

Photo CD PCD to Tiff Migration Case Study

Louise Renaud, Canadian Museum of Civilization, Gatineau PQ, Canada

Jacqueline M.E. Vincent, Brechin Imaging Services, Ottawa, ON, Canada

Abstract

Kodak Photo CD, for the last 20 years, was used by many museums and archives around the globe as a means of digitally archiving their image collections. Since 2002, the Canadian Museum of Civilization Corporation (CMCC) has moved toward digitizing its material in a non-compressed TIFF format but prior to that, 85% of its digitized collection consisted of PCD proprietary formats on Kodak Photo CDs.

In early 2005, noticing degradation of its earlier digital collection, with some CDs being impossible to open and Eastman Kodak no longer supporting Photo CD technology, CMCC took measures to ensure the preservation of its collection by migrating the PCD images into TIFF 6.0.

This paper describes actions taken by CMCC to overcome and prevent further deterioration. It deals in a practical manner with day-to-day issues encountered and technical challenges faced during the migration process. It offers useful guidance that could be replicated throughout similar projects.

Introduction

The PCD to TIFF migration process was completed in partnership with Brechin Imaging Services, a digitizing company, who worked closely with the CMCC migration team to complete the image conversion and document the relevant conversion metadata. As a final count, 342,080 images were converted from 4,925 Kodak Photo CD Master, Pro Master, Photo CD Portfolio II and Kodak CD-R Ultima Gold discs. They are temporarily residing on 98 Fuji LTO-1 tapes, for an approximate size of 9.5 terabytes of data.

What is Photo CD?

Kodak Photo CD was a proprietary imaging technology developed in the 1990s for digitizing 35mm, medium format and 4x5 negative and transparency films using the Photo CD Imaging Workstation (PIW). It was popular among cultural institutions for digitizing photo collections because it could store high quality, high-resolution images in an encoded PhotoYCC colour space using visually lossless compression on specialized archival media. “.PCD” images were originally written on either Photo CD Master or Pro Master discs. These gold-layered discs were considered an “archival” storage media that could potentially last 100-200 years.

On a Master CD, each PCD image file contained 5 image resolutions in a single file called an Image Pac with an average file size of 4.5 MB. The lowest thumbnail resolution, Base/16, was 128 lines x 192 pixels and the highest 16Base component was 2048 lines x 3072 pixels, which produces an 18.5 MB file when opened. Master discs were used for 35mm slide and negative films and held about 100 images each.

Pro Master discs were used for high resolution 35mm, medium format and 4x5 film scans and held about 27 to 30 images.

These PCD images offered a 6th resolution of 4096 lines x 6144 pixels from the 64Base component, which produces a 72 MB file. However, an Image Pac could only hold 5 resolutions so the necessary 64Base component for the 6th resolution was stored in separate folders as part of the Image Pac Extension (IPE).¹

In addition, the Museum collection also had a third Photo CD disc type, called Kodak Photo CD Portfolio II that was designed mostly for multi-media presentations of mixed file types and sizes rather than solely for scanning films. It enabled a PIW operator to write lower PCD resolutions and other file formats to it such as TIFF, but it could also be used for the direct scanning of 16Base and 64Base images.

Finally there was a fourth type of CD, the Kodak CD-R Ultima Gold, that had to be used when the Photo CDs were discontinued.



Migration Methodology and Workflow

TIFF 6.0 with Adobe RGB (1998) colour space was chosen as the destination file format and colour space to which the PCD images would be converted. TIFF 6.0 was, and still is, the current standard for image archiving and digital preservation. Though it is slowly becoming more popular for archiving among some groups, JPEG2000, at that time, was not considered an option for conversion as a standard for long-term preservation.

For sake of clarity in this paper, the Migration Methodology and Workflow has been broken down into five main parts: pre-migration administrative steps; preliminary software assessments; the actual file conversion methodology; Quality Control procedures and finally, the metadata documentation process.

Pre-migration Administrative Steps

The image conversion was done off-site by Brechin, therefore the first step was for the CMCC team to establish a process to monitor the circulation of all Photo CDs. A tracking exchange sheet was developed and both parties, the Museum and the Contractor, signed the document as CDs went out for first migration or resent for corrections.

The second step was to decide upon a file-naming standard for the converted TIFFs and tapes that would enable access to them within existing CMCC collection databases and tape storage devices.

Names were pre-assigned to the new TIFF files as PCD images on Kodak Photo CDs had identical sets of numbers ranging from IMG0001.pcd to IMG0100.pcd that would have resulted in 4,925 “001.” TIFF images. To differentiate the new images and keep a reference to the original CD identification, it was decided to tag the CD number along with the image “001.” name. The CD identification itself was also standardized as it had been labeled through the years with two or four digits. New TIFF images were therefore named “CD” followed by 4 digits to indicate the year– followed by 4 digits for the disk number – followed by the 3 digits image number–.tif which overall referred to the image position on the original CD (ex: CD1998-0096-003.tif).

The tapes were allocated pre-assigned barcode identification labels and those numbers were incorporated into the administrative conversion metadata throughout the duration of the project (ex: TAPE003 – barcode: 101003). Barcode identification was critical for tape and image retrieval once deposited into the existing CMCC collection.

Pre-conversion Software Assessments

The two key goals of the migration process were to capture the conversion metadata and to preserve as much of the original image integrity of each PCD file as possible when converting to TIFF. This included colour balance, saturation, contrast, highlight and shadow detail and overall dynamic range. Because of the sheer number of images on CDs to be converted in a very short period of time, a software program was needed that would take all of these requirements into account.

Adobe Photoshop CS was selected as the image conversion software after various programs were tested and compared for best overall results and functionality. Adobe RGB (1998) colour management profile was chosen as the destination colour space to which to convert from PhotoYCC as it had a wider colour gamut than sRGB, was cross-platform between Mac and PC and is the current standard colour space for image archiving. Finally, Monaco was the colour management product used for calibration.

Preliminary testing and evaluation

Apart from Photoshop 7 and CS, software tested for converting PCDs included SilverFast PhotoCD from LaserSoft, Jasc Paint Shop Pro 9, Graphic Workshop Professional from Alchemy Mindworks and IrfanView. Other software products including open source command-line program ImageMagick were reviewed but not considered for testing.

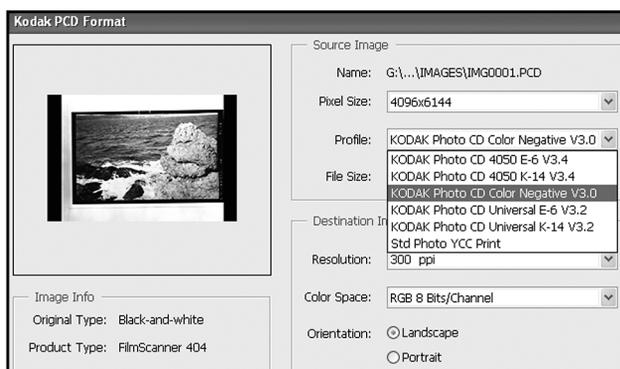
Image types tested were black and white (continuous tone), colour negative and colour transparency and were converted by the above programs and saved as TIFF. The converted TIFFs were then opened and compared in Photoshop. Highlight and shadow details, colour saturation, balance and overall dynamic range were compared for both visual and numeric quality using the “Info” palette tool and the Histogram palette and levels.

Except for Photoshop and SilverFast PhotoCD, all tested software converted the images using one internal profile with no options for a user to choose a film type source profile. SilverFast Photo CD did provide the option to create custom profiles for conversion but this was based on setting up profiles by manually choosing colour settings, balance, density, etc. rather than selecting standard profiles that were based on original film type.

Photoshop CS utilizes a Kodak Photo CD plug-in that does allow an image to be opened by selecting a source profile based on original film type. This feature has been improved over previous versions prior to Photoshop 7.

Why Photoshop CS?

At first, SilverFast PhotoCD was seriously considered but after the testing phase, it was decided that Photoshop CS was a better choice for three main reasons. First, the Photo CD plug-in allowed a user to select one of six standardized source profiles. The main concern with SilverFast was that there could be too many variables for setting profiles and therefore more chance of error when batch processing if a particular customized profile was less suitable for some images than others. In addition, for long-term preservation it would be imperative that these customized profiles somehow be preserved along with the migrated data.



Secondly, actions and batch conversion processes within Photoshop CS were more efficient to use on a project of this size than was the SilverFast Job Manager.

Finally, CS enabled the embedding of Photoshop’s History information into its Adobe XMP metadata for capturing source and conversion technical metadata. The “Browse” function of CS allowed all data to be displayed in its metadata window. The idea was that this information would be permanently stored within each TIFF file and the XMP data could be extracted should it be required in any given time.

PCD Conversion Methodology and Process

Batches of Photo CDs were taken off-site to the contractor’s location accompanied by a hard copy and MS Excel tracking list of CDs. Excel files were modified and used by the contractor throughout the migration for recording completed work and documenting everything related to the CDs that would be used for creating the associated administrative and technical migration metadata.

Seven PCs were allocated and used simultaneously for converting PCD images and one server-PC was set up for writing the LTO-1 tapes. Once the CDs arrived, each disc followed the conversion workflow process described below.

Preparation of folders and recording of CD info

CDs were migrated in numerical sequence and a folder was created in Windows Explorer for each CD based on its ID number. For example: folder name for CD96-234 would be CD1996-0234.

All relevant information about each CD was recorded on the CD tracking list: Date; Kodak Barcode number; whether images were colour, black and white or a mixture, and if a mixture then which profiles were used on one CD (usually only applied to Pro CD); whether disc type was Pro, Master, Portfolio or Ultima Gold; number of images on CD and final count successfully converted; images that could not convert properly or had quality issues; orientation of converted images if vertical; OS version (XP Pro SP1, 2 or Windows 2000); and finally any other information needed for completing the technical metadata later.

Photoshop actions and batch process set-up

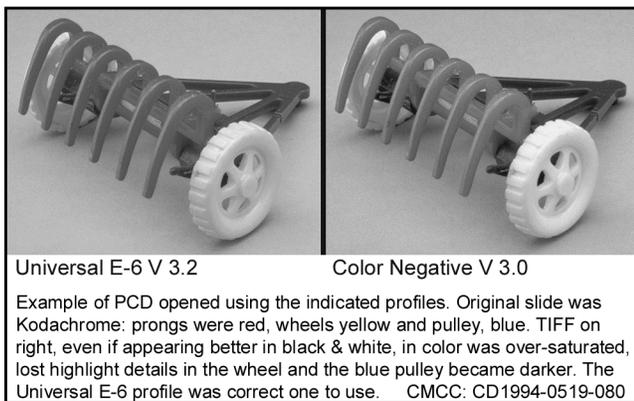
Photoshop actions were created for Master CD colour slide and colour negative and black and white film types. One action was created for all Pro CD images. A disc was put in CD drive, Photoshop Batch window opened, folders selected for source and destination files, parameters set for Pro or Master, then the batch conversion was started and TIFFs saved to PC hard drive. Master CDs ran till finished, Pro CDs ran semi-automatically with some manual user intervention to select Photo CD options for each file.

Source profile options for opening PCD images

The Photoshop Photo CD plug-in presented several options that needed to be selected to open a PCD file. Pixel dimensions of the PCD source images were set to 2048 x 3072 for Master CDs and 4096 x 6144 for Pro discs. Destination resolution was set to 300 dpi, colour space to RGB and orientation set to horizontal.

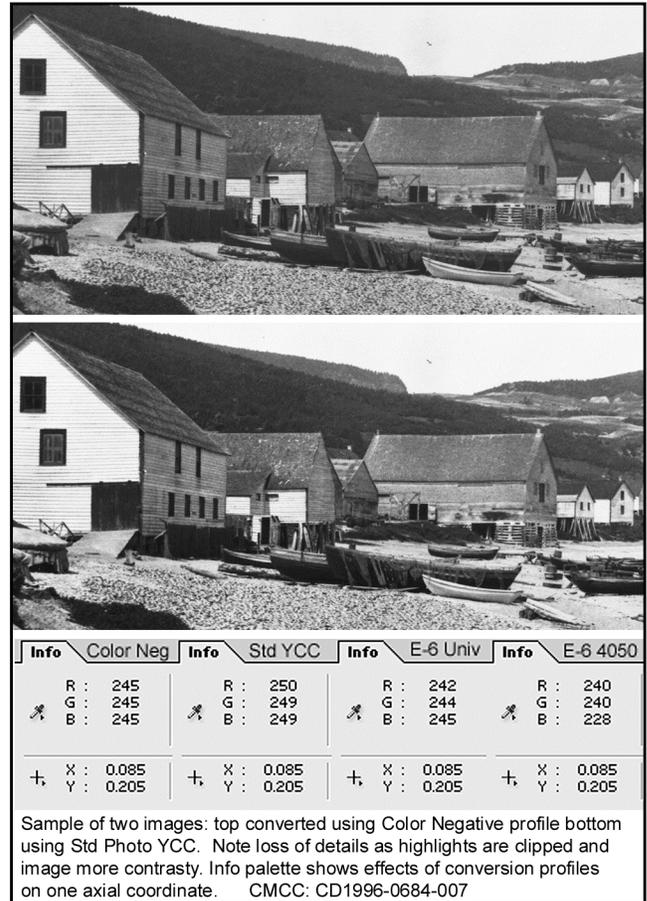
The “Image Info” section within the Photo CD plug-in window indicated “Original type” of film source. There are six source profiles from which to choose to open a PCD file.

“Color reversal (slide film)” type images, both E-6 and Kodachrome, were converted using “Kodak Photo CD Universal E-6 V3.2” profile. Though the plug-in does not actually indicate original type as Kodachrome, tests using the Universal K-14 V3.2 profile to open PCD images known to be from Kodachrome slides gave poor results with green/cyan colour shifts. For E-6 images, if either the Color Negative V3.0 or Std Photo YCC print profile was selected, test images resulted in highlights being too bright, details being lost in the deepest shadows, colours being over-saturated, particularly the reds and blues, and overall contrast being high.



Best overall results for “Black and White” type images were obtained using “Kodak Photo CD Color Negative V3.0”. Even though the original negatives were continuous tone black and

white, the PIW scanned these negatives as neutral colour images not as grayscale. For these PCDs, if the other Color profiles were selected, the resulting images was not neutral gray and, in fact, depending upon which incorrect profile was used, the highlights and other tones could be shifted towards cyan or green. The “Std Photo YCC print” profile did produce neutral tone TIFFs but they were higher contrast than when using the Color Negative profile.



Quality check of converted images

Once a CD was finished, the TIFFs in its corresponding folder were checked using the thumbnail view in Windows Explorer and compared with the Photo CD Index print to ensure all images had converted successfully. File naming and the CD’s ID was double-checked for accuracy in case of error when recording in the Photoshop Batch window. Ten percent of the TIFFs were opened to check image quality. This was first, last and every ten or so in between. If there were any problem PCD files or discs, they were either manually converted or if they could not be opened at all then this was noted on the CD tracking list and reported to CMCC. In some cases, a corrupted image could be replaced with the corresponding image from a duplicate CD.

Writing TIFFs to LTO-1 tapes

Each Ultrium LTO-1 Tape could hold 100 GB of TIFFs and there were 98 tapes numbered sequentially from TAPE001. A tape containing only Pro CD converted images of 72MB each amounted to about 1350 TIFFs whereas a tape of 18 MB Master CD TIFFs

held up to 5400. A mixture of both was anywhere between those amounts.

The first step was to create and sequentially name a new tape folder on a portable Maxtor hard drive and move enough CD TIFF folders into it to amount to 100 GB. For logistics of storage and retrieval, TIFFs from each Photo CD were kept together so some tapes might actually be slightly less than 100 GB. The Maxtor was then removed and attached to the Tape Drive server-PC for writing.

It was imperative for database access that there be no “folders” on the tapes, only sequential TIFF files. Therefore, TIFFs were then moved up from their CD-identified folders and stored loosely within the tape folder and empty CD folders were deleted. One last check in Windows Explorer was done to ensure all files had moved up properly, file names were accurate, file sizes were what they were supposed to be and images had not become corrupted when moved from CD sub-folders up to tape folder.

Using CA BrightStor ARCserve Backup r11.1 tape software, TIFFs were then written to tape and a subsequent back up tape was also done. Once finished, the BrightStor logs were checked to ensure that the backup had been successfully completed. An image quality check was then done to ensure the files themselves had backed up properly. To do this, 4 or 5 TIFFs were randomly spot-selected on the tape and “restored” back to the hard drive and then opened to make sure that they had not been corrupted during backup.

A designated Fuji barcode label was then affixed to the tape and submitted to the Museum with the migrated Photo CD discs.

Quality Control Procedures

The success of a project of this scope required an extensive file validation and Quality Assurance Program from both parties, the Contractor and the Museum, to ensure that all images were converted properly, that accessibility on tapes was secured and the correct Photoshop Photo CD profiles were being used for the corresponding film types.

CMCC General Workflow

Upon delivery by Brechin Imaging Services of the completed tapes and Photo CD returns, the following workflow was implemented at the Museum.

1. A list of the tape contents was produced and compared, using an Excel formula, with the original CD file inventories to ensure that all images had been properly migrated on the tape. Any missing data would cause the tape to be returned to the contractor for correction.
2. If all images were on the tape, pre-selected images were restored from it and downloaded onto the server.
3. Each Photo CD was opened and thumbnail views were displayed on one monitor.
4. The corresponding TIFF images were viewed through another monitor and image quality was assessed. File size, color mode and file naming were also reviewed at same time.
5. If the conversion was not adequate, both tape and CDs were returned for editing and would then be reassessed from the start.

6. Finally, the metadata records were reviewed and compared with both the latest tape inventory and the initial TIFF file naming list. If mistakes were found, corrections were done by Museum staff after clarifications and discussions with the Contractor.

CMCC Workflow image review – Quality control

When assessing the image quality, TIFFs were restored to shared folders on the network and two workstations with 80 GB hard drives were used. Each of the workstations were supplied with monitors that allowed simultaneous visual comparison, as one was used to display the PCD images and the other one used for the TIFFs. Image Quality control was done with each image opened at 25% and when there were discrepancies, the image was then reviewed at 100% of its size.

Mirroring Brechin, the software selected for visual manipulation was PhotoShop CS and screen calibration was done using Monaco Systems.

Because of time constraints and the quantity of images to be reviewed, only 15% of the overall images were assessed. In fact, the Museum reviewed every 7th image on the Kodak Master CDs, images nos. 2, 9, 16, 23, 30, 37, 44, 51, 58, 65, 72, 79, 86, 93 and 100. The images validated on the Pro CDs, were nos. 2, 9, 16, 23 and 30. Those numbers dropped when assessing the tape back-ups, as the only images reviewed were nos. 2 and 9 on any given CD.

The selection, displayed diagonally throughout the Photo CD index prints, allowed the Museum staff to immediately spot image gaps, as it altered the remaining image order and highlighted shifts in image numbering. This also facilitated their retrieval both from tapes and from Kodak CDs as it was fixed and became routine.

Metadata Documentation and Process

Since the technical metadata embedded in the PCD image proved to be impossible to retrieve, the Museum and the Contractor identified a number of fields critical for long-term preservation. It was also requested that the data itself be returned in two formats: an Excel spreadsheet (delimited cells) and a “.txt” file (plain text), in order to facilitate any future data migration.

Furthermore, the recording of the metadata was reinforced as the Museum needed means to differentiate original TIFFs that were added to the collection after 2002 and TIFF images that have gone through PCD conversion and may have suffered slight alteration in the process.

Administrative Conversion Metadata

The first step at the Museum was to register initial information, such as CD number, CD barcode, new TIFF file name, and the CD old number. The data was then transferred to Brechin Imaging Services where the fields associated with the conversion were added. Finally, the entire metadata file was returned to the Museum for validation.

Data associated with the image location was recorded (ex: tape number, tape barcode, and tape back-up identifier) along with some administrative fields (ex: Producer’s name, date, total number of images, comments, etc).

Technical Conversion Metadata

Other than the data associated with the CD, the following information was recorded in relation with the image and its source:

Field	Example
File format	TIFF 6.0
Resolution	300
Resolution unit	Inches
Pixel height	3072
Pixel width	2048
Level of compression	None
Source Media	Eastman Kodak Photo CD
Source File	PCD
Source Color Space	YCC

In association with Color management, the hardware and software, the following information was recorded:

Field	Example
Colour mode	RGB
Bit Depth	8
Color Space	Adobe RGB (1998)
Color management software	Monaco
Capture software	Adobe Photoshop CS Windows
Version	8
Computer	IBM 2.8 Mhz
Operation System	Windows XP Pro SP2

Regarding the conversion itself, the conversion profile was recorded (ex: Kodak Photo CD Universal E-6 v3.2) as it was essential in understanding the overall image quality and its conversion result.

Data was recorded throughout the conversion process and compiled along with actual tape content logs in Excel, to complete the final conversion metadata, and saved as text files – one file for each tape. Other fields for future use included File Modifications, Versions Generated, Institution and Date Changes.

As a temporary measure and until the information can be captured into the institutional collection management database, the entire Excel metadata file has been exported in a SQL database. This tool, now available through the Intranet, facilitates search and tracking of TIFF images residing on the tapes and allows exchange of information among museum staff.

Challenges, Issues and Problems

Overall most of the PCD files migrated without too much difficulty and Photoshop CS proved to be a good choice, although two issues with its batch processing caused some original migration plans to be altered.

The conversion project did encounter a number of technical and production challenges that had to be overcome. They usually involved a significant amount of manual intervention in the process particularly when migrating Pro and Portfolio CDs.

Photoshop Batch Processing Issues

First, scripting issues impeded with the fully automatic batch conversion of Pro CDs. This probably was due to the 64Base components being stored in separate IPE folders from the Image Pac PCD files and either code within the Photo CD plug-in or Photoshop itself was not able to access the components in batch

mode. To overcome this difficulty, a Photoshop CS action was created that still enabled semi-automatic batch processing.

A second coding issue allowed only the original PCD file name and .pcd extension to be embedded into the TIFF XMP metadata sidecar file (XML Packet) not PhotoShop's complete History when files were batch converted. This was disappointing because a PCD file's conversion metadata would have been embedded in its TIFF. The metadata had to be recorded manually and input into MS Excel file.

Missing IPE Folders and 64Base Components

Some Pro CD images when originally written to disc by the PIW did not include the IPE folders with 64Base components so those images could only open to the 5th resolution. This occurred randomly with no particular pattern and in some cases the first few images did not contain IPE folders, while in others it was the latter half. In a few cases it was 1 or 2 images in the middle of the disc.

XP Pro Service Pack 2 (SP2) and CD Drive Issues

Operating System updates created problems when some 6th resolution PCDs on Pro discs would not convert properly using some CD-reader drives on PCs with XP Pro SP2. Photoshop CS kept crashing once part of a disc had been converted, often around image 20 or 21. The same discs converted with no problems on Windows 2000 and XP Pro SP1. The problem was limited to a particular barcode series of Pro CDs not all the Pro discs. The problem was resolved by attaching a portable CD-writer drive, which appeared more robust.

This XP Pro SP2 issue also caused the same Photoshop crashing problems when converting some 5th resolution images that had been written to Portfolio CDs and Ultima CD-R. These Master CD images may have burned to Portfolio and CD-R due to a worldwide shortage of Master CDs at time of original scanning.

Media Problems and OS Issues

Some CDs, even after multiple attempts on various platforms, were impossible to open. Overall, their number represented less than 0.25% of the entire CD collection and there may have been many reasons for this difficulty: e.g. initial blank CD defect, CD rot, damaged CD surface, etc.

The Museum, in some instances, was able to revert to duplicate CDs, in other cases, new scans of originals will have to be done later. Even though the ultimate reason for this failure remains unknown, it serves as a reminder that optical support has its limitation and duplication is never a waste of resource.

Support Issues

Calls were made to both Adobe and Kodak technical teams to try and resolve the issues mentioned above but at that time neither had a solution other than the manual workarounds that we had already decided upon.

Tape issue

Tapes are not recognized as archival media but the decision to migrate onto such support was mainly based on the institution's current storage capability and the desire to recover the collection before it became obsolete.

Such a set-up, however, was complicated by the fact that the tapes have to be replaced on a regular basis and cross-reference between the tape content and its ever-changing barcode requires a good tracking system. To overcome the rotation challenge, the tape name along with its new barcode, have been included in the metadata registration and are updated on a regular basis.

Accessibility issue

Accessibility was also an issue. To view images from an LTO-1 tape, the data must first be read (merged) and the files then restored on a server. This process takes an average of 2 ½ hours per tape and any modification or correction to an image cannot be done on the tape itself. It has to be done through the file on the server and the new image burnt on another tape.

Conclusions

Overall, the Canadian Museum of Civilization Corporation has demonstrated along with Brechin Imaging Services that some processes can be successfully implemented to convert a large amount of Photo CD images to TIFFs in a timely manner without incurring great expense or loss of image quality. A methodology was established to capture relevant metadata to meet the Museum’s requirements.

Challenges associated with the Photo CD format, the effects of operating system upgrades, deteriorated CDs, PIW technical errors and other media problems had a negative effect on the batch conversion process but these challenges were overcome.

Managing the sheer volume of image conversions and monitoring the circulation of the collection outside CMCC facilities were challenges of their own and required good tracking procedures.

The process, which was completed in 5 months, does not solve the on-going preservation issue. It is another step, however, in ensuring the migration of such collection and one first attempt in salvaging 350,000 PCD images.

CMCC is now at the next phase, actively looking in acquiring sufficient server space and adequate software to manage the images currently on Tapes and facilitate their accessibility. It is the Museum’s objective to complete this transfer within the current fiscal year.

References

- 1 Fully Utilizing PCD Images (Eastman Kodak, Rochester, NY, 1994) Article 4, PCD-045.

Author Biographies

Louise Renaud: *Head, Photos & Copyright at the Canadian Museum of Civilization for the past 4 years. Louise is very much involved in the development of Digital Image Management and works in establishing preservation strategies and corporate guidelines toward securing the digital assets of the Museum Corporation.*
www.civilization.ca

Jacqueline Vincent: *Jacqueline has a background in traditional photography and digital imaging in the museum and archive fields. She was a photographer for the 1994-1998 CMCC Photo CD Digitization Project and scanned some images on the Kodak PIW to Master and Pro Photo CD. She is a partner in Brechin Imaging Services, a company that specializes in digitizing for the heritage and archive community.*
www.brechingroup.com

PCD - TIFF Migration Flowchart

