Floppy Disk Migration Pilot Project: Phase One Report
Completed 2/13/11 by Tracy Popp

Project Overview:

Computer magnetic disk media, like other magnetic media, is vulnerable to many preservation issues such as physical decay and hardware obsolescence. A pilot program to address these issues within the University of Illinois libraries collections began in the fall of 2009 with a mass email call for disks. Disk transfer began in January 2010 and the first phase of the pilot program was completed in May 2010.

Phase One Goals:

The intent of Phase One was to determine if we could transfer the contents of floppy disks to a networked directory in order to increase access and preserve the disk contents. If successfully accessed, disk contents were transferred to a networked directory on the University of Illinois' server[1]. Transferred files were not tested for readability. That is, we did not attempt to open any transferred files. As there are many factors which could influence whether or not a file is readable such as software availability, and hardware configuration file access and readability testing will be conducted at a later date.

Population:

Disks were attained from departmental libraries through a self-identification process. In response to a mass email sent to departmental libraries, collection managers indicated their interest in the program. Disks were collected from interested participants and returned to the collection managers once disk transfer was attempted or completed. Disks were collected from seven departments: BEL, Funk-ACES, Vet Med, Biology, History, Philosophy, and Newspaper, Music and Performing Arts, and Geology. Although there were eight respondents, we were unable to accommodate the eighth interested participant, John Hoffman, due to time and equipment constraints.

The project scope was limited to 3 ¼" and 5 ½" floppy disks as these are the most common disk sizes found in library collections. This project did not include optical disks such as CD-Rs. Disks were not limited based on file-structure; we collected many different disk formats including Macintosh and FAT-16 formatted disks.

Transfer Environment:

Disks were transferred to the library server location using a desktop computer running the Windows XP operating system. We modified this desktop computer by installing a 5 ¼" floppy drive, a 3 ½" floppy drive and a Catweasel[2] floppy disk controller. Our hardware and software configuration was based on

[1] (location as of 2010: \\libsys5.library.uiuc.edu\groupfiles\PresConsPublic\DiskConversionProject)

[2] See: http://lib.stanford.edu/digitial-forensics-stanford-university-libraries/catweasel-universal-floppy-drive-controller for more information about the Catweasel.

research and consultation with Cornell University[3] and Stanford University[4] both of whom have researched floppy disk preservation.

Floppy disks which were readable by our Windows computer were transferred using the XCopy command line utility. The command line was used in favor of cutting and pasting disk contents from the drive to the server in order to capture files on the floppy disk that may have been hidden or archived. A batch file script was written to automate some of the copying and directory creation processes.

In instances where the floppy disk was unreadable by the Windows XP operating system we used the Catweasel floppy disk controller and software to make an image of the disk. The proprietary Catweasel file was then converted to a non-proprietary format using a freeware computer forensics software imager.

Directory and File Organization and Naming:

Transferred files are organized into directories named after the originating library. For example, disk contents from floppies received from ACES are located in the ACES folder on the library server. Within each of these directories there are subdirectories where the transferred floppy disk contents are stored. In most cases, the directories are named after the call number found on the originating floppy disk or the disk title. If the floppy disk did not have a call number or title, the directories are named after the originating library and a sequential number. For example, the second disk with no call number that originated from the ACES library would be in the directory name starting with: ACES_0002. In some cases, text was appended to the file name to aid identification.

In instances where a call number corresponded with a series of disks another subdirectory level was created beneath the call number/title subdirectory. These sub-subdirectories are named based on the number of the disk in the series. For example, if the disk has a call number of 630 706762 and there are a series of disks associated with this call number, subdirectories would be created for each disk in the series. The subdirectories were named based on the disk's order in a sequence. For example, the content of the first disk in a series was transferred to a directory named "1_Disk".

> **Full Example:**
> *ACES has a series of three disks associated with the call number 630 706762. A directory named 630_706762 is created in the ACES folder. In the 630_706762 directory three folders are created named, "1_Disk", "2_Disk" and "3_Disk" to hold the corresponding disk contents.*

Outcome:

During our testing period nearly 500 disks were transferred. The majority of the disks were 3 1/2" floppies formatted with the FAT32[5] file system architecture which allowed for transfer with the XCopy

[3] http://www.library.cornell.edu/iris/migration/

[4] Spoke with Michael Olson from Stanford Library's Digital Forensics Lab on 1/29/10.

[5] http://en.wikipedia.org/wiki/File_Allocation_Table#FAT32

utility. Of the nearly 500 disks received, only 18 were 5 ¼" floppies. All but 21 disks –of which all were 3 ½" disks - were readable. Twelve of the disk read failures were due to physical issues; the sectors on the disk may have been physically damaged in some way preventing the floppy disk drive or Catweasel floppy controller from accessing the disk contents. The nine remaining disks were due to incompatible disk formatting (8 Macintosh formatted disks, 1 FAT16 formatted disk). These disks were accessed and returned prior to the arrival and installation of the Catweasel controller. The Catweasel floppy controller was not installed until midway through the transfer project due to several delays. I presume that these disks could be transferred using the Catweasel floppy disk controller as I was successful in transferring other disks with non-FAT32 file systems once the Catweasel arrived. All 5 ¼" floppies were transferred using the Catweasel. See Table 1 for percentages of successful disk access.

With the FAT32 formatted Windows compatible disks the average transfer time was approximately two minutes. Disk transfer using the Catweasel imaging was considerably longer averaging about 20 minutes per disk. The two-step process of creating an image using the proprietary Catweasel imaging software and then making the bit-stream level image accessible by processing with the forensics software greatly increased the transfer time.

Table 1 - Disks Transferred

Number of Disks Transferred	**Total Number of Read Failures**	**Percentage Accessible**
500	21	95.8%
Number of Disk Transferred per Disk Size		
18 5.25" floppies	0	100%
482 3.5" floppies	21	95.64%

Conclusion:

The percentages of successful transfers are very high. This may indicate that the floppy disks in the collections are not as threatened by physical failure as they are by hardware and software obsolescence. Indeed, it could be the lack of hardware to access these disks which has limited their use and kept them in relatively good condition. However, hardware and software obsolescence is a pressing issue as the resources needed to make the disk contents accessible and usable are becoming increasingly scarce. The greater challenge of reading the transferred disk contents lies ahead.