus search, NYU Web authors can take advantage of the NYU search engine by including a query form in their HTML code to perform limited scope searches. These searches find documents on servers or sections maintained by the organization offering the search.

For example, if the GSAS Webmaster wants a search engine which will only return results from pages within the GSAS Web site (that is, in the sites beyond a limited list of NYU servers. The search engine catalogues only New York University Web space and does not aim to index personal Web pages which may be stored on university servers. The sites indexed include these 15 servers:

**The R1 robot searches**
- www.nyu.edu
- www.stern.nyu.edu
- www.bookc.nyu.edu
- www.med.nyu.edu
- www.math.nyu.edu
- cims.nyu.edu
- www.sce.nyu.edu
- www.cns.nyu.edu
- cvisions.nyu.edu
- www.mrl.nyu.edu
- www.cat.nyu.edu
- edgar.stern.nyu.edu
- www.cs.nyu.edu
- www.tsoa.nyu.edu
- www.itp.tsoa.nyu.edu

**The R2 robot searches**
- www.cs.nyu.edu
- cvisions.nyu.edu
- www.mrl.nyu.edu
- www.cat.nyu.edu
- edgar.stern.nyu.edu
- www.cs.nyu.edu
- www.tsoa.nyu.edu
- www.itp.tsoa.nyu.edu

The search engine indexes only those pages that are linked, however distantly, to NYU Web pages. You can keep a page from being indexed by making it a stand-alone page—one that is not linked to or from anywhere else on NYU Web. Alternatively, you can use a “robots.txt” file to prevent pages or sections from being indexed. For more information on controlling how the robots index your pages, see http://www.nyu.edu/search/robots.txt.html.

Meta tags provide another method for controlling the indexing of your pages. The NYU search engine recognizes Meta tags, allowing you to specify information about a Web page. For example, documents have an “expiry” time associated with them in the search engine index. A document’s index entry is removed when its expiry time has passed. By default, each document is assigned an expiry time of one day. In the header of an HTML document, the author can insert a specific time for the document to expire, such as: <META NAME="RD-Expires" CONTENT="Thu, 11 September 1997 12:00:00 GMT">. This date overrides the default expiry time. The page will be maintained in the index until September 11, 1997. After that date, when the page is retrieved by a robot the default expiry time of one day will again be applied.

New Tool for Webmasters

In addition to the full campus search, NYU Web authors can take advantage of the NYU search engine by including a query form in their HTML code to perform limited scope searches. These searches find documents on servers or sections maintained by the organization offering the search.

For example, if the GSAS Webmaster wants a search engine which will only return results from pages within the GSAS Web site (that is, in the
A Web Search Glossary

boolean query
In the case of the NYU Search Engine, a search phrase consisting of two or more terms separated by the words and, or, or not.

Catalog Server
The Netscape Communications Corporation search engine offering.

hyperlink
A reference contained in one document which fetches another document.

indexing
Keyword indexing is offered in the NYU search engine. Every word in each document is examined and put into a list in such a way that the document containing the keyword can be located and retrieved.

keyword
A word that is sought within a document.

query language
A word or group of words that can be used as operators and modifiers to compose searches.

resource description
A summary of the contents of a document in a special format suitable for creating indexes of documents.

robot
A special program that obtains HTML documents from Web sites and follows the hyperlinks in those documents to obtain more documents.

search engine
As used in this article, a facility that allows a user to find information.

search form
An electronic form that allows a user to enter a search term and query.

search page
An HTML document containing a search form.

stand-alone page
An HTML page that is not pointed to by hyperlinks on any other page. Thus, the page is sort of an “island.” Such a page will not be retrieved by a search engine that follows hyperlinks.

“www.nyu.edu/gsas” hierarchy), she can change the search engine scripts to require the “gsas” subdirectory name in each result. The GSAS-only search engine would only give results that both include the search term and are in the “gsas” subdirectory.

Limited scope search pages can be given any look and feel by customizing the title and graphics on the results page returned to suit the site. Documentation and sample search pages are available online at http://www.nyu.edu/search/howto.html.

The search engine is operated by the Distributed Computing and Information Services Group at NYU’s Academic Computing Facility. For more information, contact webmaster@nyu.edu.

Using Eudora to Change Your NYU-Internet Account Password

If you use the Eudora client e-mail program to read your electronic mail from one of the NYU-Internet systems (is through is6) or from the ACF’s general purpose UNIX system acf5, you can now change your account password from within the Eudora program itself.

ACF account passwords expire every six months as a protective security measure. Up until now, to learn that your password was about to expire, you would have had to log in interactively with telnet. The only way to change your password has been to log in similarly. Yet many Eudora users have never gotten in the habit of interactive logins, and have had no easy way to change their passwords.

Within Eudora, the Special menu contains a Change Password option. To change your ACF account password, select this option. You will be prompted for your current password, and then (twice) for a new password. Once you have changed your password, you must use the new password to retrieve e-mail messages from the system.

Note: Passwords on ACF systems must be at least six alphanumeric characters in length. Password on these systems are case-sensitive. When choosing a new password, do not use a word that can be found in the dictionary, and do not use any of the last five passwords you have had on the account.

If you have questions about this new capability, contact the ACF HelpCenter by sending e-mail to acf.is.questions@nyu.edu or by calling 998-3333.

—G. Chapman
Computer Advocacy @ NYU held its second annual Computer Awareness Week from March 24 to March 27, 1997. The events spanned a wide range of topics related to computing and networking. All of the events included both the NYU community and the general public.

An introductory talk, “Understanding Hardware,” explored how Macintosh and IBM PC-compatible computers work. At a demonstration entitled "Talk to Me,” the president and founder of Empower Practice Consulting, Mark Hollis, discussed two voice recognition systems: PowerSecretary and DragonDictate. These systems work in tandem with other software, such as word processors, spread sheets, and databases, as well as e-mail and other Internet applications.


Dave Garaffa, author of the BrowserWatch Web site (http://www.browserwatch.com), discussed the development and relative merits of various World-Wide Web browsers, from the first cross-platform line mode browser developed at CERN, through the emergence of the graphical browser Mosaic, up to the present browser competition. He suggested that as we demand more from the browsers we run on our computers—and more from the authors in shorter intervals of time—security concerns are likely to become problematic.

Ilya Slavin and Tim O’Connor of NYU’s Academic Computing Facility addressed one security issue in a talk on e-mail encryption and the PGP program (Pretty Good Privacy). This public key encryption program is one of the strongest means available for private citizens to secure their electronic communications, yet relatively few members of the NYU community have adopted it.

“Social Transformation in an Internetworked Society” brought New Jersey Institute of Technology Professor Murray Turoff and NYU Professor of Sociology Caroline Persell together to discuss the impact of network technology on every day life. Prof. Persell raised questions regarding the real penetration of network technology in a world where significant portions of the population have never used a telephone, let alone a computer. Prof. Turoff noted how the options for developing regions have changed with advances in technology, to the point where it is possible to “leapfrog” several steps in the movement toward a technologically enabled society.

Two sessions addressed health and safety considerations for computer users. In the session “Diagnosis and Treatment of Cumulative Trauma Disorders (CTDs),” Dr. George Piligian, Co-Director of the Program to Prevent and Treat Cumulative Trauma Disorders and Clinical Instructor at the Mount Sinai Irving J. Selikoff Center for Occupational and Environmental Medicine, demonstrated how injuries occur, and how postural and work practice components can exacerbate them.

An unusually wide range of disciplines was represented on the panel “RSI: Complementary and Alternative Treatments,” including physical therapy, biofeedback, nutrition, acupuncture, medical massage, Hatha Yoga, Alexander Technique, ergonomics, speech pathology, and psychology. The panelists addressed the “biopsychosocial” individual in treatment development, stressing the importance of tailoring comprehensive treatment plans rather than using “off the shelf” solutions.

With the help of Computer Advocacy’s cosponsors Academic Computing Facility, the Department of Sociology, and the Henry and Lucy Moses Center for Students with Disabilities, Computer Awareness Week ’97 was a great success, and we plan to continue this annual event.
VRML in Scientific Visualization

Adel Hanna
adel.hanna@nyu.edu

Advances in computing technology allow computations on a scale heretofore unknown. Terabyte capacity storage and gigaflop computers provide researchers with the ability to solve complicated problems which generate very large data sets. Among the areas which are producing very large amounts of data are remote sensing (satellite and imaging), medical science (radiology), molecular modeling (chemistry and biology), computational fluid dynamics, statistics, cartography, archaeological reconstruction, and oceanography. The challenge arises in finding the correlations and relationships between different parts of the data sets.

As Richard Wesley Hamming points out in his Numerical Methods for Scientists and Engineers, “The Purpose of Computing Is Insight, Not Number.” Reams of numerical output cannot give one insight into the data. By displaying the data sets in understandable form as a three-dimensional model, one may gain an intuitive grasp of the conclusions. A three-dimensional model can be turned, spun, peered into, backed away from, and examined from all angles. It matches how we see things in the real world, allowing us to apply familiar skills of perception.

Software capable of rendering three-dimensional models of data has been available for decades. However, sharing such a model with other researchers has required that your colleagues work on the same high-speed hardware with the same proprietary software that you do. Transferring the data itself was often a better job for Federal Express than for the Internet. These obstacles to distributing your results (processor power, proprietary software, and bandwidth demands) may be overcome by Virtual Reality Modeling Language (VRML), a standard language for describing interactive 3D objects across the Internet.

VRML is a three-dimensional description language for the World-Wide Web, much like the text formatting language HTML (Hypertext Markup Language). HTML text takes up very little bandwidth in transmission. Once the HTML is received by a browser (say, Netscape Navigator or Internet Explorer), the browser uses rules described in the HTML specification to format and display the information.

A VRML representation of the P450 enzyme, from “3D Molecular Graphics on the World Wide Web,” by H. Vollhardt and J. Brickman
mation. Similarly, VRML text describes 3D objects and their relationship in space using VRML scene description language. When a VRML browser such as Cosmo Player or the Live3D plug-in receives the VRML file, it generates a three-dimensional exploratory model using the rules described in the VRML specification.

VRML is not the only way to deliver three-dimensional effects to a broad audience, but it is one of the best. You can create three-dimensional effects in HTML by embedding a video clip on a Web page. The video clip mimics an exploration of a three-dimensional object, but this has some limitations. Video files are very large and consume a great deal of bandwidth. Video is stored in a literal pixel-by-pixel, frame-by-frame way. VRML stores the same image information as a set of points and vectors from which an image can be rendered, reducing the file size by orders of magnitude. By using VRML, the bottleneck for viewing three-dimensional models moves off the network and on to the local computer. Instead of being limited by how fast data can be transferred before it is viewed, one is only limited by how fast the receiving computer can render the information received.

When distributing a three-dimensional object by using a video clip, the person viewing the object is limited to seeing only what was specifically recorded. If the top of the object was not included in the video, that aspect of the object will remain forever a mystery. The resolution of a video clip can not be changed once the video is made. If the clip only shows surfaces close-up, the viewer will never be able to back away and see the big picture. In VRML, the user has complete control over the angle and the magnification of the object, making it a truly interactive experience. There are an infinite number of possible views, every one of which is available to the viewer when the object is delivered as VRML.

VRML, like HTML, is an open standard. VRML browsers are available for UNIX, Mac, Windows95 and Windows NT based computers. The same VRML code can be used on any of these platforms without modification. This is a great improvement over the use of proprietary software, which requires any two people sharing data to commit to a single platform, and automatically excludes any one not using the same hardware/software combination. Also, as a single open standard, any VRML viewer can be used to examine three-dimensional objects, regardless of what kind of computer was used to create the original code. There is no need to learn a new user interface for every file type encountered.

Now scientists and others are beginning to recognize that VRML greatly facilitates their efforts to gain support for research and to help them get needed funding. Publishing their work in 3D on the Internet in a way that clearly conveys the meaning of the data not only helps explain the researcher's idea but also casts a glow of both credibility and capability.
Classes and Talks

Academic Computing Facility — New York University

The full contents of ACF Classes and Talks is on NYU Web at http://www.nyu.edu/acf/classes/

Summer '97 Schedule

<table>
<thead>
<tr>
<th>Listing by Date</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wednesday, May 7</strong></td>
<td>NYU-NET Software (PC) C-3</td>
<td></td>
</tr>
<tr>
<td><strong>Wednesday, May 21</strong></td>
<td>NYU-NET Software (Mac) C-3</td>
<td></td>
</tr>
<tr>
<td><strong>Thursday, May 22</strong></td>
<td>Intro to SAS C-3</td>
<td></td>
</tr>
<tr>
<td><strong>Saturday, May 24</strong></td>
<td>Using a PC at an ACF Lab C-2</td>
<td></td>
</tr>
<tr>
<td><strong>Sunday, May 24</strong></td>
<td>Using a Mac at an ACF Lab C-2</td>
<td></td>
</tr>
<tr>
<td><strong>Tuesday, May 27</strong></td>
<td>Using a Mac at an ACF Lab C-2</td>
<td></td>
</tr>
<tr>
<td><strong>Wednesday, May 28</strong></td>
<td>Internet and E-mail C-3</td>
<td></td>
</tr>
<tr>
<td><strong>Thursday, May 29</strong></td>
<td>Using a PC at an ACF Lab C-2</td>
<td></td>
</tr>
<tr>
<td><strong>Friday, May 30</strong></td>
<td>Choosing Your Computer C-2</td>
<td></td>
</tr>
<tr>
<td><strong>Saturday, May 31</strong></td>
<td>Using a Mac at an ACF Lab C-2</td>
<td></td>
</tr>
<tr>
<td><strong>Monday, June 2</strong></td>
<td>Using a PC at an ACF Lab C-2</td>
<td></td>
</tr>
<tr>
<td><strong>Tuesday, June 3</strong></td>
<td>Using a Mac at an ACF Lab C-2</td>
<td></td>
</tr>
<tr>
<td><strong>Wednesday, June 4</strong></td>
<td>Intro to SPSS C-3</td>
<td></td>
</tr>
<tr>
<td><strong>Thursday, June 5</strong></td>
<td>NYU-NET Software (PC) C-3</td>
<td></td>
</tr>
<tr>
<td><strong>Wednesday, June 11</strong></td>
<td>Internet and E-mail C-3</td>
<td></td>
</tr>
</tbody>
</table>

**Friday, June 13**
Understanding Your Mac C-2

**Wednesday, June 18**
NYU-NET Software (Mac) C-3

**Friday, June 20**
Understanding Your PC C-2

**Wednesday, July 2**
NYU-NET Software (PC) C-3

**Monday, July 7**
Using a Mac at an ACF Lab C-2

**Tuesday, July 8**
Using a PC at an ACF Lab C-2

**Wednesday, July 9**
Internet and E-mail C-3

**Saturday, July 10**
Using a Mac at an ACF Lab C-2

**Wednesday, July 16**
NYU-NET Software (Mac) C-3

**Friday, July 18**
Understanding Your Mac C-2

**Friday, July 25**
Understanding Your PC C-2

**Wednesday, July 30**
Internet and E-mail C-3

**Wednesday, August 6**
NYU-NET Software (PC) C-3

**Wednesday, August 20**
NYU-NET Software (Mac) C-3

About the ACF Classes and Talks

All members of the NYU community are welcome at the ACF's classes, workshops, and talks. There is no charge for any of the ACF instructional sessions, but participants should have a current, valid NYU ID card. In some cases, as noted just after a course description, an appropriate computer account is required.

Seating capacity: To avoid overcrowding, we have listed maximum seating capacities for each class. We recommend that you arrive a few minutes early in order to secure a spot.

Classes by arrangement: Faculty members may sometimes arrange special classes for a specific course or research group. These do not necessarily have to be given at an ACF site. For classes in statistics, call Frank LoPresti (998-3398); for other applications, call the ACF Innovation Center (998-3044).

Additional information: A list of pertinent ACF locations and phone numbers, and pointers to further information about ACF services and resources, can be found on page C-4.

—Vincent Doogan
Associate Director, ACF
vincent.doogan@nyu.edu
Choosing Your Computer (Mac and PC)
This talk is intended to help you select the best personal computer for your needs. It will cover the basic components of a computer, as well as the other hardware required for various tasks. We will also discuss how you can assess your particular needs to establish your criteria for selecting computer tools. NYU Computer Store staff.

Warren Weaver Hall, room 313
Seating capacity: 30; first come, first served; talk.
- Friday 12:00–1:30
  - May 30

Understanding Your Computer (Mac and PC)
This introductory talk will help you learn about your computing equipment. It will focus on such basic operations as setting up your computer, setting up a printer, and configuring your operating system with the fonts and tools you need. Discussion will include troubleshooting techniques and other strategies for dealing with problems you might encounter while using your PC. NYU Computer Store staff.

Warren Weaver Hall, room 313
Seating capacity: 30; first come, first served; talk.
- Friday 12:00–1:30
  - May 30

2. For PC Owners
Fridays 12:00–1:30
June 20
July 25

Using a Mac at an ACF Lab (Mac)
A hands-on introduction to the Macintosh computer. Topics include the ergonomics of proper computer use, working with the graphical user interface, understanding the file system, choosing printers, file servers, and other devices, and launching software applications. ACF staff.

Education Building, 2nd floor
Seating capacity: 25; first come, first served; hands-on class.
- Tuesdays 11:00–12:00
  - May 27
  - June 3
  - July 8
- Saturdays 11:00–12:00
  - May 24, 31
  - July 12

3rd Ave. North Res. Hall, level C-3
Seating capacity: 15; first come, first served; hands-on class.
- Mondays 1:00–2:00
  - June 2
  - July 7
- Wednesdays 11:00–12:00
  - May 28
  - June 4
  - July 9

Using Unix at ACF (Unix)
An introductory class on using the Unix operating system, variants of which run on several different types of computers at the ACF. Most are accessed at ACF labs through PCs, Macs, and terminals, but the SGI workstations also use Unix. The basics will be covered: logging onto the host machines, organizing files, editing text, printing files, and using applications. ACF staff.

ACF Unix account required.

Tisch Hall, room LC8
Seating capacity: 24; first come, first served; hands-on class.
- Thursdays 11:00–12:00
  - May 29
  - June 5
  - July 10
- Saturdays 11:00–12:00
  - May 24, 31
  - July 12

3rd Ave. North Res. Hall, level C-3
Seating capacity: 15; first come, first served; hands-on class.
- Mondays 1:00–2:00
  - June 2
  - July 7
- Wednesdays 11:00–12:00
  - May 28
  - June 4
  - July 9

Using a PC at an ACF Lab (PC)
A hands-on introduction to the PC — the "IBM-type" personal computer. Topics include the ergonomics of proper computer use, working with the user menus on the PCs in the labs, understanding the file system, choosing printers and file servers, and launching software applications. ACF staff.

Tisch Hall, room LC8
Seating capacity: 24; first come, first served; hands-on class.
- Wednesdays 11:00–12:00
  - June 4
  - July 9
E-Mail and Network Services

www.nyu.edu/acf/accounts.office.html
www.nyu.edu/acf/help/

Introduction to the Internet and Your ACF E-Mail Account
(NYU-Internet Account, Unix)

This talk-demo introduces new holders of the NYU-Internet Account to its menu interface and components. E-mail concepts and commands will be explained and demonstrated. The account runs on ACF's DEC mini-computers and is connected to NYU-NET and the Internet. Lisa Barnett and Vincent Doogan.

Warren Weaver Hall, room 313
Seating capacity: 30; first come, first served; talk-demonstration.

Wednesday 12:00–1:30
May 21
June 18
July 16
August 20

NYU-NET Software
(Mac, Windows)

This talk is intended for those who have TCP/IP connections to NYU-NET from their office or home. The TCP/IP and PPP protocols will be discussed, and software based on these protocols will be demonstrated. The software to be discussed includes Netscape, Eudora, and Fetch. Lisa Barnett and Jane DelFavero.

1. For PC Users
Warren Weaver Hall, room 313
Seating capacity: 30; first come, first served; talk.

       Wednesdays 12:00–1:30
May 7
June 4
July 2
August 6

2. For Mac Users
Warren Weaver Hall, room 313
Seating capacity: 30; first come, first served; talk.

       Wednesdays 12:00–1:30
May 21
June 18
July 16
August 20

Statistics, Databases, and Spreadsheets

www.nyu.edu/acf/socrcl

SAS
(Windows, Unix)

This series will progress from the basic description and operation of this statistical package to advanced concepts and usage. Robert Yaffee.

Warren Weaver Hall, room 313
Seating capacity: 30; first come, first served; talk.

Thursdays 4:15–6:00

1. Introduction to SAS
May 22

2. Intermediate SAS
May 29

SPSS: SPSS for Windows
(Windows, Unix)

SPSS (Statistical Package for the Social Sciences) is a comprehensive, integrated system for statistical data analysis. These presentations will use either the Windows or the newer Unix version of SPSS, but the programming concepts are applicable to all versions of SPSS. Frank LoPresti.

Warren Weaver Hall, room 313
Seating capacity: 30; first come, first served; talk.

Using the ACF Computer Labs
http://www.nyu.edu/acflabs

NYU faculty, staff, and students in degree programs may use the PCs and Macintoshes in the ACF computer labs without charge during designated hours or, with ACF access accounts, during all hours of operation. In most labs, only those with access accounts are admitted during the high-demand times from noon to 8:00 pm on weekdays.

Faculty members who wish to use the labs during restricted hours, or who expect their students to use the labs for either coursework or independent projects, should be sure to obtain coursework accounts from the ACF Accounts Office (Warren Weaver Hall, room 305). This type of account assures access to the labs during all hours of operation.

For complete lab hours, pick up a brochure at any lab or see:
http://www.nyu.edu/acf/nyu-events

ACF Classes and Talks    Spring 1997    C-3
Important Dates for ACF Users

May 18  ACF Summer Hours begin
May 19  Summer Session I begins
May 24  Spring Class accounts expire
May 26  Memorial Day*
May 29  Applications for fall semester Class Accounts due
June 9-20  Applications accepted for account extensions from students expecting Incompletes in Summer Session I (Instructor's signature required)
June 10  Responsibility Statements due for Summer Session I
June 14-27  Summer Session I Class Account holders should store all files needed past June 28
June 27-28  Summer Session I ends; Class Accounts issued for Summer Session I expire
June 30  Summer Session II begins
July 4  Independence Day*
July 22  Responsibility Statements due for Summer Session II
July 22-Aug. 1  Applications accepted for account extensions from students expecting Incompletes in Summer Session II (Instructor's signature required)
July 29-Aug. 8  Summer Session II Class Accounts holders should store files needed past August 9
Aug. 8  Summer Session II ends
Aug. 15  Class Accounts issued for Summer Session II expire
Aug. 22  Fall Class Account Applications for late registration due
Aug. 30-Sept. 1  Labor Day weekend†
Sept. 1  Labor Day*
Sept. 3  Fall semester begins; ACF Fall Hours begin
Sept. 30  Responsibility Statements due for Fall Semester

* NYU holiday: Labs and offices closed.
† Holiday hours in effect. For details please see www.nyu.edu/acf/nyu-events/

Additional Information

Class Locations
http://www.nyu.edu/acf/classes/
The following are the street addresses of the locations referred to in the course descriptions.
Warren Weaver Hall, 251 Mercer Street
Education Building, 35 West 4th St., 2nd floor
Third Avenue North Residence Hall, 75 Third Ave., C-3
Tisch Hall, 40 W. 4th Street, lower concourse

HelpCenter
http://www.nyu.edu/acf/help/
251 Mercer St., 2nd floor  998-3333
Troubleshooting; software distribution; information about ACF services and academic support.

Accounts Office
http://www.nyu.edu/acf/accounts/
251 Mercer St., 3rd floor  998-3035
Faculty and staff account applications and information: individual, coursework (class), and NYU-Internet accounts. DIAL account applications for faculty, staff, and students.

Innovation Center
http://www.nyu.edu/aflic/
251 Mercer St., 2nd floor  998-3044
Discipline-oriented resources and services for faculty and advanced students; instructional computing support; new and emerging technologies.

Student Computer Labs
http://www.nyu.edu/acf/labs/
Education Building  998-3421
3rd Ave. North Res. Hall  998-3500
Tisch Hall  998-3409
Student NYU-Internet Account applications; computer and Internet access (see ACF flyers and above Web address for hours and rules of access).

Publications
http://www.nyu.edu/acf/pubs/
Pamphlets, flyers, brochures, and the magazine Connect for users of NYU computer and network services. Printed copies are available at the HelpCenter and labs; online editions are at the above Web address.

News and Announcements
http://www.nyu.edu/acf/nyu-events/
Updates on hours and services; special events and other notices of interest.
T


twenty years ago, a nascent networking effort by academia and the federal government began to link associated governmental and university research centers so that data files and messages could be sent and shared among a small number of researchers and administrators based in the United States. By the early 1980s many of these networking projects existed, and were united by the adoption of a standardized set of software protocols named TCP/IP. The internetworking of these smaller networks had been accomplished, and it was christened the Internet.

One would have to be a modern day Rip Van Winkle to have missed the cacophony of enthusiasts and pundits who have explored the information and communication services of the Internet during the 1990s. The exponential growth of the Internet was a consequence of the wide-spread adoption of its text-based services, such as electronic mail (e-mail), discussion groups (newsgroups and bulletin boards), and database services such as library catalogs. Another surge of interest arose with the introduction of its newest component—the World-Wide Web—in 1993.

The World-Wide Web is the collection of computers (servers) on the Internet able to store and retrieve linked multimedia-based documents, which are then viewed on a desktop computer with such graphical browser programs as Netscape and Internet Explorer. Since the first graphical browser was introduced just four years ago, there are now an estimated thirty-one million Web-based multimedia documents (Web pages) available to the point and click of your mouse.

The presence of audio and video on Web pages has brought both delight and frustration to many. The lack of maturity in the infrastructure of the Internet presents these powerful and exciting media without the viewing and the listening quality to which we are accustomed. The deficiencies in the delivery of network services to the desktop are very apparent to academic users whose applications now compete with all other Internet traffic.

Since Fall 1996, several initiatives, including the Next Generation Internet and the Internet 2 project, have been announced. These initiatives promise to provide advanced networking capabilities for new Internet activities and services.

The Next Generation Internet (NGI), announced by President Clinton and Vice President Gore, is a three-year $300 million investment that will create the foundation for the networks of the twenty-first century. Specifically, the NGI initiative will connect universities, national laboratories, and research institutions with high-speed networks that are a hundred to a thousand times faster than today's Internet. NGI will also invest in research and development that will enhance the capabilities of the Internet for real-time services, and demonstrate new applications designed to support important national goals including scientific research, distance education, environmental monitoring, and health care.

During this same period, a group of universities also joined with partners from government and business to jump-start the future of the Internet. Dubbing the project "Internet 2" (I2), one hundred charter members met in San Francisco in January 1997 to lay

Vincent Doogan is the Associate Director for User Services at ACF.
the foundation for a prototype network to support a new generation of academic and research applications. At the heart of I2 is GigaPOP, a new technology to provide advanced communications services. This technology is a point of intercommunication between institutional members and service providers. Merging old and new technologies, the I2 prototype may well become the basis for commercial solutions. While the I2 project is a collaborative effort to support primarily academic objectives, clearly, the NGI and I2 projects are related substantively and politically.

New York University has a strong initial position in both initiatives. As a charter member of the Internet 2 project, NYU was represented at the San Francisco meeting by George Sadowsky, Director of Academic Computing Facility, and by Gary Chapman and Vincent Doogan, Associate Directors. NYU is also a full partner in the regional NGI project, NYSERNet 2000. The New York State Educational and Research Network (NYSERNet) sponsors this project.

Significant funding from government and industry will be available to I2 participants and members of NYSERNet 2000. The University's next step is to identify specific faculty members, activities, and programs that require a new level of information and communications technology to help invent the future. Recently, George Sadowsky asked NYU deans for their "help in casting a very wide net for potential Internet 2 applications. We are interested in both faculty already using the current technology to its fullest capacity, as well as those from non-traditional areas with needs which require network media capacity beyond what is currently available."

Examples of such activities include:
- simultaneous digital video connections to a number of universities for remote collaboration or "virtual" academic conferences, with support for interactive white boards and multimedia;
- multimedia projects with unprecedented content depth, interactivity and production values;
- immersive digital environments with high speed connections allowing individuals at different locations to share a single virtual reality;
- distance education technology to market NYU programs, or to import experts from outside of New York to supplement NYU programs;
- complex scientific computation and visualization harnessing resources at multiple locations;
- further development of services of digital libraries; and
- performances which link real and virtual artists and audiences at multiple theaters.

Faculty and other interested parties should contact George Sadowsky or the following ACF managers: Vincent Doogan (Instructional Technology), Edward Friedman (Sciences), Philip Galanter (Arts), Lorna Hughes (Humanities), or Frank Lopresti (Social Sciences) for further discussion and information. Also, see our Web page at http://www.nyu.edu/acf/usg/internet2.

Since I2 and NGI initiatives are concurrent with ongoing development projects in the Internet community both in the United States and world-wide, it is apparent that there is a growing desire and will to improve the way we do things. So be prepared, Internet 2 is coming soon to your desktop computer!
Dynamic Growth in Remote Access to NYU-NET

Carlo Cernivani
carlo.cernivani@nyu.edu

Some twelve years ago, NYU-NET's modem services consisted of two dozen 2400 baud modems interfacing with three Ungermann-Bass Network Interface Units (NIUs). That modem service provided only telnet access via the campus broadband network to various systems connected to NYU-NET. At that time call volume was nominal, Kermit was the software of choice, and users were anxious to get their hands on the fast new 2400 baud modems just hitting the market.

Today Academic Computing Facility's NYU-NET modem services support 720 modems distributed between two different modem pools offering a wide range of networking protocols (telnet, rlogin, tn3270, PPP, SLIP, AppleTalk Remote Access). Our efforts to meet the ever growing demands for fast, reliable remote access to NYU-NET have resulted in one of the better services offered by any university.

The 995-4343 service provides generic terminal access for users wishing to use the telnet, rlogin or tn3270 protocols to access NYU-NET hosts and resources. In this service, there are currently 240 modems. These connect to NYU's telephone system via POTS (Plain Old Telephone Service) lines and interface with NYU-NET via four Xylogics Annex III terminal servers. The modems themselves are rack mounted MultiTech units. Approximately eighty percent of the service is populated by V.34 modems (28,800 bps) while the rest is comprised of V.32bis (14,400 bps) units. This service is the direct descendent of the original twenty-four 2400 baud modems.

On a typical day this service now handles over 12,000 calls/connections.

The most versatile and evolving aspect of our remote access services comes by way of the DIAL (Direct Internet Access Link) modem pool. This service, which is currently supporting 480 digital modems, is the cornerstone of our remote access capabilities. The service is composed of 10 fully loaded Cisco AS5200 Remote Access Servers. These servers are at the leading edge of integrated modem/server technology. Each AS5200 is composed of a modular chassis containing three card slots and a high-speed backplane. All three card slots are used. One slot holds a Dual T1/Primary Rate Interface (PRI) card with integrated channel service units (CSUs). The other two slots hold modem carrier cards, each of which is loaded with two 12-port modem cards.

All these pieces work together every time you call the DIAL service and complete a connection. First, the call is carried to us across one of 20 Enterprise Flexpath T1 circuits from NYNEX. Each T1 contains 24 individual DS0s (phone lines). Two T1 circuits connect to each of the server's T1 interface cards. The interface card takes the incoming call and makes a connection via a TDM (Time Division Multiplexing) bus to an available modem in the server. The modems in the AS5200 are Microcom MNPl 0 V.34+ units (33.6 kbps). Once connected, your call interfaces with the AS5200 server itself, on a separate data bus, so as to initiate a connection to NYU-NET. At this point the modems have made their connection and control is handed over to the communications software on your end and the command level interpreter on the server end.

This integrated hardware platform provides us manageability from both the telephone and data network side of the service. We can control access from the outside world to our modems by changing how any or all of the 720 modems appear to the public phone network. Modems can be set to appear "busy" in preparation for resetting or switching in a different modem card. Meanwhile, we can also control, monitor and secure the network interface between the DIAL servers and NYU-NET. The servers currently run Cisco's Release 11.1 IOS (Internetwork Operating System) software which gives us a tremendous range of configuration options. Both AS5200 IOS software and modem firmware are upgraded as new revisions of either become available. This allows us to take advantage of all bug fixes and feature enhancements.

We have engineered the service into several separate subnets to maximize data throughput. This is

---

Carlo Cernivani is operations associate at ACF. He has primary responsibility for the NYU-NET modem pool.
accomplished by distributing traffic loads, thereby minimizing excessive data collisions on any one of the individual network segments used by the servers. The DIAL service is currently handling between 6,500-7,000 connections daily, with new users being issued accounts to the service every day.

To keep it all straight on our end, we took advantage of an organization framework with which you may be familiar: “The Simpsons” television show. Our DIAL servers are named Homer, Marge, Bart, Lisa, Maggie, Itchy, Scratchy, Apu, Krusty, and Milhouse. The main system providing both authentication and domain name service is Smithers. Infrastructure devices (such as ‘hubs’) play a role here too. One such hub is named “Springfield” while another one is called “Shelbyville” after Springfield’s neighboring rival town. There is a Mr. Burns lurking around somewhere and will make “his” presence felt as an SNMP based monitoring/management workstation (SNMP—Simple Network Management Protocol).

The modem services are an invaluable presence on NYU-NET. Thousands of users connect to our services daily to access the Internet, World-Wide Web, e-mail, newsgroups, Bobcat, and more.

As an indication of the degree of growth over the last several years, back in February of 1993 our modem services handled 59,512 connections. This past February we handled 479,615 connections—an eight-fold increase in four years.

**What’s on the horizon?**

We expect the DIAL service to continue expanding to meet growing demands and hope to add servers in a timely fashion so as to insure availability and reliability for our users at all times.

Many users have asked if we intend to provide 56K modems in our DIAL service. It is inevitable that we will support 56K protocol modems at some point. However, there are presently two divergent “standards” in the market place. One implementation of the 56K protocol is the USRobotics X2 DSP (Digital Signal Processing) chipset. The other is a joint venture between Lucent Technologies and Rockwell Semiconductor Systems which has been dubbed the K56.flex protocol. The USRobotics and Lucent/Rockwell DSPs are not compatible. Modems based on one of these chipsets can not interoperate with modems based on the other at the advertised 56K speed. The majority of modems used in the world will be based on the Lucent/Rockwell design, as will our Microcoms, but until modem manufacturers move closer to setting an international standard, we are reluctant to commit resources to this technology.

We expect to be working along with Cisco on the next generation of access servers. These servers will have greater modem capacity with faster processors and network interface speeds. They will be designed to accept high capacity data channels from the public telephone network. This is yet another evolutionary step that will eventually lead to a more advanced telephony network infrastructure such as a SONET (Synchronous Optical NETwork).

On our side of the phone line there is a world of new protocols, software releases, and hardware platforms as rapidly changing technologies in the realms of telephony and data networks converge to form a powerful data transport infrastructure. To you, the user, there are different concerns—whether the service answers when you call, whether you have to contend with busy signals, and once you get through, whether your window into NYU-NET and the Internet is fast and reliable. That is our ultimate goal—providing the best remote access services possible to NYU-NET, and beyond.
Security Is My Business
It ain’t a pretty racket, but somebody’s gotta do it — Part II

Tim O’Chandler

[The story so far: In Part 1, a nervous little man in a blue suit showed up on a muggy night at the office of our private investigator. The visitor was the manager of a computer network that had been broken into by hackers. The interlopers wired most of the money out of the firm’s account, but they didn’t stop there—they also sent e-mail, signed with the nervous manager’s name, to the president of the United States. If you missed Part 1, you can catch up with the saga at http://www.nyu.edu/acf/pubs/connect.]

“There were certain words used,” he said.
I asked quietly, “Like—‘kill’?”
His eyes bugged out even more as he nodded.
“‘Shoot’?”
He nodded again.
“They used your personal account to do this?”
Another nod.
“Did they send these lovely messages to anyone else?”
“The vice- vi- vi- vice-president.”
“And?”
“Their families.”
I winced. Whoever was responsible had a real slash-and-burn attitude. I said, “Did they share these thoughts with anybody else?”
“Some members of Congress. And the Secret Service.”
I shook my head and laughed at the gall. “You know your goose is pretty well cooked, right?”
He nodded.
I said, “Making threats like that is a felony. You don’t have to actually do anything real. Just the threat is a serious crime.”
He looked as if he might have an idea. “Hey!” he said. “What about freedom of speech?”
I closed my eyes. I felt like I was reciting from a textbook. “It’s not protected when you make threats like that.”
“But I didn’t make them!”
“Like I said before, prove it.” I spat the last nutshells into the garbage. “As far as the federal cops know, and from the perspective of your guys in the expensive suits, and even your own system, you are the guilty party. They don’t want to know about forged mail. All they want is to find the person making the threats.”
His eyes looked like they were going to pop all the way out of the sockets.
“Oh, don’t worry,” I said. “They’ll make your life miserable for a while. They’ll probably audit you, come in with legal paperwork. They might even seize your systems.”
He relaxed a little.
“But I wouldn’t want to be around when your bosses figure out what’s going on. That money is gone to Switzerland or some offshore stash. Your people will be lucky if they’re still in business next week.” I wasn’t trying to be cruel, but he was so far gone it was almost a game. “Say,” I said genially, “what business are you guys in?”
“Money management and information brokering.”
“What the hell is that?”
“Like when you want to get some information we get it for you—for a price.”
“Ah,” I said. “So now that you have no financial assets, you’re probably going to lose your hardware and software any minute now, and the company’s credibility is about to end up in the toilet.”
“We’ve got a damage-control team.”
“My friend,” I said, “you can put the best liars in the world on the case, but nobody with half a brain cell will trust you with day-old bread after this is over.”
“But that’s why I came to you! I heard about you in the Omlor case. You—you tracked down the guy who turned every icon on every computer in that company into a black square. He demanded ransom to fix it, and I heard you caught him in an hour.”
I shrugged. “I can’t talk about my clients.”
“What about the feud between the movie agents, when one of them kept faking mail as the other?”
“It wasn’t the same. The stakes were different.”
That case had made the papers, so I didn’t mind acknowledging my involvement.
“Everybody I talk to mentions you. They say you know all the angles.”

As his alter ego Tim O’Connor, the author works at NYU’s Academic Computing Facility.
"I know a lot of angles," I told him. "But when you see a tornado heading your way, you don’t complain that it wasn’t mentioned in last night’s weather forecast."

He didn’t say anything.

"What kind of precautions did you take?" He didn’t answer. "Did you pick obscure passwords? Did you check your systems thoroughly before you put them in service? Did you monitor them for suspicious activity?"

He stared dumbly at me as if I were a talking frog.

"You know how to keep system logs, don’t you?"

"I don’t have time for all that."

"Did you safeguard your system," I said, "so a cracker would trip some kind of alarm if he broke in?"

The expression on his face screamed no at me.

I said, "Did you teach your users to pick good passwords? Did you probe your own systems at least a little, to see if they were weak? Did you pay attention to who was using your system and how they used it?" The guy didn’t answer. "Did you think people outside are so much dumber than you are?"

Apparently he did. I watched him seethe.

I said, "People like you try to cut corners and take the easy way out, and figure there will always be someone like me around to fix things up for you."

I’d been in this business a long time. I’d seen more cases than I could remember, and I had the details of every one of them stored in a vault, on disks and tapes I had encrypted with the strongest security software available, but I’d never seen a guy in as hopeless a situation as this little man in the blue suit.

I leaned forward across my desk, with my hands clenched on the desktop, in case he decided to jump me. In my entire professional career, I had waited for the chance to say what I was about to say, but even as I gathered my energy, I saw him shrinking into himself. I dropped my voice as low and husky as it could go, just like in the movies, and I said, menacingly, "You’re ... taking ... the ... fall." Then I let myself fall back into my chair.

He seemed to stagger, even though he was still only sitting there, and his eyes bugged out again, and his head tilted forward like a rag doll’s, and for a moment I wondered if he had outright died on me. But the veins in his neck were still throbbing, and when I quickly slipped a pocket mirror under his nostrils, I saw that he was breathing, even though he wasn’t reacting at all.

The breeze blew hard through the window. It riffled the pages of the book I had been reading. It flipped the necktie of the little man in the blue suit, but there was nothing relaxing about the breeze. It was only hot and sticky and hopeless, and I thought of all the advice I had given out about making computer systems secure. These people never wanted to listen, people like the little man and the fools who wrote his paychecks. To them, security was a waste of time. Most of them liked to believe they didn’t need the extra protection. It wasn’t until they had to clean up after a disaster that they saw my advice differently.

The kind of intrusion my visitor had experienced was a system manager’s worst nightmare. The cracker could do virtually anything to the systems, and could likely inflict severe damage to other computers on that network, and possibly to computers outside, on the Internet. Some of those computers outside belonged to very powerful and humorless people.

I liked to remind my clients about the movie called Invasion of the Body Snatchers, where each victim’s body is taken over by an alien. I went out of my way to point out that when it all happens online, within computers and networks, it is infinitely worse. In the movie, at least, you could see the blank expression on every victim’s face. But if the people running the systems don’t bother to take adequate safeguards in advance, they probably have no way of knowing if their systems or networks have been infiltrated.

Nearly all of them decide I am a little too paranoid. Most of them, if they heed my advice at all, assure me that my recommendations will be an integral part of Phase Two of the project, which most of them never get around to doing. They pay me and send me away, and that is where my responsibility ends.

I opened my desk drawer to get something strong for us to drink. My visitor looked like a man who needed a drink. I felt like I needed one myself. Then I remembered: I don’t drink. I don’t even keep a bottle in my desk. I pulled a laptop computer out instead. I thought it might be handy to take notes. Then I put it away again, knowing it would only make things more confusing. A bad system compromise is like ———

Some of those computers outside belonged to very powerful and humorless people. ———


that. It rattles everybody it touches, even if it’s not your own system that gets hit.

Then I realized, without thinking hard about it, that this situation was impossible, and there was nothing I could do for this little man. The best he could do was turn himself over to the authorities, explain what happened, and learn some new job skills, because only an idiot would keep him on the job, or give him another chance in the same line of work.

My office was hot. I felt the heavy sweat in my shirt, in the small of my back, and above my lip. All we could do was wait, then. If my little visitor had not been tailed on the way from his office to mine, it was only a matter of time before the people looking for him deduced where he was. I have a certain reputation of my own, and people who know their way around this town understand that if it’s a late-night security crisis and the key person is missing and the lights in my office windows are on, the chances are good that the missing person is at my desk, pleading for me to do the impossible.

The only difference here was that my little guy in the suit couldn’t ask anyone for anything. When I leaned over near him, I thought I heard him muttering, “I shoulda, I shoulda, I shoulda.” It made me wonder if he had ever heard me give my little speech on security tactics. They listen to it, a lot of them choose to ignore it, and most of them end up saying “I shoulda”—in one way or another.

I put my feet back up on the desk and hefted the book back onto my knee. Hell, I figured, I was getting paid to look it over. I might as well put some effort into it. I wiped my face with a handkerchief and checked the clock and picked up at the spot I had been reviewing when I was interrupted. I knew my office would see a little action very soon.

Off in the bowels of the building, I was sure I heard the sound of the inevitable elevator as it creaked up to my high floor. And then it would be over quickly, and I would make my way home to some kind of quiet.

---

**NETWORK BOOKBYTES**

**Deeper: My Two-Year Odyssey in Cyberspace**

John Seabrook  
New York, N.Y.: Simon & Schuster  
1997, 288 pp., $25

**Review by Tim O’Connor**  
tim.oconnor@nyu.edu

In the field of computing, people who know answers to complex questions often rely on analogy to explain difficult concepts. Hence, the Internet is forever immortalized as the Information Superhighway. Local Internet connections over phone lines are dubbed “on-ramps” to this highway. So, we develop a structural framework that allows us to discuss matters when some people in the discussion are “technical” and others know less about the underlying technology.

John Seabrook has hit upon another engaging analogy in his first book, *Deeper: My Two-Year Odyssey in Cyberspace*. Taking as his models those 19th century writers who concerned themselves with frontiers (external frontiers in the case of Francis Parkman, internal frontiers in the case of Henry David Thoreau), Seabrook approaches the terrain as the greenest of novices. Like Thoreau walking away from his experiment on the shores of Walden Pond, Seabrook sums up his experience with the assurance of one who has mastered the wilderness.

At a glance, *Deeper* sounds like a snoozer of an idea: What fool would want a ringside seat at a two-year narrative about a man sitting in front of his computer?

But Seabrook seizes control of the situation early on, with an engaging tale of how he fled from computers while at boarding school, for fear of being branded a “nerd.” In his easygoing voice, he manages to make the story compelling. After all, don’t most of us know how outcasts and losers are treated in high school?

So, the young Seabrook shuns the technology, because he would be put to public shame in his effort to understand the logic and technique of computing.

Cut ahead fifteen years or so, and here is Seabrook again, this time enraptured by a Macintosh Powerbook, a computer that has a personality Seabrook feels he can work with. He acquires a modem as part of an assignment to write a *New Yorker* profile of Microsoft chairman Bill Gates. He
conducts most of that interview by e-mail, and after the profile is published, he receives a few nasty messages from a notoriously unpleasant journalist who has co-authored a biography of Gates. (He discreetly avoids identifying the journalist.) Seabrook manages to capture the feeling of one who has been tormented by e-mail messages. Suddenly the online world has veered directly into the path of his real life, and he is not sure which world he occupies, as he tries to put the abusive message in some context that makes sense to him.

Under these circumstances, Seabrook begins to grapple with the basics of life online. It is oddly reassuring to see how he gradually moves away from his faceless CompuServe account and into the world of online communities that are more intimate, like those at the WELL and ECHO. He eventually settles with a small, local Internet service provider—a nice touch that suggests the easy familiarity of a country store, in contrast to the atmosphere of a supermarket.

Along the way, although the story is nominally about his online experiences, Seabrook weaves a sly, modern allegory. As he becomes more involved with the online world, we can see how it begins to alter his physical world. This is the one place where the writer’s voice strains to be naive, with mixed results, because we suspect that even as he reports his “gee-whiz” adventures, he knows exactly what his online explorations are doing to his relationships with his wife, his family, his friends. It is easy to identify, as Seabrook hilariously does, the mannerisms of the junkie who is becoming hopelessly hooked. At one point in the book, it would almost sound plausible if the narrator described—and rationalized—moving out of the apartment he shares with his wife, in order to set up a new household just for him and his Mac.

But sanity, home life, and the culture of The New Yorker (where Seabrook is a staff writer) prevail. The author emerges from his wilderness as confidently as any 19th century explorer might return to the landscape of society. Another man has mastered another untamed territory, and the show goes on.

Along the way, perhaps his most fascinating path is the one he travels as a writer. Through his account on the WELL, he joins an online forum that discusses The New Yorker. Without announcing his own presence, he observes other participants evaluating his work. Like a man mesmerized, Seabrook returns again and again to the WELL forum. Here in miniature is the story of every writer who puts words into type and then waits breathlessly to see what reaction occurs. Seabrook’s experience is like a writer’s dream come true: He can nearly see into the minds of his readers and learn exactly what they think of his writing, though this is not an entirely pleasant observation. Eventually spotted lurking, he is drawn into the discussion by the other participants in the forum, which may well be a writer’s nightmare come true. The participants in the discussion show no mercy with him, but after he acquits himself honorably, he earns the acceptance of the community, and knows he has cleared a significant hurdle.

Deeper is not, and does not pretend to be, a guide to life online, any more than Walden is a manual for backpackers. There are plenty of those how-to books around, many drearily similar in scope, sound, and design. It is refreshing to find a writer move beyond the state of “how to” into the state of “why to,” and to follow his improvisations from that perspective. Seabrook has created a travelogue of a most intriguing sort. Here, finally, we begin to see what happens when a writer recounts life online. Put a computer in front of a computer jock and ask for some prose, and more often than not you will get the kind of deathless material that is enshrined in so many technical manuals. But put a computer and a network connection in front of a writer whose interests lie beyond the limits of the computer monitor, and you begin to see how rich—and at the same time how absurdly limited—life at a computer can be.

Deeper is essential reading for anyone who wonders whether it is worth taking the plunge into the traffic jam on that fabled superhighway. Perhaps the answer can only come from one’s own experience, but Seabrook shows that there is more to life online than the mundane and the prosaic and the intangible. His essential lesson is that it is possible to be immersed online and in real life, and that one side of the divide can complement the other.

(Author’s Note: A couple of years ago, during an online discussion about computers and society, I wrote a small message in which I wondered aloud how ubiquitous computing might affect the powerful among us. Seabrook quoted a passage of that musings in Deeper. I otherwise have no connection to the author.)
Statistics and the Social Sciences

Exploring NUD•IST for Qualitative Analysis

Frank LoPresti
frank.lopresti@nyu.edu

Developing theories from a wide collection of interviews, notes, research, and inspiration is equal parts art and science. Managing such qualitative, non-numeric data can be made easier with the assistance of qualitative data analysis (QDA) software. If you have boxloads of note-covered index cards, or piles of articles with densely annotated margins; if your field notes are stored in word-processor documents organized in directories within directories, their inter-relationships slowly fading from your memory; if your reading material is stored on the shelf but your reading notes are kept on your computer, you should consider using some sort of QDA software. QDA software, such as QSR NUD•IST, indexes and annotates both on-line and off-line documents, helping you evolve and confirm theories based on widely varied and unstructured sources.

This article briefly introduces QDA methodologies and some of the software packages that have proved useful to social scientists and humanists in the past. It then offers a description and review of QSR NUD•IST (Non-numerical Unstructured Data Indexing Searching and Theorizing), a QDA software package that is gaining popularity among qualitative researchers.

Methodology in Qualitative Data Analysis
Qualitative studies involve the storage and organization of textual and graphical data, with the goal of developing and testing theories based on that data. As an example, a researcher may collect texts and interviews, developing notes on the individual parts of the research as it is accumulated. Portions of the text may then be coded with keywords which indicate relationships among the segments. By drawing related text segments together based on the keywords, theories can be developed about the underlying causes and effects, and conclusions can be drawn.

As the quantity of text increases, researchers may turn to computer programs which provide automated search and retrieval of text. Relatively simple content analyses such as word frequency counts and relative position measures may be used to characterize some of the data. Advanced analyses might call on graphics technology to display relationships among the data and to build theories and evaluate hypotheses.

Methodologies for non-numeric research are as diverse as the range of academic disciplines. For example, Huberman and Miles describe three linked subprocesses following data collection: reduction, display, and theory construction and verification.

Reduction—categorizing a body of text by keyword—is used by all researchers. In essence, you reduce data whenever you label a file folder or store similar documents in a directory. Display, the second subprocess, is illustrated by the use of keywords to focus attention on critical relationships and categories. With some types of data, such as interviews, case clustering methods aid in the display process. Theory construction in non-numeric research is supported by coding and retrieval methods. The ability to examine and review similarly coded cases can be critical to understanding causal relationships.
Software Applications in Qualitative Analysis

Qualitative researchers have found innovative ways to suit common software products to their purposes. Word processors, text search packages, relational databases, and hypertext applications such as HyperCard have all served a role in organizing textual and graphical data. From these roots special purpose programs for QDA have developed, including code-and-retrieve software. Since the functionality of the older types of software is probably familiar, comparisons with their capabilities will help clarify what a QDA package like NUD•IST can add to qualitative analysis.

Word processors Word processors have the ability to handle multiple documents on-screen, to perform pattern searches, to annotate text, and to include charts and images within documents. The linking and publish-and-subscribe features of some word processors enable the marking of content for inclusion within passages of a text, allowing marked text to be automatically updated in the document being developed. Linking features eliminate the need to cut-and-paste, or to hunt for a given section of text each time it needs updating.

While these are useful features, word processors do not perform other tasks needed by qualitative researchers. Word processors do not effectively manage lists of keywords attached to documents; nor do they manage notes which may be attached to keywords as investigations proceed. They do not automate the retrieval of sections of separate documents with related keywords or themes.

By the same token, most researchers will probably continue to rely on word processors for maintaining source documents even after adopting a QDA package. Word processors provide advanced editing features not available in the simple editors generally provided with QDA software.

Text search software Text search packages enable the study of occurrences of themes in large bodies of text. Some packages are used to study word frequency and position; others, such as Gofer and ZyINDEX, enable the use of complex Boolean searches. The UNIX command grep is considered more powerful than most text search packages. QDA software includes varying levels of search functionality.

Database and statistical software Relational databases and statistical packages like SPSS and SAS manage data in two-dimensional arrays. These arrays are conceptualized as tables in which each row is an observation and each column is a variable.

I REMEMBER WHEN ...

It was only four or five years ago that I remember needing to find an important quote in a 1000-page log I was analyzing. In the back of my mind I was sure it was in there somewhere, but where? The quote was by a mother in my educational home computing study and included a phrase something like “cream cheese.” She used this phrase to refer to her son’s complexion in her characterization of his intensive use of their home computer. My search for the quote was exactly like looking for a needle in a haystack. Such a manual task sent me reeling.

In this same study I often had the urge to search for and accumulate evidence about various attitudes parents had about their children’s uses of home computers for educational purposes. The job of manually going through the log as many times as would be required to satisfy my curiosity was daunting. In addition, I frequently had the urge to compare the boys in the study to the girls in a variety of ways. But again, the task of reading through and then physically cutting and pasting the accumulated evidence for each comparison was too formidable.

Looking for needles in haystacks, isolating key themes, and comparing subgroups are just three of the many qualitative goals that NUD•IST has made routinely achievable, and in a matter of minutes. NUD•IST cannot do your thinking for you, anymore than SPSS does one’s thinking in a statistical analysis. But like SPSS with quantitative data, NUD•IST makes inquisitive, active, and extensive qualitative analyses far more possible. It promotes a much greater willingness to routinely test out speculations that only a few years ago—as in my case—were avoided or may have come to mind but were abandoned. In short, NUD•IST enhances the completion of richer and more accurate qualitative research.

The doctoral course I teach on the analysis of qualitative data using NUD•IST is entitled “Software Tools for Qualitative Analysis” (E10.3111). It is offered during our three-week summer intersession, which runs from middle May through the first week in June.

— Joseph B. Giacquinta
Professor of Educational Sociology
School of Education
Consider a database in which each row is an interview. A researcher could code a case with dichotomous variables like "guilt" and "worry," giving them a value "yes" if those feelings are noted. Some databases handle text data and have the functionality to do text searches.

Sorting and statistical procedures like cluster analysis are also of interest in qualitative studies. In SPSS, an Autorecode function can be applied to open-ended string data to collect all string responses and create a new numeric variable, allowing an interviewer to code open-ended questions with keywords. The researcher does not need to list the entire range of answers before coding begins. For example, if an interviewee has been asked to report a favorite activity, the researcher enters the response as a word or phrase and Autorecode creates a new variable for each activity as it is entered, assigning "1" for the alphabetically first answer, "2" for the second, and so on.

These packages work well with structured, discrete data such as questionnaires. If relational database structuring is adhered to, tables from several studies may be joined at key fields and useful complex analysis can be done. Where these tools fail, though, is in the study of varied, unstructured content which does not lend itself to strict question-response categorization.

**HyperCard** A hypertext application, HyperCard organizes information in what can be conceptualized as stacks of cards. The researcher can link cards in any fashion. A stack might be linked at common words, so that clicking on a keyword would bring you to the next card containing that keyword. HyperCard is good for storing and searching text and for attaching memos. Features added to HyperCard software since its first appearance include the ability to generate reports of the results of keyword searches.

**Qualitative Data Analysis Software**

QDA software emerged from the functionalities that qualitative researchers found useful in the assortment of tools described above. Techniques for storing, processing, and retrieving knowledge have led to the development of several types of specialized QDA software which incorporate methods for developing and testing theories. Among these are code-and-retrieve programs such as Ethnograph, which replicate the manual code-and-retrieve process. Specified text segments are coded into a database, from which relevant text can later be retrieved based on the coding information. Studying these codes enables the researcher to relate concepts to text. Ethnograph also handles memos, text searches, sub-setting, and the generation of statistics about text searches.

**QSR NUD•IST**

NUD•IST is based on code-and-retrieve techniques, but incorporates features of index-based software. At its most elementary level, NUD•IST keeps on-line and off-line text organized and portable. Work is done within a "project" that you create and name. As on-line text is introduced to a project, NUD•IST copies the text to the project directory. For this review, I created a project called "wine." Every document I introduced about wine is stored by NUD•IST in the wine directory. I included text from documents taken from CDs, from my old computer files, and from the World-Wide Web. NUD•IST provides a facility for keeping track of off-line documents, so I was able to include pointers to parts of my library. If I copy the wine directory to another disk, all the documents come along automatically in data subdirectories, as does any other work done within NUD•IST on my wine project. This organizing feature was enough to sell me on using the package for my own work.

When documents are introduced to the project, the researcher decides on the text "units." A unit can be a word, a line, a paragraph, or an entire document. The unit that you select will affect how NUD•IST deals with your data—for example, the measurement of position and the study of distances between pairs of words within the text. NUD•IST also uses units in indexing and searching. By default, the program assumes that units are marked off by hard returns—so, as most documents are formatted, the software will take paragraphs to be the units. This is easily altered with a word processor when preparing the text. For example, to make lines the unit, you can save it as DOS Text with hard returns on every line. Inserting a hard return after every word will make words the units.

The france node and its child nodes displayed as a tree.
NUD•IST Tutorial: Help for the Fainthearted

NUD•IST has an on-line tutorial which offers a project in four stages, with detailed instructions for each stage. The data for the project includes short transcripts of interviews and focus group discussions on the issue of introducing a smoke-free workplace. The tutorial’s first stage handles introducing on-line and off-line documents; running document reports (which allow you to view your text with the headers and subheaders you have assigned, among other things); and introduces the concept of text units (e.g., the distance between one hard return and the next). The “documents” menu also affords the opportunity to alter the text and to write memos as you are inspired by your data.

Now that your data is on-line you can begin the really creative work: coding. Throw out all your three-by-five cards. NUD•IST does not need them to work its magic. The second stage of the tutorial shows you how to apply code or, as it is referenced in the tutorial, index your data. You can use your mouse to drag through portions of text and assign it to a category (categories are called nodes in NUD•IST).

Like any researcher worth their salt, you should give some thought to how you want to code your data. With NUD•IST, you can use your creativity as you move through the text. You no longer have to duplicate data which fit into more than one node; you just apply more than one index and NUD•IST will do the work for you. The real beauty of the indexing is that as you realize that you need to change your indexes and move nodes around, a simple cut-and-paste procedure does the job. In the tutorial, for example, tracts of the interview are indexed as IMAGES to indicate images of smoking. As other categories appear to emerge within the node IMAGES, it is possible to create sub-nodes (termed “children”). For example, some reported images of smoking are pleasant and others are unpleasant. These text tracts can be coded IMAGES/PLEASANT and IMAGES/UNPLEASANT. Likewise, entire interviews are indexed as either SMOKER or NONSMOKER to indicate the preference of the interviewee.

At any time you can run a report which will show you the data you have indexed at any node. Running a report on the node IMAGES/UNPLEASANT will show all the references which were made to smoking being unpleasant. As your nodes and child nodes emerge, you are also able to view a tree-like diagram which graphically displays the nodes, their names, and their interrelationships.

The third stage of the tutorial introduces the concept of text search: exploring the tree diagrams and searching for text in on-line documents. It is possible to search for words, strings, or patterns and save the findings in separate nodes. The tutorial searches for the word “dirty.” The results are displayed in a clipboard window and saved as a node. There is also a feature that allows the search to be spread beyond the word or string to include a specified number of text units (for example, 5 lines on either side of the word) so that the context in which the word is used is also displayed.

The last stage of the tutorial focuses on asking questions. This stage introduces ways of looking for connections in your data, and answering questions that the data pose. For example, by using a search operator called intersect, the tutorial examines whether unpleasant images of smoking are held by smokers and non-smokers alike. Basically, the software looks at the intersection of nodes SMOKERSNO and IMAGES/UNPLEASANT, and also at the intersection of nodes SMOKERSYES and IMAGES/UNPLEASANT. The findings are displayed in a clipboard window, showing all the references made by smokers to smoking as unpleasant, as well as the references made by non-smokers. This clipboard can be saved as another node.

This type of search and retrieve operation allows researchers to test emerging theory. This is only one example of the search and retrieve capabilities of the NUD•IST index search system, termed the “heart of the system” by the system developers. There are many other possibilities to be explored through the use of command files and such.

This software can be used by the fainthearted—which is an accurate description of this author when it comes to databases and anything that resembles complex computer stuff. It seems to me that this is an exciting tool for qualitative researchers: it will not do the work for you, but it will help you organize the data so that you can spend more of your time on the important work.

— Georgina Byrne
Qualitative Researcher, Doctoral Candidate
Wagner Graduate School for Public Service
**Indexing with NUD•IST** Project information is indexed by manually entering codes, or by performing a text search and then using the **autoindex** function. An index entry, or node, can refer to an entire document or to a part of it. You might attach a full interview to the index node “female” if the interviewee was female; or you could attach some smaller unit—say, a response to the question about her job—to node “job_type.” A range of text units including her responses about her job could be attached to another node, “bad_job” or “good_job.” The same document or document part may be indexed by multiple nodes.

Taken together, the index nodes of a project form a tree. Notes can be attached to the nodes themselves and separately to documents, allowing you to record your thoughts on both concepts and content. NUD•IST automatically adds log entries to node memos as you manipulate the tree structure.

Text searches include standard features, such as the use of wild cards, integers, and Boolean operators in search strings. You can limit the range of documents searched, or search the entire project. Search results are stored in a “node clipboard” which can be saved or used to create new nodes and branches.

For index searches, NUD•IST offers eighteen separate operators for specifying relationships between indices—including co-occurrence and proximity, as well as all the standard Boolean operators. For example, the overlap operator finds text coded with at least two occurrences from a list of codes that you provide, and the near operator allows you to further refine the search by specifying an acceptable distance between two co-occurring codes. Thus, “NEAR 5 ‘red wine’ ‘food’” would gather nodes that index text with the two strings “red wine” and “food,” and that are within five text units of each other.

**Exploring structure with index-based techniques**

With the command **make-node-report** you can display a report on a node and its subtree. Command parameters allow you to control output details like node names, attached documents, and their text, memos, and headers. The **display-tree** command also opens a window on the index system, centered on the node of your choice.

These tree windows provide point-and-click functionality. By clicking in a window, you can manipulate a tree without having to resort to menu commands or having to write a command file. NUD•IST also supports a command language that parallels the point-and-click actions. Results of tree manipulations appear in separate windows which can be saved and edited into command files. Unfortunately, full syntax rules are not included in the minimal on-line help provided in the package—so the manual is a must.

The nodes form a dual layer to the text collection. That is, the nodes may be thought of as text and concepts, and may be investigated in a parallel fashion to the textual data. In this sense, the index system in NUD•IST can be used to extend the code-and-retrieve technique. The user explores the text, the nodes, and the relationships provided by the coding. Such meta-study techniques can then be used in theory development.

**(Author’s Note: NUD•IST Version 3 for Microsoft Windows was reviewed for this article; Version 4 is soon to be released. NUD•IST is distributed by Qualitative Solutions and Research Pty Ltd., at La Trobe University, Victoria, Australia (nudist@qsr.com.au) and in the U.S. by SCOLARI at Sage Publications (nudist@sagepub.com). The cost to academics is $333.**

In writing this article I relied heavily on the User’s Guide for QSR NUD•IST, published by QSR, and on Computer Programs for Qualitative Data Analysis, by Eben A. Weitzman & Matthew B. Miles (Sage Publication Software Sourcebook, California: 1995). Additionally, I referred to the Handbook of Qualitative Research, edited by Norman Denzin & Yvonna Lincoln (Sage Publications, California: 1994), in particular the chapters “Data Management” (Huberman and Miles) and “Computers in Qualitative Research” (Richards).)

---

**New Statistics Group Media Machine**

The ACF Statistics and Social Science group has acquired an experimental dual processor NT workstation. This machine, dubbed the Statistics Media Machine (SMM), will serve as a demonstration machine for faculty and researchers interested in evaluating new computing technology. As a media "kiosk," the SMM supports many input and output devices. Several popular statistical and geographical packages will be installed, as well a CD recorder and modern high density floppy and tape devices.

Please note that contrary to information supplied by this magazine in its Spring 1997 issue, the SMM workstation will allow researchers to download the old format 1/2 inch tapes and cartridges still distributed by government agencies. For further information on data transfer, or any of the other capabilities of the SMM workstation, please contact Frank LoPresti (998-3398).
SUDAAN 7: Statistical Analysis Software for Complex Sample Designs

Robert A. Yaffee
yaffee@nyu.edu

The ACF Statistics and Social Science Group at New York University is currently acquiring a copy of Survey Data Analysis (SUDAAN version 7) for Windows 95. Produced by Research Triangle Institute in Cary, North Carolina, SUDAAN is a user-friendly package that saves time and assures analytical precision in analysis of complex survey data. SUDAAN is widely used by U.S. government agencies, including the U.S. Centers for Disease Control, the U.S. Food and Drug Administration, the National Center for Health Statistics, the National Center for Education Statistics, the National Institute of Drug Abuse, the Substance Abuse and Mental Health Services, and the Department of Agriculture.

SUDAAN facilitates analysis of data sampled with complex procedures, such as multi-stage cluster samples, stratified samples, and area probability samples. The program has multiple applications in a wide variety of crossover and repeated measures designs; epidemiological and clinical studies where adverse effects are observed over time on the same respondent or patient; and in sundry meta-analyses where precision is at a premium. It is customarily employed to analyze large scale national surveys, including the National Health Interview Survey, the National Health and Nutrition Examination Survey, Drug Abuse Resistance Education survey, and the Behavior Risk Factor Survey. These studies, unlike those with a simple random sample, generally employ sampling without replacement, stratification and/or clustering. Although simple random samples performed with replacement use a normal distribution, random samples without replacement are based on a hypergeometric distribution. SUDAAN employs an algorithm to obtain a robust variance estimator regardless of this sampling technique. If the sampling is performed without replacement and needs a correction for it, SUDAAN implements it.

Stratification is utilized to focus the survey on the target population. The population may be stratified by income, if income is deemed crucial in the analysis. A cumulative frequency distribution of median household income may be computed and the distribution may be divided into low, medium, and high segments. Random samples may be conducted within each of these three strata to assure a proper representation of the population income distribution. Although standard statistical programs such as Statistical Analysis System (SAS) or Statistical Package for the Social Sciences (SPSS) presume a simple random sample, variances within the strata are more homogeneous than those found in a simple random sample, a fact which can lead to biased significance tests. SUDAAN makes the adjustments needed to properly compute the variances and standard errors for such stratified samples and handles adjustments for single- as well as multi-stage stratification and clustering.

When large surveys are conducted, there is often clustering of the population into areas. Clusters are selected by random sampling, then random samples are taken within the selected clusters. The elements within the cluster are generally more highly correlated than those between the clusters. As the cluster size and intracluster correlation increase, cluster variances increase more than one would find in a simple random sample. In short, these effects lead to loss of precision and reduction of effective sample size. Unless these differences in variances are taken into account, the significance tests performed by the standard statistical packages will have a tendency to yield false positive results. SUDAAN permits the analyst to make the proper adjustments to obtain accurate significance tests.

SUDAAN can also handle surveys with correlated data from recurrent events, longitudinal data, or repeated data. SUDAAN allows the analyst to specify how the data are correlated or weighted. Because SUDAAN does not include a dedicated data editor, the analyst employs a separate editor to write the SUDAAN program. SUDAAN accepts ASCII and version five SAS data files. When the sampling weights vary substantially within the survey, the

Dr. Robert Yaffee is a research consultant with the ACF Statistics and Social Science Group.
variance of the sampling elements usually increases. SUDAAN can adjust for this unequal weighting. It can fit marginal or population averaged models using generalizing estimating equations. A Taylor Series approximation allows the analyst to obtain the robust variance estimator and produce the proper significance test.

Employing a syntax similar to that of SAS, SUDAAN can perform its own descriptive analysis. It can also perform crosstabulational analysis—computing frequencies, percentages, odds ratios, relative risks and their standard errors, along with chi-square and Cochran-Mantel-Haenszel tests for stratified crosstabulations. SUDAAN can estimate design effects of clustered samples and employs a variety of hypothesis tests designed to take into account the weighted analysis. Examples are the adjusted Wald F, which compensates for the liberal Wald statistic; the Satterthwaite adjusted Chi-square statistic, based on a weighted covariance matrix and used for cluster samples; and the Satterthwaite adjusted F, with its degrees of freedom correction.

SUDAAN analyzes a number of more sophisticated models as well. It can fit log-linear models. It can analyze ratios and their standard errors. Ordinary least squares, weighted least squares, binary logistic, cumulative logistic, and multinomial logit regression models may be run with SUDAAN. It can perform survival analysis with proportional hazards regression models. This version can run generalized estimating equations, with complex variance structures, as well. What's more, SUDAAN permits customized formatting of output.

SUDAAN accommodates the peculiarities of these sampling procedures by specification of the type of sampling design. SUDAAN handles three principal kinds of designs. Each type of design has specific first and subsequent stage capabilities. The first type has a first stage sampling with replacement (equal or unequal probability of selection) and a subsequent stage sampling with or without replacement (equal or unequal probability of selection). The second type has a first stage simple random sampling without replacement (unequal probability) and subsequent stages with or without replacement (equal probability). The third type has a first stage sampling without replacement (unequal probability) and subsequent stages with or without replacement (equal probability). Of course, SUDAAN can accommodate variations of these three basic themes.

In order to handle the various design types, SUDAAN design syntax includes a weight statement, which identifies the variable whose values are the sampling weights to be used in the analysis. It

<table>
<thead>
<tr>
<th>Region</th>
<th>School sample</th>
<th>Student sample</th>
<th>No. of schools in subgroup population</th>
<th>No. of students in subgroup population</th>
<th>Sampling weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>101</td>
<td>20</td>
<td>100</td>
<td>(3/20*3/100) = 222.22</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>102</td>
<td>20</td>
<td>100</td>
<td>222.22</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>103</td>
<td>20</td>
<td>100</td>
<td>222.22</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>201</td>
<td>25</td>
<td>125</td>
<td>(3/25*3/125) = 347.22</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>202</td>
<td>25</td>
<td>125</td>
<td>347.22</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>203</td>
<td>25</td>
<td>125</td>
<td>347.22</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>301</td>
<td>30</td>
<td>200</td>
<td>(3/30*3/200) = 666.67</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>302</td>
<td>30</td>
<td>200</td>
<td>666.67</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>303</td>
<td>30</td>
<td>200</td>
<td>666.67</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>401</td>
<td>10</td>
<td>150</td>
<td>(3/10*3/150) = 166.67</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>402</td>
<td>10</td>
<td>150</td>
<td>166.67</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>404</td>
<td>10</td>
<td>150</td>
<td>166.67</td>
</tr>
</tbody>
</table>
also includes a nest statement, which indicates the design stages. Total count variables indicate the variables whose values are the population counts at each stage of sampling, and sample count variables indicate the sample counts at each stage of sampling.

Sampling fraction corrections and variance adjustments are made with the help of a modification of the data set. The data are sorted according to the sample design nesting variables indicating the sampling stages. The SUDAAN data set includes special additional variables needed by SUDAAN to perform the adjusted analysis. The sample design statement specifies which variables contain the sampling weights, nesting, and total counts necessary for SUDAAN to perform its analysis. An example of a SUDAAN data file containing these nesting variables, including the sample and population total counts and the sampling weight variables along with the other variables, has the structure shown in Table 1.

Even though SUDAAN performs all of these kinds of wonderful analyses adjusting for correlated or stratified data, there are some caveats. The input data must be numeric. It must be sorted according to the nesting design variables. SUDAAN possesses limited data management capability, and it is suggested that the data set construction be performed in a standard statistical package prior to its use with SUDAAN. As for SUDAAN accepting SAS files, SUDAAN accepts only version five SAS files at this juncture. SUDAAN is properly advertised as handling some repeated data analysis—e.g., survival analysis, crossover repeated measures designs, etc. However, it only handles fixed effect designs. It does not handle the random effects in either mixed or random effect designs. Data sets have to be modified for SUDAAN analysis. Persons wishing help with technical problems concerning SUDAAN may obtain manuals from SUDAAN at the Research Triangle Institute in Cary, North Carolina, or may contact Robert Yaffee at (212) 998-3402 or yaffee@nyu.edu.

---

**Finding Your Computer’s Place on NYU-NET**

All computers connected to NYU-NET, and therefore to the Internet, have unique names or addresses. If your office or dorm computer is directly connected to NYU-NET, you may need to find this information—particularly if you want to check your computer’s network connection. Specifically, you will want to find your computer’s Ethernet card number and IP number.

Every computer directly connected to NYU-NET has a card installed in it called an Ethernet card. This is a piece of hardware you need in order to connect your desktop computer to NYU’s network. Each Ethernet card, regardless of manufacturer, has a unique twelve character name, composed of a combination of letters and numbers (you can ignore anything grammatical which breaks up this name, such as blank spaces, dashes or colons when you refer to an Ethernet number).

When your computer is added to NYU-NET, your computer’s Ethernet card number is entered into a table which links it to a unique IP (Internet Protocol) number. ACF designates the IP number. For instance, 128.122.abc.efg is a typical number for NYU computers. An IP number is sometimes described as a “dotted quartet,” because it has four parts, each separated by a period, or dot.

ACF also assigns a host name to your computer. This is just another naming style which designates your computer’s location on NYU-NET. For instance, one example of a host name for a computer might be computer10.schoolz.nyu.edu—there often is some reference to the department or school in the hostname.

In the box at the right you’ll find instructions for finding Ethernet card and IP numbers of typical desktop computers on NYU-NET.

—Lisa Barnett
Finding Your Computer's Place on NYU-NET

PC Compatible Computers

Windows 3.1 and 3.11

1. Open Trumpet Winsock (otherwise known as TCPman).
2. At about the fifth line the Ethernet address is listed.
3. A few lines below that the IP address is listed.

Windows 95

1. Choose Run from the Start menu, then type \winipcfg and click on the Run button.
2. This opens a dialog box which lists the Ethernet card number and the IP number.

Macintosh Computers

Mac OS before 7.5.2

1. Open MacTCP in the control panels.
2. While holding down the option key, click the mouse on the Ethernet icon on the top part of the control panel. The Ethernet card address is revealed underneath the Ethernet icon. The IP address will be listed lower down in the control panel.

Mac OS 7.5.2 and later

1. Open the TCP/IP (ver 1.1 or 1.1.2) control panel.
2. Click on the Info button on the bottom left side of the dialog box. The Ethernet card address and the IP address will be listed in the next window.
Washington on the Web: Federal Depository Libraries Go Electronic

David Hellman
hellmand@elmer1.bobst.nyu.edu

In a single term of Congress, Washington has entered the digital age. Perhaps the most startling aspect of this migration from paper to computer-based government information is that so many researchers are unaware of the fundamental changes underway.

We are all aware of the impact electronic information storage and retrieval systems have had on research, and we usually applaud the arrival of new systems and services, but the gradual trend toward using electronic resources to provide better access to government information has become a mandated torrent that threatens the very existence of the Federal Depository Library Program (FDLP). The real loss to citizens could be full public access to vital government information.

The Federal Depository Library Program

The FDLP was created to provide free and open access to government information. Under the control of the Government Printing Office (GPO), FDLP provides free copies of many government documents to over 1400 libraries. Participating libraries, including Bobst, agree to safeguard these documents and make them publicly available. The system of “depository” libraries was established through the Printing Act of 1895 and the Depository Library Act of 1962. The GPO maintains its authority to produce and disseminate government documents through Title 44, Chapter 19 of the United States Code.

Legislative Mandates and the GPO Study

In the current climate of budget-cutting, the GPO has become a prime target. A 1996 ruling of the Legislative Branch Appropriations Act (P.L. 104-53) presents the GPO with a mandate to replace its print publications with electronic versions by the end of 1998. If government agencies wish to continue producing paper versions, they will essentially be forced to use their own budgets to do so. The GPO was also directed to investigate and report on ways of easing this transition to electronic format. The resulting report, issued in June 1996, identifies the potential impact of the transition from paper to digital production and offers a number of proposals geared toward easing the process. It also recommends extending the transition period from print to electronic format to eight years, from the mandated two and a half.

The GPO has done a commendable job, considering the pressures it faces—there is now a possibility that Title 44 may be revised to eliminate the GPO altogether. While it seems possible that Congress will accept the longer transition period, it also appears unlikely that it will be persuaded on the issue of offering or financing multiple formats.

Uncle Sam Plays Catch-Up

Despite the role played by the Department of Defense in the creation of the Internet, most federal agencies have only recently begun to use sophisticated Internet applications. With World-Wide Web becoming the buzz phrase of the nineties, the federal government felt it had to catch up.

Believing that a “paperless government” will save time and money, Washington is following the worlds of education and business in embracing the new
technology. The only problem is that many federal agencies, many depository libraries, and much of the public itself are not technologically prepared for the transition.

**The Library Community's Response**

Many depository libraries and the public at large are not necessarily prepared for the rapid transition to electronic format. A 1995 survey conducted by the GPO (published in June 1996) showed that less than half of all depository libraries had access to the Web, and a quarter had no Internet access at all. Also in 1995, a Nielsen survey indicated that only 16.6% of the public uses the Internet. That means over 80% of the population would depend upon libraries for access to electronic information.

The initial response of the library community was one of great concern, but most are now taking a wait-and-see attitude. Some libraries are ahead of the game, while others are far behind. The real problem is the imbalance between those that are ready for the transition and those bringing up the rear. The once-reciprocal system of depository libraries—in which a depository library that lacked a certain item could borrow it from another, in the expectation of returning the favor later—threatens to become a system in which unfair demands are made on many of its members, precisely because they have invested the resources to be ahead of the game.

**Access to Electronic Federal Documents at Bobst**

Bobst Library’s U.S. Depository collection (on the sixth floor), is making a progressive effort to meet the challenge of disseminating documents electronically. Our Government Documents Workstation is upgraded regularly, and plans for a public-access Internet terminal are in the works. Currently, information from the GPO can be viewed and downloaded via telnet sessions on the library’s in-house network. In addition, some of the most important government resources are made available on Bobst’s Web site at [http://www.nyue.edu/library/bobst/research/soc/govtpol](http://www.nyue.edu/library/bobst/research/soc/govtpol).

Bobst Library’s challenge is to keep up with the ever-growing and changing flow of electronic government information. A source that appears today in paper may be available only in electronic format tomorrow, and accountability for these changes is far from perfect. If the current trend continues, U.S. government documents could very well become the first truly virtual collection within the library community.

The picture is not entirely bleak. A good deal of the federal government information on the Internet is useful and relatively easy to find. The GPO has developed an excellent World-Wide Web home page ([http://www.access.gpo.gov](http://www.access.gpo.gov)) that offers various ways to find government information. In addition, a number of federal agencies have produced exemplary Web sites, and many academic and a few public libraries have become user-friendly gateways to the world of federal information.

The problem remains that the process itself does not take into consideration certain needs and limitations of the public and the institutions that serve them. For the public, the task of finding information is now coupled with the challenge of mastering new formats. The challenge for depository libraries in this new world of electronic information will be managing the shifting burdens of finance and service that will accompany the transition.
NYU Computer Store Offers New Models for Spring and Summer

Kathy Bear  
beark@is2.nyu.edu

The NYU Computer Store is heralding summer with the introduction of new products and new vendors to enhance our offerings to the NYU community.

This past February, we celebrated the addition of Dell Products to our PC lineup with “Dell Day.” Representatives of Dell demonstrated their products in the Computer Store and raffled off a Dell Optiplex system to one lucky customer at the end of the day. You had to be there to win and the Computer Store was jammed with hopeful customers.

New PC models

Dell has grown to be a very successful player in the PC market, based on its competitive pricing and dependable systems. Under our contract, we can offer NYU excellent pricing on the Optiplex line of microcomputers. This is their mid-range line of computers based on Intel processors and all coming network ready with a 3Com 10Base-T Ethernet controller already installed. Processors range in speed from 133MHz up to 200MHz and models running from 166MHz up can be upgraded with an MMX processor. We are stocking three models in the Computer Store with a variety of options which should meet most computer shoppers needs. If you have a specific configuration in mind, different from the models that we are keeping in stock, we can order an Optiplex to your specifications. It takes Dell approximately ten business days to fill a custom order. We are in the process of becoming service certified on the Dell systems, and by the time you read this, should be an authorized Dell service center.

New Apple Models

Apple’s recent product announcements have added some new innovations to their lineup. The most notable of these is their PowerBook 3400. Powered by the 603e processor with 256K Level 2 cache, models of this laptop achieve speeds up to 240MHz, which is the fastest laptop processor currently on the market. This is a truly high end laptop with a 12.1-inch active matrix display, integrated floppy drive, CD drive on most models, video in and out, and a four speaker sound system. This full featured system is designed for the professional who needs access to a complete computer solution at the office and while traveling.

In addition to their new laptops, Apple has announced enhancements to their Power Macintosh line of desktop computers. Most of the new products are faster, enhanced versions of their existing lineup. With these announcements, Apple has innovated once again by including internal Zip drives on three of their models, making them the first major manufacturer to include these popular drives in a standard configuration. Both the Power Macintosh 6500 and the Power Macintosh 8600 models come standard
with the internal Zip drive. In the Power Macintosh 6500 models, Apple has added additional features to the entry level product. The 6500 comes in 225MHz and 250MHz models, running on the 603e processor. It comes standard with 32MB of RAM, a minimum 2GB hard drive, 12 speed CD drive, internal Zip drive, and Ethernet installed—offering a full featured package at a moderate price.

In the Power Macintosh 8600, Apple offers one model running at 200MHz. Since this model is powered by a 604e processor, it runs faster than the 6500 models which are built around a less powerful processor. It also includes 32MB of RAM, a 2GB hard drive, 12 speed CD, internal Zip drive and Ethernet.

With video in and out, this model offers a truly robust desktop machine capable of meeting most customer's needs.

Upgrade for the Summer
If you are thinking of summer tune-ups for your older equipment, consider taking advantage of our RAM sale going on now on specific SIMMs and DIMMs. With RAM prices increasing, this is an excellent opportunity to get more mileage out of older models. If you are thinking of a new computer, consider our trade up program which allows you to trade in your old working equipment for credit toward new Apple equipment.

ACF HelpLine Q&A

Q: I know that I have an e-mail address and that I can send and receive e-mail. But what is the difference between Pine and Eudora, since they both do e-mail?

A: Yes, it is true that Pine and Eudora are both electronic mail programs, however, they run on different computers. Pine is a text-based e-mail program which runs on UNIX computers, such as the ones used for the NYU-Internet accounts (is though is6). In order to use Pine, you would have to be logged into an account on a UNIX platform, such as one of the NYU-Internet computers. Eudora is an e-mail program which runs on Macintoshes and Windows-based machines; it functions like any other graphical desktop application. However, in order for Eudora to send and receive e-mail, you need a direct connection to NYU-NET from either an office or a dorm room (ResNet) or from a DIAL account which you can access using a modem.

Eudora automatically downloads your e-mail to your desktop computer; Pine, on the other hand, does not download e-mail your desktop. When using Pine, your mail stays on the UNIX host (is thru is6) until you delete it. There are advantages and disadvantages to using either mail program:

Eudora

Advantages: Eudora downloads your mail to your desktop computer, so anyone who has access to your computer has access to your e-mail messages, much like they have access to your files. This is especially important if you share your computer with a co-worker or roommate. Eudora can be configured to handle more than one user, but configuration and maintenance are not simple, and sharing such a setup requires a high level of trust among participants.

If you use Eudora on multiple computers, you will have to take extra precautions to sensibly manage your mail, because you may end up with particular messages available on one computer and not the other. There are strategies to get around this problem. You may choose to leave your mail on the server, or to save downloaded messages to a disk which you use across machines.

Pine

Advantages: As long as you can open a text-based telnet session to your NYU-Internet account, you can get to all of your saved e-mail messages. So, you can connect from a multitude of locations and still have access to your ongoing online correspondence.

Disadvantages: Because Pine does not save e-mail messages to your hard drive, you have to "export" them to a regular UNIX file and then use appropriate software, such as Kermit, Hyperterminal or FTP, to download the file to your desktop.

— Lisa Barnett
Non-linear digital video editing has changed the face of post-production. For a guide to choosing the NLE system to meet your needs, see page 3.