NASA Langley Research Center's Distributed Mass Storage System

Juliet Z. Pao and
D. Creig Humes

MS157A
NASA/Langley Research Center
Hampton, VA 23681
pao@subserv.larc.nasa.gov
humes@quickdraw.larc.nasa.gov

Abstract

There is a trend in institutions with high performance computing and data management requirements to explore mass storage systems with peripherals directly attached to a high speed network. The Distributed Mass Storage System (DMSS) Project at the NASA Langley Research Center (LaRC) is building such a system and expects to put it into production use by the end of 1993. This paper presents the design of the DMSS, some experiences in its development and use, and a performance analysis of its capabilities. The special features of this system are: 1) workstation class file servers running UniTree software; 2) third party I/O; 3) HIPPI network; 4) HIPPI/IPI3 disk array systems; 5) Storage Technology Corporation (STK) ACS 4400 automatic cartridge system; 6) CRAY Research Incorporated (CRI) CRAY Y-MP and CRAY-2 clients; 7) file server redundancy provision; and 8) a transition mechanism from the existent mass storage system to the DMSS.

1. Introduction

The Distributed Mass Storage System (DMSS) project at the NASA Langley Research Center (LaRC) integrates emerging technologies from the areas of data storage hardware, high speed communications, and mass storage system software into a system that overcomes the limitations of the current approach to mass storage. The DMSS is characterized by peripherals attached directly to a network, and a workstation acting as the file server. The file server will no longer be an active participant in most data transfers because they will occur directly between the peripheral and the requesting client.

The first phase is a prototype system to provide a proof of concept. It will also provide a base for testing ideas, and measuring and tuning performance. Once the prototype system is successfully completed, the production phase of the project will be initiated. This phase will include the procurement of necessary production storage and the addition of other functionality, such as network-attached tape.

2. Background

The Analysis and Computational Division (ACD) is responsible for providing a Mass Storage System (MSS) to meet the storage needs for both central and distributed computing systems at the NASA LaRC. The current production MSS is implemented on LaRC's CRAY Y-MP. The system consists of a CRAY disk and three STK 4400 robotic tape libraries. The disk is managed by CRI's Data Migration Facility (DMF) software. When it fills to a site specified threshold, the DMF automatically moves selected files to the STK libraries. Files that reside on tape are transparently moved back to disk upon access.

The main access method to the MSS is through a set of LaRC-developed Explicit Archive and Retrieval System (EARS) commands (masput, masget, masls, etc.) which allow the users to put,

get, list, move, remove, make and remove directories, and change attributes of MSS files. Files are transferred over the local area network to and from the CRAY disk. Users may also use the File Transfer Protocol (FTP) which is available for most network-attached machines.

The current MSS is typical of large scale mass storage systems in use today. Each transfer results in data flowing through the file server before arriving at its destination. In order to meet high performance demands, this server is usually a supercomputer or mini-supercomputer. Because of the high cost of this class of machine, the current system has limited expandability, scalability, performance, and availability.

3. Goals

The primary goal of the DMSS project is to move away from costly proprietary hardware and software solutions towards an open systems approach that does not limit expandability or scalability. The hardware and software purchased and developed for the DMSS must adhere to industry standards. This will facilitate expandability, scalability, and changes to hardware and software platforms. Software used and developed must be portable so that LaRC efforts and experiences can benefit other sites with common mass storage requirements. The system must be capable of providing high-speed access to files for selected client machines (i.e. the supercomputers), while not penalizing the performance of other clients.

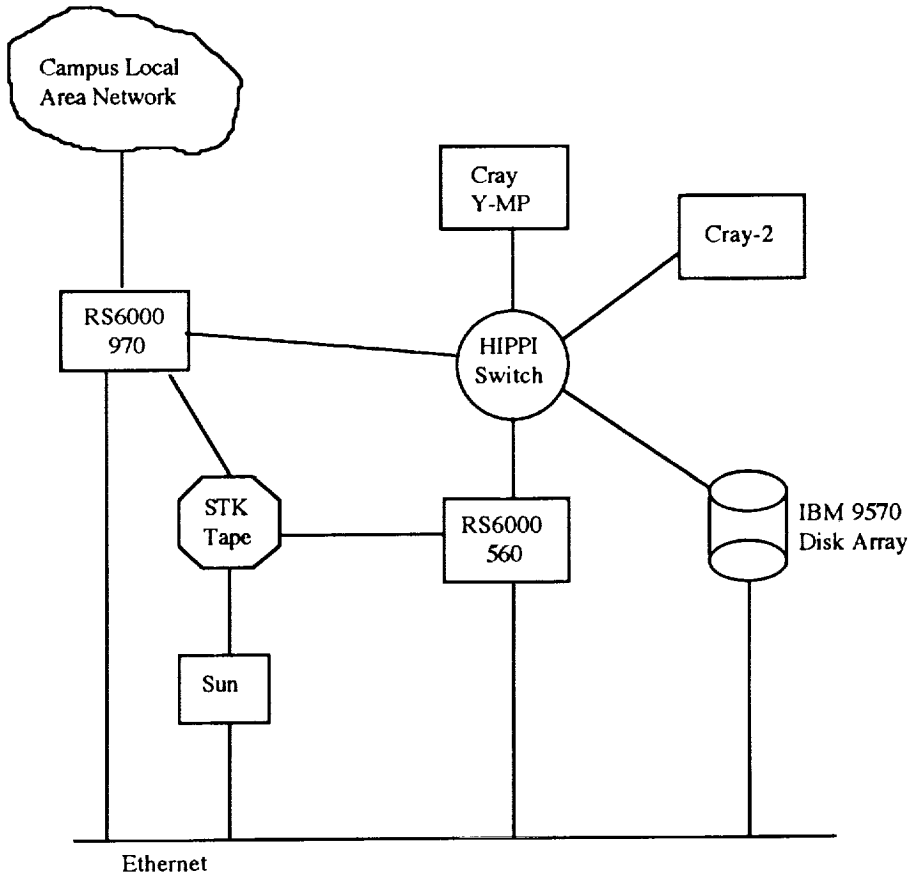

Figure 1

DMSS Prototype

4. DMSS Prototype

4.1 Equipment

The DMSS prototype [Figure 1] consists of an International Business Machines Corporation (IBM) 9570 disk array, two IBM RS6000 workstations (models 560 and 970), a CRAY Y-MP, and a CRAY-2. All of these pieces are connected to a Network Systems Corporation (NSC) PS32 High Performance Parallel Interface (HIPPI) Switch [1,3]. The workstations are also connected to the existing STK 4400 tape libraries through a SCSI interface. A separate ethernet network connects the workstations and the disk array. This ethernet is used for disk array control and tape mount requests to the STK Sun workstation.

The disk array uses the Intelligent Peripheral Interface (IPI3) protocol [4]. IPI3 commands may be submitted to the disk array via either the HIPPI interface (using HIPPI/IPI3) or the ethernet interface. Data can be directed to flow through either interface. The current disk array supports the Redundant Array of Inexpensive Disks (RAID) level 3 and supplies 40 GB of storage.

The file servers for the prototype system are IBM RS6000s. Each file server currently has 3.5 GB of local disk, 128 MB of memory, and HIPPI and ethernet connections.

The CRAY supercomputers act as clients in the DMSS prototype system. They request data transfers from the file servers. The CRAY-2 has one HIPPI channel and the CRAY Y-MP has two.

The PS32 HIPPI Switch allows up to 32 machines or peripherals to be connected. The switch allows multiple HIPPI connections without any degradation to standard HIPPI performance. Switches may be hooked together to provide more connections.

UniTree, a product of OpenVision, is a mass storage system software package which manages a storage hierarchy for files. UniTree is available on almost all open system platforms. We are currently running version 1.0 of the National Storage Laboratory (NSL) UniTree. The NSL modified version 1.7 of the general UniTree product and made numerous enhancements. The enhancements of particular interest to the DMSS project are support for HIPPI-attached disk arrays and multiple dynamic storage hierarchies. UniTree provides FTP and NFS interfaces to its filesystem and also supports distributing pieces of the system to different machines (i.e. one machine can support tape functions while another supports the disk cache).

4.2 Data Flow in the DMSS

Throughout the rest of this paper, components of the DMSS will be discussed in terms of the IEEE Mass Storage Reference Model (MSRM), Version 4, and the current evolution of Version 5 [5,7].

Clients of the DMSS that have HIPPI channels and the appropriate software drivers can take advantage of the speed of the disk array. These machines have bitfile client software which sends UniTree file transfer requests to the file server. UniTree then instructs the disk array to transfer data to/from the HIPPI port specified in the file transfer request. The disk array then initiates the data transfer with the requesting client's software component, called the mover, which moves data between the proper memory address and the HIPPI channel. The protocol used to accomplish the data transfer is IPI3 third-party [8].

Other clients of the DMSS, which do not possess HIPPI channels, cannot trade data directly with the disk array. For these clients, one of the file servers acts as an intermediary. The file server receives requests from them through a standard protocol (FTP or RCP). The file server then transfers data between the client (through FTP or RCP) and disk array (through IPI3 third

party). It is worth noting here, while hundreds of these clients exist and make use of the current MSS, they only account for approximately twenty percent of all data transferred.

The STK libraries are connected to the file servers and do not have HIPPI connectivity. During a file migration, a file server acts as a HIPPI client (as described above) to get data from the disk array before it writes the data to the tape. During a file recall a file server reads the data from tape before sending it to the disk array.

The initial user interfaces supported by DMSS include FTP, RCP, and EARS. All of these interfaces are explicit file transfer mechanisms which transfer complete files sequentially.

4.3 Redundancy

The approach for providing high availability is through redundant equipment. The production system will consist of two disk arrays, two workstations, and two HIPPI switches. This allows for the loss of any single piece of equipment without incurring lengthy down time. There are external SCSI disks that house the NSL/UniTree databases. Upon the loss of one server the other can be reconfigured to take over the functionality of the unavailable server, with access to the most up to date databases. The redundancy of equipment also allows for new system testing and development without impacting production use.

5. Prototype Development Work

The prototype system required LaRC to undertake development and integration work. The areas that needed development were IPI3 third party movers for the CRAY machines, user interfaces, and a mechanism to transition our current production system data to DMSS in an efficient manner.

5.1 Mover for the CRAY Y-MP with Model E Input/Output Subsystem (IOS)

In order to provide third-party transfer for the supercomputer client, movers have been developed for both user space and kernel space. The kernel version has been chosen for production use because it allows access to DMSS from multiple processes and fair sharing of the mover's system resource, the HIPPI channels. The user space version only allows one process to access the HIPPI channel at a time.

Mover Interface

The bitfile client, which is a set of NSL UniTree functions, communicates with both UniTree and the mover. It communicates with the mover by issuing transactions which consist of the following information:

function - action to be performed (such as read, write, or cancel)
transaction identifier - a 32-bit integer which uniquely identifies the transaction
buffer - a pointer to a buffer
length - the data length in bytes of the transaction
device index - the device index of the HIPPI device used for this transaction
status - pointer to a status structure associated with this transaction

When the bitfile client issues a transaction to the mover, it also issues a companion request to the file server which results in the file server issuing one or more IPI3 third-party transfer requests to the disk array system. The disk array system then sends the waiting client's mover one or more Transfer Notification Responses (TNR), each of which contains a Transfer Notification Parameter (TNP) with the following information:

transaction identifier- a 32-bit integer which uniquely identifies the transaction
offset- offset in bytes of this segment relative to the beginning of the transaction

length- data length in bytes of this segment
last_transfer_flag - flag to indicate that this request is the last transfer for the transaction identifier

The mover uses the TNP information to take action to complete the third-party transfer. One transaction request from the UniTree bitfile client may result in multiple TNRs due to file segmentation and system resource sharing requirements. The mover makes no assumptions as to the order of arrival or segment length of these TNRs. It also does not assume that all TNRs for a particular transaction identifier must arrive before it can handle the TNR of another transaction identifier. [8]

Mover Design

The mover maintains transaction queues and other information necessary to manage requests from multiple processes. The mover also maintains two kinds of internal buffers. It owns three large buffers used to receive the TNR and data, and many small ones used to store the HIPPI-FP (Framing Protocol) header and IPI3 command for a write request. The buffers are necessary because the mover must always be ready to accept a TNR for any transaction in the system.

The size of the large buffer limits the amount of input data coming from the disk array system via UniTree. As the buffer size increases, the number of HIPPI packets needed to perform the transfer decreases. An appropriate buffer size must be chosen to maximize performance and minimize waste of memory. The raw HIPPI driver on the CRAY Y-MP can handle a HIPPI write that has data split between two buffers. Therefore, the mover only needs to provide small buffers for the HIPPI-FP header and IPI3 command, and the user data does not need to pass through an intermediate buffer on a write. The size of the output packet is slightly larger than the user buffer size and is only limited by the maximum size of a HIPPI packet supported by the Model E IOS.

There is a set of commands to provide the following operational capabilities for the control of the mover:

- Initialize the mover environment.
- Halt all mover operations immediately (without shutting down the supercomputer client).
- Disable the submittal of transactions.
- Drop all active transactions.
- Close all HIPPI devices.
- Clear mover internal tables.
- Disable the submittal of transactions; all current transactions will be allowed to complete.
- Re-enable the submittal of transactions.
- Provide dynamic configuration capability for message logging options.
- Provide dynamic configuration capabilities for changing the time interval length for a transaction to be considered as timed-out and the time interval length to do the periodic checking.

5.2 Mover for the CRAY-2

The mover for the CRAY-2 is similar to that of the CRAY Y-MP, except for the handling of the third-party write. The raw HIPPI driver does not support a two buffer write. As a result, the mover's large buffers are used to pack the HIPPI-FP header, the IPI3 write command, and data into one contiguous area to be sent out with one HIPPI packet to the disk array system. So the bitfile client on the CRAY-2 can only submit requests to UniTree for transfers of size equal to or less than the large buffer size., Currently, the user space mover for the CRAY Y-MP has been ported to the CRAY-2. The porting of the kernel code began in June, 1993.

5.3 User Interfaces

The EARS commands have been rewritten for DMSS clients with HIPPI channels. These commands submit requests to NSL/UniTree using the supplied libnsl library. This library acts as the bitfile client and uses the LaRC developed mover for data transfer. This version of EARS is supported on the CRAY Y-MP, CRAY-2 and IBM RS6000.

Non-HIPPI attached machines have to retrieve their files from one of the file servers. These machines can get data either through FTP, RCP, or EARS. FTP is provided with UniTree. Two options are currently under investigation for providing RCP access. The first uses a locally modified version of RCP that understands how to talk to UniTree and the disk array (much like the EARS commands for the CRAYs). The second is to NFS mount the UniTree file system and use the regular RCP. The modified RCP currently works, but NFS with the disk array does not, so no comparison of performance is available at this time. The EARS interface is available to all distributed machines and is built using RCP for file transfers.

5.4 Transitioning From the Present DMF/UNICOS System to NSL/UniTree

The current LaRC MSS has more than a million files which comprise 1.5 terabytes of data on the STK ACS 4400 tape library under DMF management. LaRC has developed software that provides a mechanism for users to access any data in the current mass storage system on the first day of DMSS usage. The transition of DMF data into the DMSS is transparent to the users and requires minimal down time for the current system.

The day before DMSS production, the current mass storage system will be shut down for the transition process to take effect. First, on the CRAY Y-MP, a database called LaRCDB will be created using inode information of the current mass storage file system, the DMF daemon database, and the tape catalog database. The LaRCDB will then be moved to the file server. For each entry in LaRCDB, an entry will be created in the UniTree name server with a special flag set, indicating that it is a DMF formatted file. When a DMF file is accessed by a user via UniTree, the DMF flag will result in the tape file being staged onto UniTree disks using locally-developed routines incorporated into UniTree. After the staging, the DMF file becomes a bona fide UniTree file and its entry in the LaRCDB will be marked as soft-deleted.

While all the DMF files are available for UniTree users when they access them, not all of those files will be accessed by the users. So after DMSS is in production, a utility will be run on non-prime shifts to transition DMF files, cartridge by cartridge, into bona fide UniTree files until all files have been transferred.

6. Current Status

The prototype system is currently in a functional state. Test files are constantly being transferred, compared, and migrated. A majority of the effort now is spent testing and stabilizing the locally developed software and NSL/UniTree. The major items still in development are the CRAY-2 kernel mover and the transition software.

6.1 Performance of the DMSS

The initial tests of accessing DMSS data on the disk array system have been encouraging. The performance figures are grouped into three parts: disk array performance, file transfer performance to and from the CRAY Y-MP with Model E IOS, and file transfer performance between a Sun workstation and DMSS. The Sun is connected to the local area network via ethernet. The supercomputer's statistics were gathered on an idle machine, whereas the statistics for the local area network access were gathered in a normal production traffic environment. The IBM 9570 disk array system is configured using a 64K block size. All file transfer performance measurements include the whole transfer time between the client disk and the UniTree-managed disk array.

Disk Array Performance

Figure 2 shows the performance for the IBM 9570 disk array in both the first-party and third-party modes. Third-party performance was gathered using the CRAY Y-MP as the client and the IBM RS6000 560 as the file server. The performance includes the overhead of the command and response packets sent over the ethernet for control.

Complete File Transfer Between CRAY Y-MP and the DMSS

The timing measured is for file sizes of .5MB, 2MB, 16MB and 64MB, which are all block-aligned. Transfers that are block-aligned occur directly between the disk array and the CRAY. For non-aligned parts of a transfer, the file server is responsible for performing the transfer with the disk array [8]. In this case, the file server gets data from the CRAY's mover and places it on the disk array. This part of the transfer has been observed to take between 0.06 and 0.5 seconds.

Figure 3 compares the DMSS read transfer rates of different file sizes using large buffer sizes of 1MB, 2MB and 4MB. The graph for the 4 MB buffer case shows a decrease of performance as the file size increases from 16MB up to 64MB. This is due to the time necessary to flush the CRAY disk cache buffer. The performance of the current system is also plotted to show the increase of performance of DMSS.

Figure 4 compares the DMSS write transfer rates of different file sizes using large buffer sizes of 1MB, 2MB and 4MB. The write scenario is not limited by the large buffer size but rather the user level program's, namely masput's, buffer size. The graph shows that changing the user level buffer size from 2MB to 4MB did not yield a proportional increase of performance. The performance of the current system is also plotted for comparison. The CRAY's disk buffer cache was cleared before each transfer.

Figure 2 shows that larger buffers give increasingly better results. This is true for data transfers between the disk array system and the client's memory, but not for disks to disk file transfers. Both Figures 3 and 4 support the choice of 2MB for the mover's internal large buffer and user level program's buffer. Choosing buffer sizes larger than this gives rapidly diminishing returns due to the CRAY disk speed and the size of the CRAY disk buffer cache.

Complete File Transfer Between the LaRC Local Area Network and the DMSS

Figure 5 gives the statistics for DMSS access from a Sun workstation on the LaRC campus local area network. Masput and masget make use of the modified RCP (on the file server) which talks directly to UniTree. The performance of the current system is also plotted for comparison.

6.2 Schedule

Development will continue through the summer of 1993, along with debugging efforts for existing components and NSL/UniTree. Internal test users will begin making use of the system sometime in August and will use the system for a two month evaluation period. If the system is stable at this point selected users from the research community will be invited for a one to two month beta-test, followed by full production use by the entire research center. A second 40 GB HIPPI-attached disk array, external SCSI disk, and second HIPPI switch will be added to the configuration before production usage is initiated.

First Party vs. Third Party Transfer Performance of
the IBM 9570 Disk Array System Involving Cray Y-MP

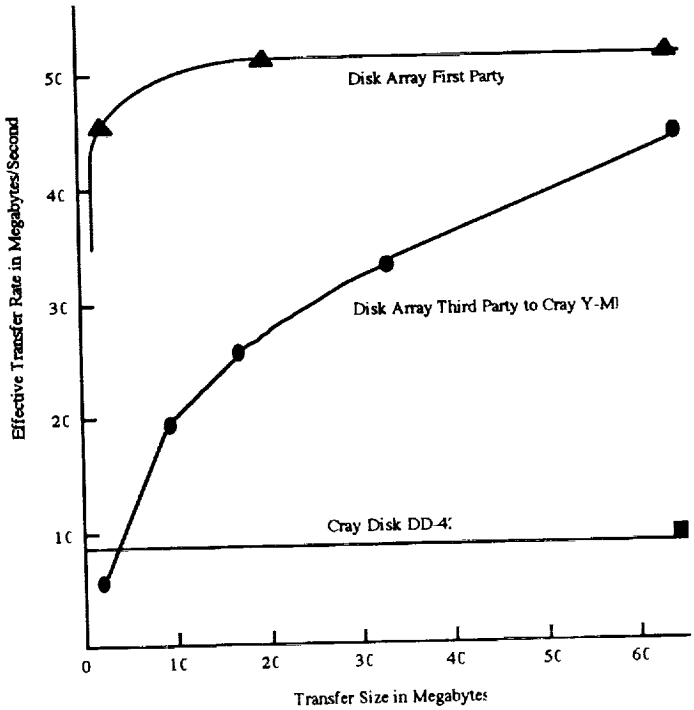

Figure2. Performance comparison among the first party disk array transfer rate provided by IBM, the third party disk array transfer to/from Cray Y-MP using LaRC mover, and the sustained transfer rate of the Cray DD-42 disks.

Transfer Rate Between Cray Y-MP & DMSS Using Masg

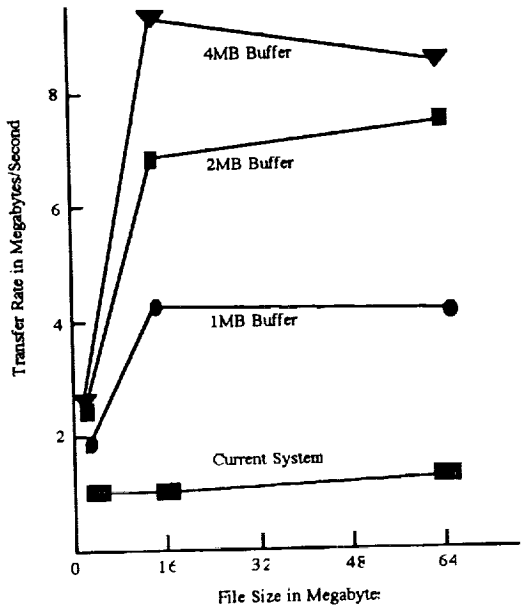

Figure 3. Transfer rate comparison of masget using differe sizes of buffers on the Cray Y-MP.

Transfer Rate Between Cray Y-MP & DMSS Using Maspu

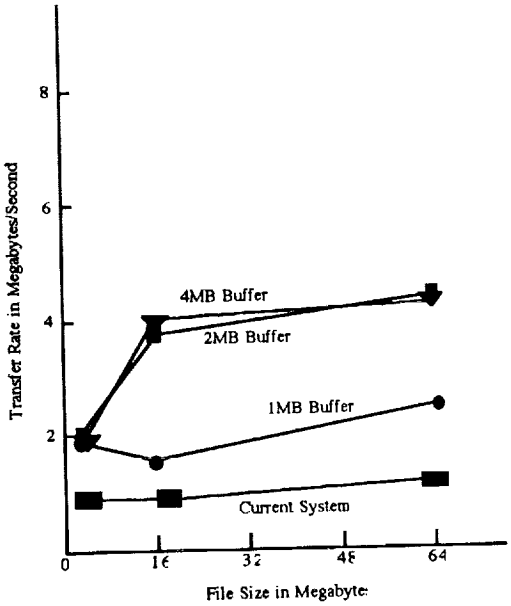

Figure 4. Transfer rate comparison of masput using differe buffer sizes on the Cray Y-MP.

Transfer Rate of Local Area Network Access
Using Modified RCP

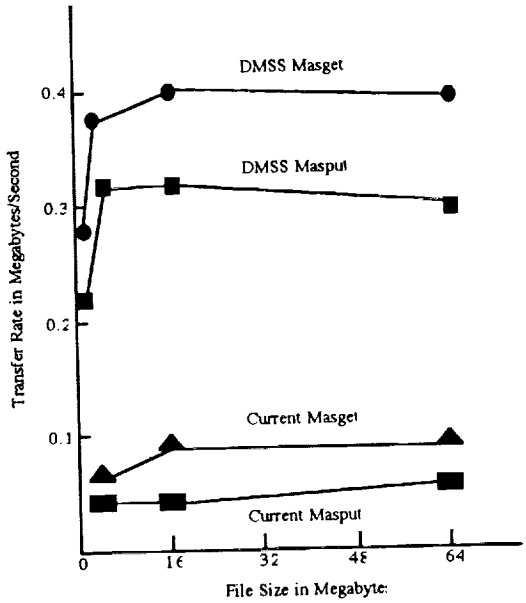

Figure 5. Transfer rate comparison of masput and masg used from machines on the LaRC local area network.

7. Future Plans

Once DMSS is stable, other features will be added. Of particular interest is a file system interface (using vnodes). The first supported interfaces are all disk-to-disk file transfers. There is also a need for high performance data transfers directly between an application on the CRAYs and the disk array. Currently the only way to do this is to incorporate the libnsl routines directly into a program. This does not give the users file location transparency, thus placing an unnecessary burden on the users. A transparent file system interface would allow for extremely good performance for jobs running on the CRAYs, while maintaining location transparency. In this way all permanent file storage for the CRAYs can be managed by DMSS.

Also of interest is a site-wide distributed file system that will be able to use the DMSS to store data. For example, this could be based on OSF's DCE/DFS.

Other machines with HIPPI attachments will have movers developed to enable high speed DMSS access. The next machine targeted is the Intel Paragon.

LaRC will also pursue adding network-attached tape to DMSS. This will relieve the workstations of more than 95 percent of the data transfer responsibilities of the current CRAY Y-MP based MSS. Migrations and recalls will occur directly between network peripherals. As the multiple dynamic hierarchies mature, applications, such as backup and visualization, will move data directly to and from the network-attached tape.

8. Conclusion

When DMSS goes into production in the fall of 1993, it will relieve the CRAY Y-MP of its function as a file server. Users of DMSS will experience performance three times better than the current system. Their access to DMSS will no longer be interrupted by the file server's unavailability due to various system maintenance functions, malfunctions, or system time. The system will be expandable and scalable. Disk and tape will be added directly to the network as the need grows. If one file server is not powerful enough to handle the workload, then the function can be split among two or more file servers.

9. Acknowledgments

The LaRC prototype DMSS system has gone through the cycle of design, acquisition, testing and software development since January 1991. The acquisition took the initial one and a half years. We would like to acknowledge Everett C. Johnson and David E. Cordner of the Computer System Branch at NASA LaRC for their help in the design and acquisition of DMSS equipment, the Unisys Cooperation for their support in software development and testing, and CRAY Research Inc. for their support on the UNICOS internals. We also appreciate the cooperation of DISCOS of General Atomics (presently OpenVision) and IBM Federal Systems Company.

References

1. ANSI, "High Performance Parallel Interface - Mechanical, Electrical, and Signaling Protocol Specification (HIPPI-PH)", American National Standards Institute, X3.183-1991.

2. ANSI, "High Performance Parallel Interface - Framing Protocol (HIPPI-FP) Preliminary Draft", American National Standards Institute, X3.210-199x.

3. ANSI, "High Performance Parallel Interface - Physical Switch Control (HIPPI-SC)", American National Standards Institute,X3.91-023-1991.

4. ISO/IEC, "Information Technology - Intelligent Peripheral Interface Part 3: Device Generic Command Set for Magnetic and Optical Disk Drives", ISO/IEC 9318-3, September, 1990.

5 Coleman, S. and S. Miller, eds., "Mass Storage System Reference Model Version 4", IEEE Technical Committee on Mass Storage Systems and Technology, May 1990.

6. Coyne, R. and H. Hulen, "An Introduction to the Mass Storage System Reference Model, Version 5", Proc. Twelfth IEEE Symposium on Mass Storage Systems, Monterey, April 1993.

7. Merrill, J., "Toward a Standard IEEE Mover", Proc. Twelfth IEEE Symposium on Mass Storage Systems, Monterey, April 1993.

8. Hyer, R., R. Ruef, R. Watson, "High-Performance Data Transfers Using Network-Attached Peripherals at the National Storage Laboratory", Proc. Twelfth IEEE Symposium on Mass Storage Systems, Monterey, April 1993.

Invited Panel: User Experiences with Unix Based Hierarchical File Storage Management Systems

DR PRATT: The Panel moderator will be Dr Sanjay Ranade, who has a bachelor's degree in aeronautics and a Ph.D. in computer science. He worked at NASA/Goddard for eight years. He helped to design and develop a high-performance network fileserver for Hughes STX, and now has his own company, Infotech S.A., Incorporated.

Sanjay?

DR RANADE: Thank you. Can everybody hear me okay? I'd like to start off by introducing the panel. The topic is User Experiences with Unix-based Hierarchical Storage Systems, and we're going to refer to these things as HSM or File Servers or whatever. But that's the main topic-Unix-based only.

The first person I'd like to introduce is Mike Daily. I won't go into a big discussion of him because he was already introduced earlier. Mike is from Mobil, and he has experience with the FileServ software.

The next person is Ellen Salmon, who works with Hughes STX supporting NASA's Center of Computational Science. She's a principal systems programmer, and she has worked one and a half years with the UniTree system on the Convex machine at Goddard. Prior to that, she has eight years software support experience, also at Goddard.

John Garon is a computer scientist at NSA. He has an MS in computer science and a BS in mathematics. He's been developing software for data archive data bases and software analysis, and he has experience with Advanced Archive Products'AMASS software.

Thomas Woodrow is from NASA Ames Research Center. He's a Scientific Analysis Software group leader. He has a BS in computer science from Hobart College and some very apt experience here, because he was recently asked to perform an evaluation of the Unix-based HSM software and he has written up a nice paper which we had a chance to look at yesterday. I am sure he will be telling us of his experiences. Included in his evaluation were DMF from Cray Research, UniTree, FileServ and Nastor.

Joe Marsala is from the Supercomputing Research Center in Bowie, Maryland. He has a BS in mathematics from Texas A&M, and he has worked with the EPOCH storage management software over the last few years.

Suzanne Kelly is from Sandia National Labs in Albuquerque, New Mexico. She is a Distinguished Member of the Technical Staff there. She has a BS in computer science from the University of Michigan and an MS in computer science from Boston University. Sue is the president of the UniTree Users' Group. She has ten years' experience maintaining HSM software storage systems. She's very well known in the UniTree community. She's involved in the HPSS software development work for the National Storage Lab.

So, having introduced everybody on the panel, I just want to give you a summary of how we are going to try and do this panel discussion. The first thing is I'd like each panel member to just introduce themselves, what they do, what their installation is like, basically give a little synopsis of their experience there.

Then we have a bunch of discussion topics. After we've been through the panel, each one describing their experience and so on, we have ten discussion topics. We will step through each one, one by one, and I will ask the panel members to comment on it. Anybody in the audience who wants to, can chime in and say whatever you like. You can ask questions at any time. Don't be shy. Just raise your hand and ask whatever you like.

Let's try to keep this really informal and productive and interactive so that we have more of a dialogue rather than people here lecturing to people over there. Let's try to keep it informal.

So, why don't we start with Mike? Do you want to say a few words about yourself and your installation, and we will go on down the line here?

DR DAILY: Well, I'm a geologist by training, so I don't know that much about all the technical aspects. As I said in the talk, we're FileServ based, with a Convex front end. The evolution that we see is that we will have direct connection in due course to things like the Connection Machine. Our installation is intended to be very diverse, so it is supporting not only supercomputer-type processing but also wide-area access by workstations, and also data archiving.

Our definition of archiving is not deep storage; it's more sort of a back end store for what will eventually be several hundred terabytes of data. We are committed entirely to open systems. So we started this thing in the Unix world and have no intention of moving from there. So in that sense, I guess we're not carrying a whole lot of baggage with us.

What were some other -- we're not going to turn to the ten questions yet, are we?

DR RANADE: No.

DR DAILY: Okay. So those will come out in due course. I guess that will do as a capsule summary of what we're up to.

MR WOODROW: I'm Tom Woodrow. I am a manager for Computational Fluid Dynamics (CFD) Visualization Developers and Parallel Software Tool Developers. I provide support for users who are trying to analyze CFD data sets which range in size from 50 GB - 1 TB. In an attempt to support users with very large data sets, I borrowed a Storage Technologies robotic tape silo, attached it to an existing Convex Visualization System and ran a UNIX-based HSM called Convex Storage Manager (CSM).

Later, when our organization needed to make a decision on whether to go into production with a home grown HSM, NAStore, or a commercial alternative, my experience and the fact that I was not involved with Storage Development made me an ideal candidate to conduct the review.

Our environment consists of 2 Cray C-90s which generate CFD data sets. We currently have 2 production HSM systems deployed at the center, one is a dedicated Cray YMP2E running Cray's Data Migration Facility (DMF), the other is one of the C-90s which runs DMF to keep scratch disks relatively free. The use of DMF on the C-90 system is tolerated because it allows us to keep scratch disk space free and the CPU load does not appear excessive. We are about to place 2 dedicated Convex C3820s into service running NAStore, a locally developed UNIX-based HSM. The volume of data and daily flow into these systems is approximately as follows:

YMP2E	1.3 M files, 5 TB, 7 GB/day
C-90	31 GB/day
Convex	2.2 M files, 3.7 TB, 4 GB/day

MS KELLY: Hi. I'm Sue Kelly, and I wanted to talk to you about what Sandia National Labs' Scientific Computing Directorate has for file servers. We have four file servers, two in Albuquerque, New Mexico, two in Livermore, California. In each site, one is doing classified file serving and the other is doing unclassified file serving.

All four systems are pretty comparable in architecture. They're all based on Convex C2 or C3 CPUs. They have on the order of 100 gigabytes of disk on each of them, and they have one or two Storage Tek silos as the archive. They interface to networks via FDDI and two of them that interface also have an UltraNet connection to Cray Y-MPs.

For client nodes, there are approximately 500 on the classified network, and 500 on the unclassified network. The nodes are one Cray on each of those networks and an assortment of HPs, Suns, Macintosh, Silicon Graphics. And that's pretty much the hardware side of it.

On the software side, all file servers use the UniTree 1.5 version from Convex. That means the access methods are NFS and FTP to the UniTree system. In the case of the Crays we have UltraNet, because we use native Ultra FTP to communicate, and that provides us higher performance. So in deference to previous speakers, we do not have performance requirements problems with UniTree systems. They satisfy our needs, and in all cases it has been networking, protocol stacks or the client that has been the slow part of the file transfers to the UniTree system.

For the rest, when I discuss these systems -- because there are four of them, it gets confusing to differentiate to you, who certainly don't care about my four systems -- I'm going to refer only to the one system which has the longest lifetime. It's been around for 18 months, has 1.2 terabytes of data on it. It averages only about 5 gigabytes a day of traffic and grows by about 1 gigabyte per day.

This system manages about 277,000 files. I think that's about all I wanted to say for the hardware and the software environment. I look forward to your questions.

MR GARON: I'm John Garon. I have been working at NSA for 18 years and, for the last 12 years, one of my responsibilities has been in the area of mass storage systems. I began by developing software to interface to the Bragaen Automatic Tape Library systems attached to CDC Cyber 176 and Cyber 84's. Although we were using commercial equipment, we had to develop all of our own software since we used a home-grown operating system and programming language. With the introduction of UNICOS around 1987, we began to explore the use of commercial software to replace the storage control software used to drive our hardware storage systems.

In the late 1980's, my office initiated the ABUNDANT requirements that Mike Shields talked about earlier. My office is no longer the customer for that project, and it is now being developed for another NSA customer.

We are still using internally developed software to perform file and volume management on our main storage products, the STK silos, using Crays as the host. My concern is that the software will become unmaintainable as people leave the project and that it will eventually not satisfy our growing requirements. In the late 1980's, we thought that hardware would be the limiting factor in solving our storage needs, but over the last few years, it appears that industry will develop the hardware and storage capacity to satisfy our requirement. The problem seems to be in the software to control the hardware, and to perform file management to those high density robotic systems.

I have had experience with the AAP product in our office on optical and Metrum VHS tape systems. Although the AAP applications do not store nearly the amount of data that goes to our silos, the functionality of the product is very close to our requirements. My problem is in the control software, and I think the product that will satisfy my requirement has its basis in the AAP product. I have plans to begin working with my systems developers at NSA to determine whether it is desirable to have the AAP software enhanced to work on Crays to interface to our silos.

MR MARSALA: Hi. I'm Joe Marsala of the Supercomputing Research Center. We're a relatively small research house, about 140 people. We have 300 workstations, a Cray 2, a TMC CM2 and a TMC CM3. Our backbone network is FDDI. We've got about 13 or 14 relatively large servers, and one of those servers, our archive server, is an EPOCH 1 Optical Jukebox System. After hearing all the massive storage requirements here, I think I'm here to provide the comic relief.

MS SALMON: Hi. I'm Ellen Salmon. I work for Hughes STX supporting the NASA Center for Computational Sciences at Goddard. We have about 1,200 space and earth science researchers who are users of our facility. So they have a great divergent group of requirements themselves.

The facility itself has a primary compute server in the form of a Cray C98, and it has six processors. That itself has a Storage Tek silo and runs DMF for a 21-day archive. After 21 days, the data is purged from that system.

We are running UniTree 1.5 on a Convex C3820. Our Convex/UniTree system has three Storage Tek silos that are within a couple of hundred tapes of being completely full. We have about 5,000 vault tapes from our UniTree system. We've got about 105 gigabytes of disk cache. We have about 3.3 terabytes stored at this point.

Our UniTree system has been operational a little over a year at this point, so we've gone from nothing to 3.3 terabytes in a year's time, and one of our big issues, of course, is handling that volume of data. We do have UltraNet connectivity between the C98 and the Convex. UltraNet is the route where most of our data comes, and the Crays are the primary storage client.

We are expecting, as far as requirements are concerned, that our transfer requirements are going to have to be even bigger than they are now. At the moment, we see in the neighborhood of 50 to 70 gigabytes of traffic a day into and out of our Convex/UniTree system. Depending on the day, we can see more gets than puts. We allow only FTP access for reasons of performance. Recently, we've been seeing on the order of 30 to 50 gigabytes of new data a day.

Probably our primary concerns at this point are issues of network robustness and the ability to write enough data tapes fast enough to keep up with the data coming in from the Cray; we're also very concerned about the fact that we also have an IBM system from which we have 1 to 2 terabytes of data to transfer into our UniTree system. Right now, we're going from 3480 technology on the IBM system to 3480 technology at this point with our UniTree system, and we'd like to see higher density. So, at least we would have our storage in a smaller area and not just moving the data from one kind of system to another.

DR RANADE: Okay. Before we get to the questions, I just want to say a couple of things to set the background, as it were. First of all, the market for mass storage systems, many people look at it as being composed of three segments: the small segment, your workstation, LAN, file server; the middle segment, which is often commercial market; and the top end, high performance, high capacity, which some people call the lunatic fringe of the market.

So that's one way to break down the mass storage picture. Another way is by the type of system that's really needed in a given case, and the four cases that I can say, there is what is called the virtual disk, which is just one machine with extended storage. There's the network file server. There's the backup and recovery, and there's client migration. There are four different kinds of software there.

I just said that because we're not comparing apples and apples, and we're not talking about the same thing. We're talking about different kinds of software for different requirements. So having said that, I'd like to ask each of the panel members how they developed their requirements. What process did you go through to come up with your requirements? Or did you go through a process to come up with your requirements?

Mike?

DR DAILY: I guess looking back into the deep dark past -- five or six years ago is when we started doing this. Originally, our use of this technology was envisioned in the grand scale, which is kind of how it's turning out. And then about halfway through its life or the development cycle, it got sort of pinched down, and then it has subsequently re-expanded. So let me just mention that.

Five or six years ago we looked at it primarily as a back end to supercomputers and as a replacement for the tape library. So the idea was that we had this pretty compelling economics of projection of a couple million tapes sitting off the tape library with capital costs of that running $20 million or $30 million just for media and $4 million to $5 million a year for managing those tapes.

I don't know if any of you have ever worked with round tapes especially. You actually have to be like these people at brindle champagne where you go in and quarter turn them every three months to straighten out the magnetic flux lines and all that, some sort of weird physics involved in these large amounts of magnetic media.

So the two drivers at the time were replacement of the tape library and the back end for the supercomputer. With E Systems we did a lot of numerical simulation about how many recorders and latencies and all that sort of thing and put that case together.

At the time we also recognized that there would be a future need for things like serving workstations over wide area networks, but that was not explicitly part of the justification. About halfway through the project the focus narrowed just to replacing the tape library, so there was little attention paid by the people that were managing at that time on these other things.

Then about a year ago, things opened back up again. So I guess the long and the short of it was that there was a lot of thinking done, constructive thinking, and now it has rewidened with all these opportunities which have come available, especially with faster workstations.

DR RANADE: Is it possible for you to say what proportion, what is the ratio between the money you spent on developing your requirements compared to the money you spent in buying the system? The reason I ask is my own experiences, having worked with the procurement, which is about a million dollars, the government spent $200K on developing the requirements and doing the spec. So how does this compare with your experience?

DR DAILY: Multiply that by ten in both cases and you're about right on.

DR RANADE: Okay.

DR HARIHARAN: (Off microphone.)

VOICE: Is that a later question?

DR RANADE: Yes, we can come to that.

DR DAILY: Do you want me to go ahead, though?

DR RANADE: Go ahead, yes.

MR WOODROW: Okay. I wasn't around at least to develop or participate in the requirement discussion for how we got going. I can talk a little bit about what model I know we use. We've been driving the requirements for how much storage space we needed basically by the solution development capability that we have in the Crays. We have an idea of how fast the systems are, what the canonical grid size is for a CFD data set and about what we can produce per day.

Unfortunately, most of the data that is produced is saved forever, whether it's good or not. So we're not terribly aggressive to go out and get people to throw away the data set that they don't really need to keep. At least I know that that's part of the model. So we're talking with users to identify how big their data set is and we multiply by the capability that we have to produce on the Cray.

One of the factors that's making things more difficult for us now is we're going from a situation where people are generating a single time step to generating a hundred or ten thousand time steps. So we're seeing that individual users are increasing their output tremendously, and what they want to look at later.

Okay. So that talks about at least how I believe we derive the requirements for the production systems that we have on the floor. For the purpose of the evaluation that I ran, I did the same kind of evaluation. I sat down with users. I talked to them about data set sizes, and I also took a look at the population breakdown for what we have on our production system. Then I put together a workload that reflected user needs and population breakdown.

DR RANADE: Does anyone in the audience have anything to say at this point on requirements? Any comments? No? Sue?

MS KELLY: Yes. We did a very detailed requirements study in order to purchase the system. It was a competitive bid, so the requirements study was translated into a Request For Proposal. We spent approximately $300,000 for that requirements study which resulted in an acquisition of $3.3 million. Different color money, however, capital versus expense.

DR RANADE: So it's 10 percent roughly.

MS KELLY: Yes.

MR GARON: I have no idea what it cost to gather our requirements. We have two sites that I am familiar with using the AAP AMASS product, and I was not involved in either procurement process. My office has the AAP product controlling two optical units. The other site has two Metrum systems, a 600-cassette and a 48-cassette system.

How we got the AAP product in our office was rather by chance. I stumbled on the two optical units that were a by-product of the ABUNDANT program. They were not being used, so I borrowed them and discovered that the systems were managed by the AAP software. So we re-initiated the AAP license and found, to our surprise, how functional the software actually was. Now we are investigating other platforms where the AAP product may be useful.

MR MARSALA: Well, see, at SRC, my group's function is primarily to support our research user population. So we basically developed requirements by talking to those users, looking at some historical data, and a scan of available technology. I couldn't give you any idea what it wound up costing. The evaluation assistance later wound up being a whole lot more than gathering the requirements.

MS SALMON: I wasn't in on the whole procurement process, but I understand ours was one that started five years before the final product was accepted, kind of a large-scale government procurement type of thing where, at least initially, I think the need for storage, *et cetera*, was greater than what was available on the market.

As far as requirements, we had an existing, and still have an existing, IBM MVS-based HSM system. Processing done on that system was primarily satellite data calibration, *et cetera*, very I/O intensive work. The other big use of data, of course, is our supercomputers, the Cray C98 at this point in time. I know the procurement process was pretty thorough in trying to understand what the satellite processing requirements were going to be and including major users and asking for their trends and trying to look into the crystal ball and seeing what the computers, the supercomputers, of the future were going to require.

That's pretty much what I can tell you about our requirements.

DR RANADE: Anybody in the audience from Goddard who has something to say on the requirements development? Because Goddard had an interesting experience. They purchased one mass storage system, and then they bought another one. And I think a lot of it had to do

with the requirements being reformulated or whatever. Anybody want to comment? No? Okay.

The second one -- and let's start with Tom -- how did he develop acceptance tests or benchmark tests? Did you have a need to have acceptance testing or benchmark testing? Did you write your own? Did you go and talk to other people, borrow theirs?

MR WOODROW: For the HSM evaluation I just completed, I created my own set of benchmarks to stress disk, tape performance and that of the HSM product. These tests included individual peak performance tests as well as a simulated user workload. I was interested in pointing out differences between several alternatives rather than testing out a system before it went into production. Our Mass Storage Groups ran extensive acceptance tests on NAStore, the system we recently placed in production. These tests were oriented towards verifying functionality and performance, reliability, stability and failure testing, and an extensive beta testing period. We had access to the experience of two production HSM capabilities on site and were able to develop extensive test suites.

DR DAILY: Going back to requirements for just a second, I wanted to know if any of the panelists had the experience of, in the process of the requirements, having seriously underestimated their total capacity? Have they filled up their systems dramatically faster than they had originally anticipated? Or were they always aware that they were dealing with a very short time constraint? Because it sounds like a number of the panelists are already pushing up against the limits of their existing systems.

MS SALMON: It's my understanding -- and once again, I wasn't in on all the details of the procurement for our particular system, but if the money had been available-- by the time things finally came through, we would probably have initially obtained two to three times the storage we have now with, of course, growth capability. So, to a certain extent some people felt that we had overestimated the rate at which we would be storing data. But it's pretty much gone according to those who felt we were going to be storing more data than what prevailed budgetwise.

DR RANADE: Okay. Anyone else?

MR WOODROW: We're also seeing data coming in from other sources than we had earlier anticipated, so that's not a major increase. But we did not expect to see the volume of data saved on the Cray that we are seeing, and it's causing us to rethink the way that we stay in production with our service.

MS KELLY: For our requirements, every capacity requirement also had a requirement for an order of magnitude expansion beyond what was already there. So we bought a system that can be expanded quite readily.

DR DAILY: I think that's pretty much our experience, too. We chose the solution that we did because of the very large dynamic range and size. I think we are pretty much on track for the sizing that we did but for the wrong reasons in the sense that we anticipated 200 or 300 terabytes a few years out. That was based on the idea that we were going to transcribe the existing million and a half tapes in the tape library, because no one has the guts to throw away existing data.

Well, since then, with the travails of the oil industry, people have gotten a lot more courageous about it. We're putting them into deep storage in salt mines. So, our guess now is we're only going to transcribe about 20 percent of that one and a half million or so, but it's being made up for by the much higher data densities that we're getting in seismic acquisition now, gigabytes per kilometer of line mile, that sort of thing.

DR RANADE: Moving to the next topic: how did you evaluate the software? What process did you go through? Can we step through, let's have a bit more speed, because we've got a lot of topics, and we're at 5:00 o'clock now, I think, aren't we?

VOICE: 5:30.

DR RANADE: But right now it's about 5:00?

Okay. Sue, do you want to start on this one?

MS KELLY: Well, it was a competitive bid, so that's how the evaluation was done. To kind of pick up on question number two, we did develop a set of benchmarks for evaluating the various solutions that were offered to us. The evaluation and the benchmark criteria were part of the $300K investment we made in the requirements.

MR WOODROW: I can say a couple things.

MR MARSALA: Well, we didn't do a benchmark per se. We took the requirements that were at a more functional level and did a validation/evaluation of all of those, including some transfer times and that sort of thing. But that was basically the extent at which we evaluated it.

MS SALMON: For us, the part of the procurement was also acceptance criteria. Basically, the product had to satisfy the acceptance criteria, and there's a list of them. We kind of had to go through one by one and show that they could be met.

DR RANADE: Mike, do you want to go?

DR DAILY: Our selection was really driven by some of the requirements for the media itself. We have pretty stringent requirements on bit error rate, like 10^{-12}. We needed bandwidths of 10 to 15 megabytes per second. The large capacity per cassette is to minimize the handling, so we wanted these 10-, 20-, 30-gigabyte cassette sizes instead of sub-gigabyte, and the scalability left up to libraries.

At the time that we really got into writing checks and things like that, about the only thing that we saw out there was the stuff that our cousins in Fort Meade are doing. So I think it ended up being pretty much of a sole-source sort of thing.

MR WOODROW: We had to justify why we would continue going with Nastor as opposed to one of these commercial alternatives that certainly are getting a lot of use. So we brought in UniTree, we brought in FileServ and DMF and ran them on systems in-house for about three months while also running Nastor. We ran a number of different benchmarks across all of them, and then we basically rated all of them for performance, functionality, ease of use, stability as much as we could determine in a short period of time, and ranked them and made a decision: in the end, to stick with Nastor since there is no additional development that needs to be done. Basically, because of a cost decision at the end, it's the lowest cost one for us to go with.

DR RANADE: It was the lowest cost one?

MR WOODROW: It was the lowest cost at the level of functionality and performance that we wanted. Basically, the result of our evaluation was that DMF, FileServ, and Nastor were all very, very close in terms of performance, ease of use, functionality, and that DMF was behind primarily on a performance basis. I'm sorry, UniTree was behind primarily on a performance basis.

Question?

MR JIMMY BERRY (DoD): (Off microphone.) How much did it cost the government?

MR WOODROW: I'm sorry. Could you repeat that?

MR JIMMY BERRY (DoD): You indicated that your own internal system was the lowest cost. What value did you assign to the government resources that were used to produce these?

MR WOODROW: We assigned a cost of $0 to NAStore. This clearly does not take into account any of the development costs that have gone into it. However, given that we are faced with a choice of several alternatives, all of them cost real dollars for us to acquire, except NAStore. These costs are not trivial, especially when dealing with a tape inventory of significant size. Most of the commercial packages are priced based on size of the inventory or on the number of robotic tape units. For an installation like ours where we have eight 3480 silos, the cost of a commercial license is large.

MR JIMMY BERRY: How do you account, then, for the subsequent releases in the operating system, changes in the environment? I mean, for example, there's some of the other people that are running on like a release 1.5, which is about two releases back on even the commercial products.

MR WOODROW: We recognize that whether we run a commercial or home-developed HSM, we need a staff who understand the product in detail. In fact, we require that the local staff can build the product from source code on site. With this in mind we believe that OS upgrades for a home developed HSM can be accommodated locally without significant additional cost.

For the four packages in the HSM Evaluation, we looked at startup and recurring costs. We estimated that we would require a local staff of 2 for a commercial package and 3 for NAStore, to find and repair problems (yes we do this for commercial packages too) and add features as required..

Based on a one-person difference between a commercial HSM and NAStore, significant start-up costs and the fact the NAStore was very strong from a performance standpoint compared with the other alternatives, we chose to go into production with it. This decision makes sense today. When we started development of NAStore in the mid 80s, there were no UNIX-based commercial alternatives. The Storage decisions we make in the future will again be a cost/performance tradeoff and will likely be tipped in favor of a commercial package.

DR RANADE: Are you happy with that answer?

MR BERRY: (Off microphone.) Well ... (*laughter*)

DR RANADE: I'm not either. I mean I'm not --

MR WOODROW: You're not.

DR RANADE: Well, let me rephrase it. I'm not unhappy with it, but what I'm thinking, isn't this the case everywhere? I mean, wouldn't this be the justification in any place where they have a home-grown mass storage system? For example, does the Census -- go ahead.

MR WOODROW: We recognize that continued development on an in-house package makes less sense in light of current commercial alternatives. We do not intend to continue development of NAStore. It is useful as is and can be sustained at a competitive cost. At this time, factoring in cost, performance, and features, the balance is in NAStore's favor. As time goes on, the commercial alternatives should improve, and the balance will tip in their favor. We welcome this and will continue to evaluate our situation in light of the market.

DR RANADE: Anyone else on the panel? No? Okay. Let's move to number four. We have now developed the requirements, we've done the benchmarks, we've evaluated and now we're up to installation. Were there any special events or something you wanted to communicate to

potential buyers about the installation phase, something that you learned and which you wouldn't know otherwise about any of the software packages?

MR MARSALA: Well, at SRC we sort of learned remembering back that our primary function is supporting our research users. While we go a lot of input about the functional things that they wanted to do, when we implemented it, we implemented it about as user unfriendly as we could have, and, of course, the users didn't use it, which brought to our attention that it wasn't being used. After a little checking, we found out that maybe if we did a little more homework, we'd have it right. We now have our archive mounted as normal user UNIX file systems, and users don't seem to have any problems anymore.

VOICE: (Off microphone.)

MR MARSALA: FTP, telnet, and, of course, they hated it. I mean, it sort of makes the assumption that you have a knowledgeable UNIX user with lots of time, and both of those assumptions are bad.

VOICE: (Off microphone.)

MR MARSALA: Right. It's now NFS-mounted.

MS KELLY: When the system went into production, I had a 3-month hard deadline for decommissioning the old file system, which had about a terabyte of data. That was by far the most painful experience, migrating the old data to the new system, while we were still learning how to operate it. Of course, we didn't quite have our administration guide and all our procedures down pat on day one. So the conflict between getting the data off the old system at the same time we were trying to learn how to run the new system was a very painful experience.

I don't know if I should elaborate too much, but we didn't spend enough effort on the scripts for transferring the data. And yes, we chose to transfer the data rather than a cut-over date where the old system went away and the new system came on-line. We didn't spend enough time on recovery on the scripts. We didn't spend enough time on statistics to tell how we were doing. Operations had to dedicate one person 24 hours a day for those 3 months, and during that time an analyst worked 7 days a week, just making sure that everything was running all right.

DR RANADE: Mike, do you want to say something?

DR DAILY: I guess the only lessons learned were the typical things that you learned when you've got a complicated system: a fair amount of finger pointing, problems with software revs with mismatches, FTP daemons misbehaving and all that sort of stuff. I think if we had to do it over again, we would have tasked E-Systems a little bit harder to be the total system integrator rather than maybe doing a few end runs around them, or we would go chat with Convex about something. Pinpointing accountability and this sort of stuff is important, especially if you're not trying to be in this business.

MR WOODROW: Two points: 1) make sure data gets out to tape daily (don't allow a backlog to develop); 2) do regular backups of the file systems. Both of these are things that seem obvious, but can pass you by a little at a time..

DR RANADE: Okay. Well, both sides learn lessons, I'm sure. Question?

VOICE: When you're dealing with a terabyte of data, how long does it take to back up a system like that, or multiple terabytes of data? It strikes me as a significant problem.

DR DAILY: In our case, a substantial amount of what is on the system is data that's been transcribed in from external sources, and we transcribe in duplicate and pull the duplicate cassettes. As we start working more with intermediate data sets that get shed out of the supercomputers, that problem is going to become much more severe. I agree.

MR WOODROW: We use a primary and backup tape for all user files. User file system backups only save metadata (the node information) to tape and are quite fast. We also do regular backups of system file systems directories, but these are quite small and the backups are similar to most UNIX systems.

MS KELLY: We only backup the metadata, also. We only make one copy of the actual user data.

MR GARON: We don't back up the data. Most of our metadata is in Sybase data bases, and we just back that up as regular Sybase backups.

MR MARSALA: We just do a rotational kind of thing. It takes us about a week before we finish backing up our optical jukebox.

MS SALMON: Well, we back up the data bases that control where things are on tape, *et cetera*, but we've made it very clear to our users that we can only afford to keep one copy and can't make backups of the user data.

DR RANADE: How about the lessons from the other side of the fence, the vendors? I'm sure they learned lots of lessons in installing big systems and small systems. Would somebody from the vendor community like to say something? Dale, would you like to say something?

MR LANCASTER (Convex): (Off microphone.)

I was just saying that I don't know if it's a lesson learned, but it's just that you want to have the customer expectations well defined so that there's no mismatch in what you're trying to do. Also, try to bring these systems up slowly, rather than try to turn them on overnight. I think that's probably one lesson that I have seen out of many installations that we've done.

DR RANADE: Yes.

MR BENDER (Convex): I'm Ed Bender of Convex, and one of the things that I've seen is that data management customers are a hell of a lot more maintenance for us, a lot more work than typical computer servers. So that's one thing we've learned. We've had to put a lot of people into keeping things working.

DR RANADE: Can you tell us why that is, I mean elaborate a bit?

VOICE: (Off microphone.)

DR RANADE: A lot of different technologies are coming together in one system, and, therefore, you have these things.

MR LANCASTER: (Off microphone.)

I guess to summarize what your question is: why is there so much work, it's that we're really a system integrator now, rather than just a computer vendor, and that's really a big step.

DR RANADE: How about -- Dave, would you like to say something from the lower end of the market? I mean, you don't have as big a system as Convex does, for example, but --

MR THERRIEN (Epoch): (Off microphone.)

DR RANADE: Well, your lessons learned from installations with your customer base. I mean, yours is more or less a shrink-wrapped thing, isn't it? I mean, it goes in and --

MR THERRIEN (Epoch): Right. EPOCH was -- I think when you go from being a turnkey supplier to being a software supplier and expecting the hardware to come from somewhere else,

the problems are magnified even more so. Because now you're dealing with product revisions that are sitting in some dusty distributor's site that don't really match your minimum requirements, and you've got to kind of manage all that.

Those are some new problems that we're facing as we're moving toward being a software-only supplier: hardware incompatibilities. So we have to maintain quite a bit of information on which revisions of which storage products and which platforms actually do work with our software on a revision-by-revision basis. It's a big problem, but it's not impossible.

I guess what it produces is a limit to how many products you can support. If we go back to some of the talks today, you just don't want to support everything out there. What you want are a collection of things that you know work from release to release. So you've got to limit what you support.

DR RANADE: And you guys do very thorough testing before you actually support it in your product.

MR THERRIEN: Sure. Sure. Right. You have to do that. If you don't, you spend all your days on the phone in customer support problems.

DR RANADE: Anybody from a systems integration company? Do you want to say something on this?

VOICE: The prototyping seems to be very important, especially when you're working with new hardware. Also, simulation seems to be a good tool. We use quite a bit of that, but the real key is when you're experimenting with new types of hardware, HiPPI switches, if that's the case, MaxStrat disk arrays, or even at the lower end, the newer disk drives, that prototyping is pretty critical to understanding how user requirements relate to system sizing.

DR RANADE: How about somebody in that segment of the audience? That's a pretty quiet segment over there. No? All right. Well, let's move on to the next one, performance, which is a big issue for many people. Whose turn is it? Joe, is it your turn to start? We are on question six.

MR MARSALA: Well, what I'd like to say about the Epoch 1 is the performance met our expectations.

DR RANADE: Okay.

MS SALMON: For us, performance is a continuing concern. I think, in general, we've gotten some strong performance out of all parts of our system. We're handling 50 gigabytes of new data a day and up to 70 or more in and out of the system. So clearly, it's not that any of these pieces is a fly-by- night kind of thing. But our users' performance requirements continue to grow, so the level of the fence that you have to jump over keeps getting raised, as well. It's something that we have to continually work in concert with the vendors to try to solve, and the users.

DR RANADE: Mike?

DR DAILY: I'd say for most operations we're within a factor of two of the nominal numbers for these things, which is pretty good for being only a year or so out of the gates. There's still plenty of room to improve, and I think in many cases we're still technology-limited. Things like the CM5 are sufficiently fast that we're going to have trouble feeding it no matter what we use.

DR RANADE: What about this problem of small files and the D2 tape drive? Is that something you've experienced?

DR DAILY: No, we tend to have different classes for the large data files that get stuffed into the Connection Machine and smaller files that sit off on other classes that serve as workstations. And we've been experimenting with some of the things that the Sequoia folks have thought up, like abstracting, and our own crude forms of clustering of data to kind of intuit what the user is going to do next to improve the perceived performance there.

MR WOODROW: One of the surprises in the HSM Evaluation was that although the same underlying storage media was used there was great variation in the disk and tape performance. Apparently simple things like keeping a slow tape device streaming were accomplished by only two of the four packages.

Another surprise was the extent to which the disk performance differed between UniTree and the other candidates.

There is a lot of variation between commercial HSMs in the types of performance optimizations built in to the package. There appears to be a lot of room for improvement for some of the packages and extensive benchmarking appears to be a very wise investment.

I spent a lot of time on performance in the evaluation report and you can see the specific differences for yourself in the proceedings.

DR RANADE: Sue?

MS KELLY: Well, I've already given my two cents' worth on performance. The UniTree system satisfies our performance needs. But I guess to give four cents' instead, when we had originally done the requirement study, we had selected the protocols of NFS and FTP. They were given. And we began an early campaign of recommending NFS for directory management and for small files and using FTP for any large file transfers.

So when we think of performance, we tend to focus in on the FTP performance. UniTree is a poor NFS server. Our NFS transfer rates with UniTree are about 250 kilobytes per second, whereas we can get up to 6 megabytes per second with FTP. Did I say that right? Six megabytes per second with FTP; 250 kilobytes with NFS for reads and writes.

MR WOODROW: That's from disk.

MS KELLY: From disk. Well, yes, that's where they come from. For our tape activity, we have approximately four new gigabytes that are written a day. I said we have five gigabytes a day of I/O; four is writes and one gigabyte is reads. With the four gigabytes per day, our tape system has no trouble keeping up. Our migration is idle a good part of the day. So it's somewhere between four gigabtyes and five that there's a problem.

MR GARON: The system that's using the Metrum AMASS software, they're very happy with what they have. They just bought it and plugged it in and it sort of worked just the way they expected it be. They're storing about eight gigabytes a day. I talked to them and interviewed them, and they just can't imagine anything much better than what they're getting.

And there are improvements coming with AMASS software. Those improvements, I'm hoping, will help me solve some of the problems that I'm going to try to use another Metrum system for coming this fall. I'm going to try to store 25 gigabytes a day and see what comes of it, see how well it does in that environment. Call me up in 6 months and I'll tell you.

DR RANADE: Well, since we have about 10 minutes left, let's skip number seven and go on to number eight. This is: what are your thoughts on cooperating servers, different mass storage systems being able to talk to each other, so to speak? This will lead into our next topic, which is the IEEE Mass Storage Reference Model.

MR GARON: The only problems I have with the AMASS software is that it does have a proprietary format on a tape and the disk, but I think that's all done for performance issues. What eventually I'd like to be able to do is be able to move that media into other software management systems.

DR RANADE: Right. What most of these do -- I mean, all of them do.

MR GARON: Right. That's the problem.

DR RANADE: They just get locked into their universe of data formats and then it's impossible right now to move data between one system and another. So in whose interest is it to have that happen and are we likely to see it? Does anybody want to comment?

For example, if there's a UniTree system or some system and there's an Epoch system or some other system, is it useful to expect them to talk to each other? Does anybody have a need like that? Yes? Do you have a need?

VOICE: I have a question about proprietary formats. By definition, a format is proprietary if it is used by one company to store its data.

DR RANADE: Okay.

VOICE: (Off microphone.)

Is it still proprietary if that plan is public, even though it's only used by one vendor? If you have access to the formats so that you could translate the data if you need to, then is it still proprietary?

DR RANADE: Well, I don't know what the definition of proprietary is, but I see what you're saying. If the format is public, then anybody who wants to can write in that format. But what I'm asking is: is there a need for this to happen? I mean, are there installations where they have two different types of storage systems and they have a need for one of them to talk to another one?

I would think that there would be such a need, but I don't know if any -- yes, Jim?

VOICE: (Off microphone.)

DR RANADE: Any comments from NASA/Langley?

MR BERRY: Not NASA-Langley, but I can give you a different comment. We went through an evaluation on doing backup and recovery for a bunch of file servers. In the paper by Mike Shields of the National Security Agency which appears in this volume, you could see there were a lot of systems back in there. One of the primary criteria for the selection was that the tapes be readable through the standard Unix utilities, which means we could take a tape that was made through the backup system, move it somewhere else, restore it, and put it back up. From the system administrators' standpoint, that was very significant for their selection criteria. There were relatively few systems that did that, but that was one of the reasons why Bud Tool was selected, for example, because it produces that type of format that you can then use through a standard utility.

DR RANADE: Right.

MR BERRY: So in that particular situation, that was a very important criteria, and it also let you exchange tapes between Bud Tool systems. So you -- or you can even -- well, the other thing we were concerned with was being able to read a tape if we didn't have Bud Tool installed on a given server so that we can move files around.

So there is a very specific situation where that's true. It's also -- in some of the situations, one of the reasons why we don't have some of the systems on our supercomputers was the ability to share those files and not wanting to be locked up inside somebody's format, so that multiples of those systems can read the same data.

And actually, as we go to a more scattered processing, that becomes even more important. We don't want to funnel it through one thing.

DR RANADE: Right.

MR BERRY: So in both of those cases where we've got production processing, we think that's an issue. And backup and recovery, I think it's an issue that's turned out to be fairly important.

DR RANADE: Backup and recovery is a big issue. So are these open systems under Unix-based HSMs -- but how many of them are really open systems? I think to my knowledge there's only one HSM that writes migrated data in a standard format.

MR SARANDREA: What format? (No reply) Which is?

DR RANADE: Which is NetStore. They write standard format optical disks when they send data off the magnetic disk.

Yes? Go ahead.

MR SARANDREA: With reference to NetStore, just to comment. You said they write open format optical disks, but what they're really writing is the UFS file, magnetic disk file system, of the system they're on, which is far from standard. UFS file system on media format changes from vendor to vendor, so that's not an open standard.

DR RANADE: Okay.

VOICE: Our FileServ product --

DR RANADE: Writes tapes.

DR DAILY: It writes tapes and it writes standard ANSI tapes.

DR RANADE: Okay.

DR DAILY: So any utility that can read an ANSI tape can read our formatted tapes. Also, there's work with POSC to standardize an interchange format for tapes, so that it's not just the format that FileServ might use on D2, but it would be a standard that anybody that wishes to adhere to could use.

DR RANADE: Okay. Moving on to number nine, we have 5 minutes left, 6 minutes left, I think. I've purposely left that one vague. It says: IEEE MSS RM - practical import. So I think when we discuss that question in the panel, what we mean is: if the IEEE Mass Storage Reference Model has been an ongoing activity for a long time, and who knows how much longer it will go on. So what is the practical relevance of it to buying a system today? I mean, if it were ready and done, would it affect the way you buy something today or would it not?

I'd like to hear from the panel and also from the audience on this, because almost every spec from the government that one sees, it says the system shall be IEEE MSS RM-compliant or something to that effect.

DR DAILY: Well, we're big fans of standards, and we're willing to pay a certain performance penalty for it. But I don't think it would be a make or break in anything we're doing. This area is still awfully immature and there are other bigger fish to fry right now. But longer term, yes.

DR RANADE: Brian, did you have something?

MR SARANDREA: Yes, Sanjay. Can you define mass storage reference model-compliant?

DR RANADE: No, I can't. That's why the question is there. Why do you think it's on the list of things to talk about?

MR WOODROW: Yes, that's why I think the problem everybody puts in their spec, but how do you determine whether when the vendor says yes, this is compliant, what is it? Certainly, this is what we look for, one of the things that we look for, but it's not one of the things that we've been terribly rigid about enforcing.

DR RANADE: Well, yes. I think the goal of it is great and we want that, but how can the user community move towards it? I mean, is there a way for the users to accelerate that? I don't know. Sue?

MS KELLY: We used the IEEE MSS Reference Model during our requirements study. We had first done our requirements in more traditional areas of functionality, performance, and capacity. We then turned it around and looked at the system, looking for requirements based on the components of the reference model. We were not able to identify any new requirements by looking at it from the reference model viewpoint.

MR GARON: We would certainly ask the question, but I don't think it would have any impact on what we bought or didn't buy.

DR RANADE: It would or wouldn't? What did you say?

MR GARON: It would not impact what we bought.

DR RANADE: Okay.

MR GARON: I think it would satisfy the requirements, and it wasn't -- it satisfied what Mike Shields was talking about: solid company, they're going to be in business for a while and we can work with these people. Then we will continue to -- that would be a big plus, not necessarily the IEEE model.

DR RANADE: Joe?

MR MARSALA: I don't think I can add anything to what has already been said. I mean, it's just not defined enough yet.

DR RANADE: Ellen?

MS SALMON: Well, I can pretty much only speak for myself and not for the folks that went through the procurement. I think that the Reference Model is an important basis, but perhaps for us it was more important that the product we ended up with could run on multiple platforms from multiple vendors. So the product being "open" was probably more important than the Mass Storage Reference Model itself.

DR RANADE: Anyone in the audience?

MR BERRY: Yes, I can give one comment. Probably the most practical import that we've seen from our basis is early on and almost continuously they've emphasized the separation of control and data. And for at least the high-performance applications, I think we've validated that that is a concept that must be present if you're going to get performance. It's absolutely critical. You can't move this data across the networks with the control. You literally need to set up things. So when -- in Mike's charts you saw HiPPI switches, eventually fiber channel

kinds of things in which the data is going to move in a path that's not out contending with network traffic; it's running TCP.

So in that sense, I would say that's -- from our standpoint, we've seen that that's really a critical factor and is how you get high performance.

DR RANADE: Now that's a very specific application.

MR BERRY: It's a very specific thing in terms of model, but in terms of the whole model, no. There's lots of things in there that don't seem to be -- you know -- who knows?

DR RANADE: Yes, sir?

VOICE: How do you verify compliance?

DR RANADE: With something that doesn't exist?

VOICE: How do you verify compliance with things like compilers, POSIX, for instance? It seems to me what you're going to need to do is you're going to have to come up with a series of tests by some group affiliated with the people that come up with the standards or the models, and the products are simply going to have to be -- you're going to have to be able to run these tests to guarantee that all of those requirements are met when in operation.

DR RANADE: Right. It's a big job, isn't it, to say if something is compliant or not and actually prove it or certify something like that.

Dale?

VOICE: I think Mike --

DR ISAAC (MITRE): Just having some experience with the reference model, I felt obligated to stand up and say something about it. There's three or four comments I'll make. I'm not sure they're all connected.

Of practical importance, I'm not sure any reference model has any practical importance, and perhaps it shouldn't. Maybe the only practical importance a reference model would have is that one of the goals of the reference model establish a common vocabulary; this way, we can sit around here and talk about migration, and everybody knows that we mean something different than caching.

So just having a common vocabulary can be practical importance, but that's about as close as a reference model can get. Its goal is, especially if you read the fine print in the front, that this is not a document that one can establish compliance with.

The goal of the reference model, the second goal besides establishing a common vocabulary, is to establish a framework for the standards that are to follow, and that's where you should look for the compliance, the rigor, the benchmarking, and compliance testing. There are three or four dots that have been spun off the IEEE P1244 project. PVR will be the first one out of the gate. You can look to have active work on that towards a standard that will get you a physical volume repository, and the major vendors are actively involved in that.

It is yet to be seen whether or not such a standard is successful. It's quite a challenge to develop a standard rather than accepting the product.

DR RANADE: See, that's my point. Go ahead. Sorry.

DR ISAAC: That's about all I said. As for the other ones, storage systems management, the identifier, storage object identifier, and storage server, there is a dot spun up that has been

launched to establish standards in those areas. So maybe down the road in another few years, we can start looking at standardization that will actually get you interoperability and some of the other things that we'd like to see.

DR RANADE: Thanks.

Dale, do you want to say something?

MR LANCASTER I think -- I was going to make one of the points that David pointed out, that the reference model is really not the standard. It is, you might say, a fleshing out of the thinking behind the need for a standard. The standard is really called P1244 dot whatever and is currently being developed. How you do compliance is one of the goals of the National Storage System Foundation, which is having somebody that says yes, you really are compliant to the P1244 dot whatever standard.

I think mainly it benefits the vendors, rather than the users. I think the users have a secondary benefit, but the vendors, you know, we're pulling our hair out trying to have five different PVMs and PVLs and PVRs and all this other stuff that we have to integrate day to day with each of these systems. So it benefits us more than the users. The users just want a system, and I think I heard that a little bit earlier today, maybe from Mike, that you just want to store lots of data quickly and easily access that, whatever that means. And you're not going to hear, I don't think, a customer say "I think I need to buy another PVL today." That just won't happen, even though the PVL will be P1244 dot something complaint. So I don't think I -- I think there's no practical import to the user, but there's a lot of practical import to the vendor, which in the end will probably save money to the users buying the stuff.

DR RANADE: Sam, would you like to add something?

DR SAM COLEMAN: (Off microphone.) ...

In the software area, UniTree is an implementation of one of the earlier versions of the reference model, and it points out some of the strengths and weaknesses of the model. But the success of that product is demonstrating the importance of the reference model.

The National Storage Lab was a direct result, an outgrowth, of the IEEE effort, and that's a collaboration with 27 companies at this point working on new architectures that were suggested by the reference model.

There's a new project in the National Storage Laboratory which is specifically chartered to be an implementation of Version 5 of the Reference Model, and that system is going to provide performance of a couple of orders of magnitude greater than what can be achieved today. That project will become one of the projects of the National Storage System Foundation that John Simonds described yesterday, and the software division of the National Storage Industry Consortium is a direct result of the work in the IEEE.

I think the most important value is that the vendors have deemed this to be sufficiently important that several dozen companies are willing to send people to meetings every two months, and we have forty to fifty people that come together to talk about the best ways to design a storage system. The IEEE provides the forum and the reference model is the basis for those discussions. And that's very important, because we brought together a lot of traditional competitors in this area. We have all of the major software developers that are working on these systems. We have IBM and DEC and HP talking about how to build storage systems. We have Ampex and Storage Tek talking to each other. We even have Convex and Cray in one room having friendly discussions on how to build storage systems.

I think that the real importance is that this storage problem has gotten to be so big that no one vendor, not even Convex and not even Cray, is going to be able to solve this problem when we have large networks of heterogeneous, massively parallel systems, and we're talking about

terabytes a day and many petabytes of storage. This is an enormous problem, and the only way we can solve it is by collaborating and cooperating. And we see good cooperation among the vendors, and I think with that kind of effort being applied to the problem, that we will be able to solve it. So I think that's the main importance of the model.

DR RANADE: Any more on the model? Okay. Let's go on to the last one, metadata.

DR HOWELL (ICI Imagedata): Sanjay?

DR RANADE: Somebody on the model? Okay.

DR HOWELL: This makes me a little horrified, hearing that the standard is actually just a vocabulary. I would agree with the previous speakers that standards, in my book, are an agreed solution to a common problem. If it's a vocabulary, let's not have it masquerading as a standard.

DR ISAAC: Should I respond to that?

DR RANADE: Yes, absolutely.

DR ISAAC: (Off microphone.)

So you'll see IEEE documents that say guidelines four, blah, blah, blah, and standard four, and this is a Reference Model four. So it's not -- there will be a standards to come, and that's what you'll get, lots more than vocabulary. But the reference model has -- besides the Reference Model activity, I think Sam pointed out well that half of the importance of the Reference Model is the Reference Model activity in the working group. Establishing common vocabularies and establishing the major components as a framework for the follow-on standards is the most important activity of the Reference Model itself.

DR RANADE: Any final thoughts on the model before we leave it for another year? All right.

On metadata, anybody on the panel want to start? We talked about it yesterday. But let me just explain what we mean by that. Metadata, we mean data about data. You have lots of files, lot of information, but how do you access it? Must you use the file name every time? Or is there a way to intelligently index what you've got stored? I think we have somebody who has actually done a pilot system. Do you have a DBMS that --

MR GARON: The only data that we store in the one main system we work with is all -- there's a Sybase data base and it points to every entity of data. The analysts never pull by file name. Well, they don't know what the file names are. They query the data base, and they query in certain columns and get their information; that gives them their file name. We have built a level of software above that does the queries for them if they know what they're looking for. It goes out and retrieves the data for them.

DR RANADE: So they ask for certain types of data and the files come to them.

MR GARON: It could be.

DR RANADE: Right.

MR GARON: By various reasons, dates, whatever. I can't tell you the rest of it.

DR RANADE: It could be content-dependent, also, like what type of data is it; you could say for a simple example--cloud cover. You know, if you want data with X percent cloud cover, you could pull those, for example.

I would think that, Ellen, in your system, where you have 60 gigs going in and out for a day, something like that would be useful, right?

MS SALMON: Well, I think one of the problems with implementing that system-wide for our facility is the wide diversity of users and the reasons that they use the data. I think the division is looking towards at least providing the tools for users to organize their data in that way, and at some point it may be the logical step for us to step up to the management of that. But that's almost going to have to be something that the user labs explicitly come to us for and say we need this, and by the way, here are the funds from headquarters to go purchase the software and things.

DR RANADE: Well, there are actually two efforts that I'm aware of that are going on to define metadata standards. One is the one in Austin, and Bernie -- is Bernie O'Lear here? He left? He just left, okay.

MR LANCASTER: (Off microphone.)

DR RANADE: Could you tell us about it, about both of them?

MR LANCASTER: There are two efforts that are actually combining. I just found out this afternoon, because I talked to Paul Singley from Oakridge, who was on that committee with Bernie O'Lear. Basically, the IEEE, the same group that Sam and I and all are involved in, especially the one that was responsible for doing the Mass Storage Reference Model, started a series of workshops to deal with intelligent access to large amounts of data. Now, I'm not sure exactly what the titles were, but that's what I call it. Or what we call simply the metadata problem, which is: you've got ten million files-- how do you find what you're looking for?

Even people at NASA retire eventually, but their data doesn't. And you wonder: well, do I need to delete this file or keep it? And you don't know, because you didn't generate it originally.

But anyway, we had a workshop in Austin that Jim Almond and I set up down at his center, and we had several people come who were highly motivated to try to get a handle on this. We have some minutes from that workshop that have been generated, and a white paper is being written by a couple of people. Robyn Sumpter and I think even Sanjay is working on that as well -- to try to define what the problem is and where we might want to go with that.

Parallel with that, there was supposed to be a workshop at Oakridge sometime in '94 to deal with something that they thought would be the data base-type problem. Well, they had their first meeting to set up the workshop, and they realized that they were really more interested in metadata; that's what they really want to talk about.

So Paul Singley and I got together just a while ago -- and I don't know if he's in here or not. There he is -- to say: well, gosh, we're skinning the same cat; let's go skin it together rather than try to reinvent the wheel.

Then I saw some papers on the Information Interchange Reference Model, again maybe defining some vocabulary; but the idea that -- it's a big problem. In fact, I think that's public enemy number one, because I think that anybody can store lots of data, but not anybody can effectively use it. And I think this is a step to get there.

So that's my 25 cents' worth, Sanjay.

DR RANADE: Thanks.

VOICE: (Off microphone.)

DR RANADE: Well, if there's no more, I'd like to thank each of the panelists for being here with us and sharing your experience, and the audience for being here and listening to us and participating. Thank you.

EVENING RECEPTION AND DINNER

Moving Images Archive

David Parker
Acting Head
Curatorial Section
Division of Motion Pictures, Broadcasting and Recorded Sound
Library of Congress
Washington, DC 20540

MR. PARKER: -- for lower check, for the way things were. I'll try not to make this autobiographical and dull; I'll try to make it official and dull instead. But I got something in the mail Saturday. It was one that didn't say "occupant" or "resident." It said something to the effect that if you get to the Library of Congress by 3:00 a.m. on Wednesday morning and stand in line or bring a cot, as you would for a Rolling Stones concert tickets, you were eligible to retire. It puts me in a retrospective mood tonight.

Well, came Wednesday and a lot of people were in line, including our assistant chief. He's been there for 30-some years. So it was retirement, retirement all day long. People who hardly knew each other, who were barely colleagues at the Library, would pass in the halls, and one of them said, "I don't want to hear another word about retirement."

At the end of the day, one of the last researchers came in, somebody I knew, and he didn't know anything about this. He hadn't been reading the paper. So he saw the assistant chief's secretary putting on her coat, and he said, "Oh, are you leaving early?", as one would ask, "how's the weather?" She said, "I'm not retiring." And neither am I.

But it does take me back to 1969 when I came to the Library of Congress. I was a film maker, and somehow they convinced me it was more important to save the original negative of ***Citizen Kane*** than whatever I might turn out next year.

I was also there in the early '70s when they changed the name of our division. It's a little bit of immortality for me, because the word "broadcasting" in the name, "The Division of Motion Pictures, Broadcasting and Recorded Sound was my suggestion. And about 15 minutes after it was officially adopted, it became obsolete for reasons that may have to do with what you were talking about during this conference.

I hold here a printout. This is my security blanket. I'm a bureaucrat, I bring this. This is ultimate truth. This is a count of everything we had as of last October. If you want a count of everything we hold in our division as of *last* October, that's why I may look a little more frayed than usual today. We're still working on it. We're going to have it ready tomorrow. It's four days overdue right now.

I guess an important milestone would be in 1964, when we got a film scholar as head of the film division, not a retired military person, which had been the tradition up till then. I mean, pledge of allegiance to the flag first thing every morning.

The film scholar decided it would be good to retain more films than fewer. So the idea of selecting only the very best of the best, chosen by whatever the standards of that year, reverted to what Archibald MacLeish, a Librarian of Congress in 1943, had envisioned at the establishment of the film area: instead of sending films in for copyright , having a clerk note some information from the film and sending them back to be dispersed and perhaps never to be collected again, the Library of Congress, as the national library, should select every year for the permanent collection films that tell us about living in that year.

And Archibald MacLeish didn't just want the best of the best of that year; in addition to films of great news events, he also wanted newsreels about whatever would be the 1943 equivalent of the hula hoop, and he spells that out, the range of production, I guess, as if it were to go into a time capsule. And curiously enough, the University of South Carolina, which now has the collection of films of the Movietone News, most of the requests are not for the hard-core news features; they're for the other parts of the newsreel: the dog who ice skates, the guy with the wooden garden, and the hula hoop. Because a newsreel of 1943 was made up of all sorts of things, and that's the mix he wanted.

We were able to select a lot of films because of the U.S. copyright law, one of the best in the world, if we wanted a film for the permanent collection, it must be surrendered. One copy instead of two. For books, two copies are required, but the Motion Picture Association and The Library of Congress made a deal, not the last one.

In the late '70s we had a shotgun marriage, and all media was put in one area. It's sort of the concept that I understand was used by the University of Maryland library. You could dial the media number -- probably still can -- and they answer the phone, "non-book." So I guess I'm in the Library's "uncola" division.

Well, that's the way it is. That's the library. the books and the media, these Johnny-come-latelies. Perhaps the reason film has become thought of as an art is that there is now television to trash, you know, because it's newer.

So we're now the Division of Motion Pictures, Broadcasting and Recorded Sound. (Presumably that's sound isn't wandering around, bouncing off the walls but is actually engraved on a support base.) In the division, they came up with the Curatorial Section. They already had the standard library administration, acquisition, processing and cataloging and added something called "curatorial". (That's not "custodial", but some days I can't tell the difference.)

I'm up to '92. The official count: In our curatorial division alone -- omitting the books and the electronic media, (machine-readable documents and CD ROM) -- only counting moving image and audio -- we hold 3,328,589 items, which take up linear shelf feet of 263,875 feet and 7/10s of a foot.

I can't give you the cubic feet they fill; because we have many formats, from miniature home movie formats to 70 mm copies of films such as "Lawrence of Arabia," each reel of which is counted as one item -- and don't drop a reel of 70 mm on your foot! In fact, if you've been with the projectionists' union so long that you have the seniority to be projectionist at the house where they show 70 mm, you might get a hernia. They ought to assign 70 mm work to projectionists in reverse order of seniority.

So we have several hundred pictures that are in 70 mm format, including reels that came from Elizabeth Taylor Warren's residence of a motion picture called ***Around the World in 80 Days***. A film studio has accessed that material to put together a new 70 mm version of that film, in the same restoration procedure done with ***Lawrence of Arabia*** and ***Spartacus***.

That kind of holdings help make us an archives, not just a library. It's not getting just having video copies for home viewing; it's also having the original camera negative of ***Casablanca***.

The Library and copyright started about 100 years ago copyrighting films. We have now some film copies manufactured made 100 years ago that are still in good shape. The others are not and I want to get to that right away, because that's the part that worries us in the janitorial part of the Library here.

We also collect 45rpm records, although there's a guy who says he has more 45 records than the Library of Congress. His trick is that he bought up all stock from the regional exchanges as they went out of business, because the computer let the record companies ship nationally from a single location -- Terre Haute, I think -- so he may have many copies of the same 45. But it's true, he has more copies of 45s than the Library of Congress. We're talking to him.

Now, when a format becomes obsolete, we don't throw it away. We give it to the Library of Congress. For instance, recorded sound on cylinders. We've got 10,500 of them in last year from one collector. So when people clear out their attics and basements and they find something very valuable, the Library of Congress has to have something to play it back on, into whatever new wonderful equipment is now the technology for the next decade or so.

Maybe the name I should have come up with in 1970 was "the Division of Motion Pictures, Broadcasting, Recorded Sound and Laser-Etched Saran Wrap, and whatever they invent next"..."non-book". the book side seems a bit more stable.

I've seen some things along the way that even with my poor eyesight I knew weren't going to pass muster. Somebody was explaining to me -- I think it was a film manufacturer-- the advantages of something new called super eight (How many remember super eight?)

He was explaining the advantages of super eight over standard eight. Does anybody remember standard eight?) He said you couldn't recognize your own grandmother on standard eight, but with super eight, you could.

So somewhere between ***Lawrence of Arabia*** or ***Far and Away*** or some showing in IMAX format and the poorest half-inch videotapes we've ever been offered, we have to decide what is appropriate or acceptable quality of preservation for the moving image. Does the film still survive when you can barely make it out as if it's transmitted by wirephoto? Or do we require a 70 mm original copy?

You can look at a movie called ***Love Story*** and hardly make out the figures of the actors, and it can still make you cry. But if you're looking at an Anthony Mann western, the landscape is very important to what the filmmaker is trying to communicate. Some film makers use such strong geometric forms in their pictorial compositions that you could send it by thermofax and the idea would get across. But the more detailed the physical surface, the more the sensuous parts of the medium are used to tell the story, in contrast to diagrammatic plots and cliche'd dialogue, the more important it is to retain the resolution, the technical quality, of the original, at least in one format so it could then be translated into the other forms in which it's going to be distributed and viewed.

So ideally the problem is getting a print from the original negative of ***Casablanca*** over the fiber optics network to Los Ceritos, California, where it can be picked up in the viewer's own home, and still look and sound like ***Casablanca***. There are perhaps one hundred shades from the whitest to the blackest black in ***Citizen Kane***. To reduce that to 20 shades of gray gives you the equivalent of a smudged carbon copy or something even worse. Let me take an example from music: listening to Mahler's ***Symphony of a Thousand*** over a 50 cent, two-inch loud speaker (like those used in cars at the drive-in movie) may work fine if you've already heard Mahler's ***Symphony of a Thousand*** in a concert hall or on a fine CD. You can bring your earlier experience to what's actually there from the 50 cent speaker from the drive-in. But if the drive-in quality of sound is your *first* experience with the work, your filling-in of what's actually not there may be relatively unsuccessful.

The Library of Congress has to worry about such considerations when we talk about compression and sampling rates, when we talk about translating it into any other formats. But mostly we worry about the condition of the physical material on which the content is recorded. Digitally we can now recopy every five years and theoretically lose virtually nothing. But if we've got 700,000 safety films, all of which may be attacked in the present or future by the vinegar syndrome, that's 770,000 cans that have to be opened by somebody has to

do something physical to each can, even if it's just to stick the rubber nipple in the first time so that a probe can be used with the nipple every subsequent time to record information about gases in the can and not have to open the can itself again.

We have 110,000 cans of nitrate film. When nitrate film is ignited by a spark or an open flame -- it doesn't explode; it just burns so fast, even under water, that you can't tell the difference, -- With nitrate film, we try to open each can for inspection once every six months. But the irony now is with the vinegar syndrome problem, we have movies made on safety film in the '50s and '60s, the original negatives of which are showing -- not on a large scale yet, but on a small scale -- throughout that entire 20-year period, deterioration characteristics quite similar to those of nitrate made from the late 1890's to 1952.

We've found there are not that many differences between the new safety films of 1915 or '52 and the nitrate, if we're talking about long-term keeping and storing and their total lifetime. Let's move closer to the present time.

Remember you couldn't recognize your grandmother in the straight eight? Now let's go to something I saw a couple years ago that made me very excited and made we want to be part of this group here to learn what I could learn. All this time we've been hearing something just as good as 35 mm and then we've been seeing, and it does not meet large auditorium, large screen showings. It may work in some other kind of presentation environments.

I've seen Kodak's new system, where you convert a 35 mm mm image -- not 70 mm, not IMAX, but 35 mm to a digital record, manipulate it in that form and etch it back onto back 35 mm film. , at least on a reasonably-sized screen, some pretty remarkable digitization. First the Kodak tests and then Cinesite, the company that restored ***Snow White and the Seven Dwarfs***.

Now we're back to an area in which I'm some kind of an expert, having memorized "Snow White" over many viewings. I've been seeing that film -- I won't tell you how many times and for how long. Every time it was reissued, I saw it, and I have some clips of a print at home which I could compare what the digital form was like with the original. If I were an art historian, I could quibble about this shade and that hue and that intensity and say that the blacks are too gray and the saturated reds are not there. But it's amazing what *is* there.

What is there is a pristine copy. If you saw it in its last reissue in 1987, produced with conventional printing techniques, in the scene where the prince first meets Snow White and she's singing to the doves -- well, in that 1987 print you could see the doves fly off to the left and the field of dust go over to the right, and both were about equally prominent visually. In the current reissue the dust has been now removed digitally, except for two specks they left on the surface of the magic mirror because, you know they made it look more like a mirror. Without the dust, it looked too transparent. That's referred to as the inability of a dog to pass a fire hydrant without stopping.

That's not fair to the people who have done a wonderful job, and they showed the "before and after" of the first reel to us at the Library of Congress,. They had an idea in mind that tied right in with something we'd been talking about ever since we knew in 1969 what NASA could do visually that was not possible for us. When large pieces of the original film emulsion with the original information fall off of that 1895 picture, leaving only the clear base. And if what's been lost is redundant information, if it's present in adjacent frames, and if you could capture it from those frames and put it back in the frame suffering the diverticulation, it could look as if it had been shot yesterday.

When Frank Capra, the director visited the Library of Congress in his later years I was privileged to set up a screening for him of one of his movies made in the mid-thirties; we had struck a print from the original negative. It was a test print, and I thought it was terrible due to shrinkage of the native -- the sharpness was not good. But Mr. Capra said he was impressed with what he saw because there were no visible scratches, and without the scratches it seemed

that the action was happening not in 1933, but right now as we were looking at it. The illusion of the movies was sustained. That could hold good for a sound recording, too, where the processing allows the original to come through with its own kind of sound.

We had wet gate to make the grain less visible. You couldn't see the grain. You could blow up 16 mm, Disney's ***True Life Adventures***, ***The African Lion*** or ***The Endless Summer***, **On Any Sunday**, and you didn't see an oppressive grain structure; that was removed. You didn't have a very sharp image there either, but you had the cues of color and shadows in your own mind to help separate planes of action and foreground and middle ground from background. As for what's not there, but it doesn't seem to matter because the psychological effort of the person who are reading the image or listening to the sound image compensates for it.

Maybe we may decide to do exactly what they did with Disney's ***African Lion*** shot in 16 mm. We can't just save everything in 70 mm, although the technology to do it is there. Here is one thing which the Library of Congress is somewhat slowed up a bit, and it's the same factor as in 1969, when we were talking about diverticulation and the patches and what NASA could do to restore visual information at that time:

With ***Snow White*** it goes something like this. I may not be quoting this directly, but this is what I remember hearing them say: For each frame manipulated, it takes thirty seconds and costs $8 to etch that amount of information into the digital format, and then when you're finished manipulating it, getting rid of the holes and patches and creases and all or maybe touching up the color a bit -- for instance, it's monochrome down to the bottom of the ocean, so you add red coloring to the coral so the audience doesn't see such a boring all-blue image. (That's being done for a new Tom Cruise picture. Look for the coral; it's digitally enhanced -- and then you get it back onto motion picture film so it can be projected in 35 in a regular-size theater, that's $6 more. Twenty-four frames a second, 90 feet a minute -- well, you get the idea. And that isn't paying for the 100 people who worked three shifts around the clock to get "Snow White" ready.

So the difference between the potential and possibilities and what resources the Library might have available for that seems to be a great chasm to bridge.

There is another demonstration I saw that cheered me up as much as seeing the digital ***Snow White***. This was a development by a professor and his graduate student, working with limited resources, using off-the-shelf materials at a university brought to the Library of Congress. It was a particular jolt for me because the man who had just been given the assignment at the Library of Congress to look into what might be technologically possible for such an application, was watching the demonstration and could see that this system was already up and running, and we were starting far behind.

Positioned on the West Coast you could view the cracks and gouges on the surface of a disc recording that we hold in the Library of Congress in Washington. It's a 78 rpm record. (How many people know about 78 rpm records?) Maybe they're in your attic, if you're not tidy and haven't done spring cleaning.

We have become reconciled at the Library that our 78 rpm records are going to get fully cataloged just about the time all those 45s also get cataloged by conventional means. It isn't going to happen soon. So let's talk about applying the low tech of 1975. We photographed each label, front and back, on each disk onto a frame of 16 mm film. It may not have the best resolution, but if they can use 4 mm for photographing recording instrument panels for test planes, we can use 16 mm film for photographing record labels.

Now, we haven't cleaned up the mistakes on the record, the typos. And beyond mere typos, you may not believe everything stated on the label: we have a Decca record that says, "Bill Haley and the Comets, `Rock around the Clock'. Foxtrot."

But catalogers can worry about what it is if it isn't a foxtrot later. What you can do now is punch a four digit number and retrieve by composer, by artist or by title every 78 rpm that we have in the Library of Congress and in four other U.S. sound archives, up until the time when the project was over, when they quit photographing labels onto those 16 mm frames. Accessing a huge data base is possible, thanks to a meat packer who'd made a lot of money, who liked opera, who wanted to find out what there might be in the way of opera on 78 rpm records. And he was convinced that everything on 78 rpm ought to be treated the way he wanted opera treated.

Now they're working on getting that data onto a CD-ROM so it can go out with all the other things the Library makes available on CD-ROM. We have videodisks of San Francisco, of New York, photographed at the turn of the century. These are the paper prints, contact sheets made for purposes of copyright. Until 1912, the only way moving pictures could be copyrighted was as still images. And between 1912 and 1943, when the Librarian of Congress said, "We ought to be keeping some of these films here in the national collection," that's the period we were trying to fill in by getting the original negatives from the major studies and making a master film copies on 35 mm to match the originals as closely as we could with the silver content of emulsions today, to retain the ability to recreate a large screen theatrical experience.

Yes, you can get a copy of ***Casablanca*** on a half-inch video copy, but it's not quite the same thing. The size, the dimension, a lot more is lost than one would know unless one saw it in reverse order, on the big screen first as I've been privileged to do, as we did for all of these films.

As you may have suspected by now, I am lost somewhere in the past, selecting films made before 1952 to be copied, because they're on a nitrate base and going to crumble into dust early. And now we're also concerned about the pictures made in the '50s and '60s because of the recently discovered threat of disintegration.

The one thing that it seems to me that all this boils down to that I've seen since '69 is the technology changes every decade or less. The Library of Congress has to keep all the information we might want to access that's recorded on the cylinders -- Brahms playing-- down to the present day. And the physical materials that these recordings are on is so fragile. If the consumer audiotape is projected to last 20 years plus however lucky we get -- and, of course, we don't control the materials chosen. We don't have the materials of our choice to work with. Often we have just what the collectors give us, .

In a play by Brendan Behan titled the ***Choir Fellow***, which takes place in a bar, The woman who runs the bar sees a man dandling a girl on his lap, and says, "Put that girl down! You don't know where she's been."

We don't know where the collections have been before they come to us, so it's harder to figure out what their additional life span may be now. We know about a man who owned the organ company and wanted to have something to look at while he played his magnificent theater organ. He got a wonderful collection of the great silent films. He lived in a castle by the sea. In it he had a vault near the seacoast in which he kept the films. By the time we learned about it, there was only one of his films that could be salvaged. It was ***Salome***. We hung it up around a room in the nitrate film building, and dried it out like wet wash. When it was dried out we were able to print it.

So we worry about compression, we worry about sampling rate. But mainly we worry about the tendency of all things laminated to delaminate, whether we have 20 years or 30 years and whether the accelerated aging tests of materials done at Kodak and elsewhere are accurately predictive. We do know tests for the longevity of films, done in the '50s at one of our sister archives, didn't prove to be accurate. So there must be other factors, such as "where it's been", that couldn't be taken into account.

We have somebody who believes in cryogenics, digs the film a hole, buries it in the hopes that the technology to bring it back will come from this group or others one of these days. That's a faith in science maybe, but beyond my powers of willing suspension of disbelief.

All of this audible and moving image material is the memory of the world or at least, the memory of the Continental United States and its territorial possessions, et cetera, et cetera, as of certain times. Of this memory of the world, we never know for certain what is going to be wanted next. Although we keep a great deal of it, we have to make "triage" decisions every now and then.

The fragility of the material, the lack of backup copies, that's the sort of thing that bothers me. But the excitement is what is possible even if the Library of Congress doesn't have the resources yet to play in that particular high-tech, high-expense ball game in that club, in that league.

The disk that you can not only hear played for you but can also look at its notches and cracks, as well as the label stating that it's a foxtrot called "Rock Around the Clock", from across the country, that's a little more exciting than just the offerings on pay TV, as easy as selecting something from your local video store. It's an example of what the Librarian of Congress may be talking about when he speaks of "getting the champagne out of the bottle", so the super digital highway is a wonderful dream of possibilities and we're all following that dream.

But the time and cost of getting from here to there is a problem, and I suppose I'm an arch conservative. I'll end with repeating what I heard at the East German film archive: (remember East Germany?) And if I suspect that it was chosen because they didn't have the high tech resources available to them, it still may be the right choice.

But what the head of the archive there said is something like this: "we've built good vaults with proper temperature and humidity controls to keep the film and tape alive for 100 years. And when you with the high technology figure out what are the optimum means of re-recording this material might be, the material will be here. We'll know where it is, we can find it and we'll make it available to you. It will have been saved."

So those are the two paths, to what I like to call "archivery". It's "thievery" and "sorcery". A bit of everything in it. I think it sounds better than "janitorial".
(*Applause*)

If there's someone I've not confused totally by what I'm saying or where I seem to be going, raise your hand. I'm open to questions.

VOICE: (Off microphone.)

MR. PARKER: What is the relationship between the Library of Congress and the National Archives? Do you mean from the firing on Fort Sumter or after that? I get this asked all the time, when I don't get asked about Kemp Niver, the guy who got the Academy Award for the paper prints being transferred to 16 mm film.

There are gray areas, which I'll not go into, but roughly it is that what the government produced, the documentation the government produces, like your Army record from 1915 or films about activities of the government -- that's how they sneak in newsreels with hula hoops into the collection -- material generated by the government goes there.

The private sector, largely, I guess, because of the books we buy and the fact that the copyright office gets materials to us, the private sector is represented in the Library of Congress. We're always getting mistaken. You know, it's like the actresses, Gale Sondergard and Judith Anderson: which one played the sinister housekeeper in ***The Cat and the Canary***

and which one played the sinister housekeeper in ***Rebecca***? After a while, the one that wasn't in ***Rebecca*** just wearily thanks the fan for the compliment and doesn't try to correct anybody.

VOICE: (Off microphone.)

MR. PARKER: Surrender the copyright to the government? Did I say that?

VOICE: No.

MR. PARKER: Sounds good. Surrender copies, two copies. Should there be a legal case, then this would be evidence. We've even sent out a videotape of a movie that was evidence in a copyright infringement case, and we put a ribbon around it and stated ion a note, "We verify this was a true copy of the movie."

VOICE: (Off microphone.)

MR. PARKER: Yes, we have two copies of one film, ***Johnny Guitar***, because one copy came from its star, Joan Crawford, and they didn't say no to her. And there is a problem of how much backup is desirable. If you have one copy and it gets torn up during the next screening, where are you?

VOICE: (Off microphone.)

MR. PARKER: Everything until about 1955 might have been shot in the three-color process. (***Foxfire*** was the last movie shot in the three-color process, that's ***Foxfire*** with Jane Russell, not ***Firefox*** with Clint Eastwood.)

VOICE: (Off microphone.)

MR PARKER: Yes and no. We are storing them for the archive they belong to, along with three vaults of other materials. Well, let me just explain about these three million items by way of Technicolor and Warner Bros.:. If you want to see ***Robin Hood***, it runs about two hours. If you want a print from the original negatives, that's forty reels. For every reel of picture you look at for ten minutes, you've got a cyan record, a magenta record, a yellow record, and the soundtrack.

So if you lose one of those -- and it happened to a reel of ***The High and The Mighty***, I'm told -- then you've got to try to reconstitute what should be there from the surviving elements, and that happens too, as was done with the restoration of ***Becky Sharp***.

Yes, we have -- we're storing MGM color pictures made during the nitrate era to my knowledge, but they're not ours. We're storing them temporarily for another archives.

VOICE: (Off microphone.)

MR. PARKER: You hold onto it as long as possible, because still and yet again, (***Snow White*** on digital notwithstanding), it is the best source material to copy from. Of course, if it *does* start ticking, it is put under water, because if it crumbles into pieces, it is much like gunpowder. If it becomes a safety hazard, it goes under water. But you try to save as much of the picture as possible. You *don't* say: "oh, this reel smells bad. I'll throw it away." You carefully cut out the deteriorating parts. It's a bit like running a cancer ward.

VOICE: (Off microphone.)

MR. PARKER: Well, we hope, you know, it will go by fiber optics to Los Ceritos, California, but we're a little way -- but with $8 a frame going out of film and $6 a frame going back to film, we're not quite there yet.

The other thing, of course, are the copyright owners. Oftentimes we have to send access seekers to the donor of the material, if it is on deposit at our place, to their lawyers to find out what rights are involved. Paranoia in the industry is classic and has not been mollified by the discovery people who have been active selling
video copies that are unauthorized. The copyright office is located one floor above us, so we're very circumspect about that sort of thing.

But we always have the success story of the guy who did everything we told him to do, instead of trying to find a way to beat the system to get access. The rights owners said yes, the publisher said yes, and he got what he wanted. It takes a little longer maybe than you wish and a little patience, but it works.-- I think the last line in my job description says: "get the stuff out so people can see it and hear it. So we've found new ways of doing that. We're having some touring shows next year for the centennial of the motion picture. Nearly every state will have a showing, over two years. The details are not worked out yet.

We're making the first batch of films available to the public. Early films by early film directors, women -- some important black cast films that are otherwise not being distributed. And those will be out in February for rental on 16 mm, 35 mm, and for sale by mail on half-inch videotape.

VOICE: (Off microphone.)

MR. PARKER: What I heard, they don't store it on digital video for "Snow White." It takes up too much space, it's impractical. If you're talking about full 35 mm resolution.

VOICE: (Off microphone.)

MR. PARKER: Well, yes. We're working with -- that's why I was interested in last year's transcripts. One thing I may have in common with this group is interest in the longevity of D1. We worry about the moisture content of the tape at the Library, too.

We make -- the analogy, I think, for our policy, quickly, would be when we make a transfer on audio, we make both an analog and a digital copy because we're trying to have something retrievable for 200 years, and because we have anecdotal evidence accumulating that's not cheering, such as not being able to read time codes and things like that. In fact, I guess our most extreme position would be the one we've taken with the Marlboro Music Festival. They've been sending us recordings of the festival for years. When they started sending us digital recordings, we said, in effect, "Thanks for the recordings. Now we want the machine they're recorded on, too," because we've got to be sure we'll be able to play them back." That may be an extreme position, but I guess that's the way our thinking goes.

VOICE: (Off microphone.) ...Movietone News

MR. PARKER: I didn't see it personally. I've talked to fellow archivists about it. I have a prejudiced, bigoted opinion of it without having enough information to even be worthy of having an opinion. However, would you like to hear my opinion?

So far, it has nothing to do with preservation. It has to do with access. The preservation part of it does not meet our criteria, to put it mildly. These films go back to c. 1919. There's yet to be any test of film in shrunken, curled or otherwise unsatisfactory physical condition being transferred. I don't know whose criteria it might meet. We'll find out as they work it out. It may mean that a lower level of preservation is acceptable for some applications. But if you've got gorgeous, breathtaking 35 mm images, to reduce them to that kind of quick, easy access only does part of the job, I think.

Although, that would be half the solution that I would see as ideal somewhere along the way. But I would say you start with retaining the information that is there in some kind of master copy and then make it available for prompt use that way. And my boss, who just

retired, was once called a bad name by a frustrated documentary film maker at the top of his voice because the Library, then working through an outside lab -- we didn't have our own in those days -- couldn't meet his deadline for television.

So that part of the problem, the Fox has got -- let me say something nice about the studios. You know, we're not -- I feel like Teddy Roosevelt: "Alone in Cuba" should be the name of my address here.

There are four archives that conserve this same kind of material in the United States, as well as the film companies. I saw something wonderful in last year's program about assets, preserving and protecting assets. That's a new idea, instead of nitrate just being this stuff that explodes on you and costs a lot of money. And one of the major companies that just built a beautiful restored vault for nitrate films, state-of-the-art facility, calls it "asset protection". Why didn't we think of calling it that in 1969?

VOICE: (Off microphone.)

MR. PARKER: It's here. I can work it out with you afterwards.

VOICE: (Off microphone.)

MR. PARKER: I could have gone on with several more formats, you know, after super 8mm and 78 records. By a reel, the industry, since the '30s, has considered a reel about 8 to 10 minutes of running time. When we get these reels, they may come off the airplane in 3,000' reels. Typically, with original 35 mm negatives, you don't store anything larger than 900 to 1,000 feet a reel.

So the average A budget picture in the '30s runs 10 to 12 reels. A Fred and Ginger musical may run 10 to 12 reels, although its running time may be only 90 minutes, because they don't want to cut right in the middle of one of the numbers of where the reel breaks go.

However, once when the Library of Congress had a total of three people working on motion pictures and the industry had changed over from 1,000 foot as its standard length to 2,000 foot, because everybody now had projectors with take-up reels with 2,000 feet capacity, there was one guy, I understand -- and I've seen some of the musicals, so I think it's true -- who had a machete. If he had a 2,000-foot reel that came in for copyright, about 1,000 feet in, he would whack it with the machete so it would fit in the 1,000 cans. He only had 1,000-foot cans. We couldn't buy 2,000 foot. And he didn't miss a musical number; it was sort of amazing--whacked right in the middle of each one. I don't know about the others, just the musicals I went through.

VOICE: (Off microphone.)

MR. PARKER: It's difficult having to operate many kinds of equipment at once, and we've had special programs transferring cylinders.

Let me tell you about the amateurs who recorded wax cylinders, because what's semi-soft wax and what's stamped celluloid, what's original wax recordings, is one of the more exciting stories we have.

Indian rituals that would be lost to the memory of the tribe today are sometimes documented and retrievable by amateurs who went out with their portable cylinder machines. Long before the folk song project of the '30s that the Library of Congress is noted for, when they took tons of recording equipment in a truck right out in the field and recorded folk songs on site.

We had a special project for transferring these disks in the late `70's, and I remember vividly when we became part of the recorded sound division in a shotgun marriage. We would have a meeting around the great green table in a recording studio. But the project couldn't stop.

In the same room they were transferring native American chants at the same time. Yes, we don't deal with all obsolete formats the same way, but yes, we try to cover the waterfront.

VOICE: (Off microphone.)

MR. PARKER: Do we buy hardware?

VOICE: (Off microphone.)

MR. PARKER: Yes. In the case of the Vitaphone system, the disk system that brought sound movies, to popularity -- they'd been around forever, like 3-D, but they weren't popular -- the Vitaphone system we now share is in a lab in Hollywood with one of the other archives.

You see, they -- this is a symbiotic relationship. They have the soundtracks for these movies and we have the movies without any soundtracks. And there is a third factor: ***Ali Baba and the 40 Thieves***, I left them out. These are the collectors, bless their hearts, without whom I'd be out of business, because a lot of these things are not available at the studios or from copyright deposits, if the movies we're talking about are from the silent era or the very first years of sound.

And there's a record collectors group now, a consortium, which negotiates with the Library of Congress, because their collectors have the soundtracks and we have the films. It's getting more interesting. If you want to know what the Ed Sullivan Show would have looked like in 1927, we're about to be able to show it to you. Because in the first years of sound, twenty-four hours a day in a studio in Brooklyn, they set up four cameras, and anybody in show business who was appearing in town came in and did their act. They didn't cut away to Alice Faye kissing Don Ameche or keep the plot going during the act. You get to see the act ungilded. So you get to see somebody who had done the same act for 30 years on the stage, and in their thirtieth year they're recorded picture and sound for the vitaphone in 1926. That's sort of a reach- back.

Not as amazing as seeing a pope who was born in 1830 on a motion picture, which we can do with the paper prints of films made before the turn of the century, but we're getting back there. We're finding out that we're not necessarily better in every way than anybody else who ever lived in this country. I guess we learned that from Ken Burns' television series on the Civil War. He found people who were sensitive and intelligent and admirable and their experiences could be moving to us from that kind of presentation, and we're finding the same sort of thing as we go back to these obsolete formats and bring them back to life.

Not all of the films and recordings are equally wonderful, but enough are so that the pride of discovery is still there and the delight in finding something that communicates to us today is there.

Well, we've got Mickey Mouses now. Disney has deposited at the Library important material, World War II and -- the people who acted out "Snow White" live before the cameras, the animators translated it into drawings and much more. Please do come to visit our division at the Library of Congress. And if you give me enough advance notice, I'll try to crack out a Mickey Mouse to look at. Thank you.