

檔案管理局

電子檔案保存管理機制委託服務案

**Electronic Records
Technical Service Center**
(1011130_01)

英福達科技股份有限公司 謹呈

民國 101 年 11 月

CHAPTER1.	HASH CODES AND ENCAPSULATION FILES	- 1 -
SECTION1.	EXPLANATION ON HASH CODES	- 1 -
SECTION2.	EXPLANATION ON ENCAPSULATION FILES	- 2 -
MEDIA MIGRATION TUTORIAL.....		- 3 -
CHAPTER1.	FORM OF MIGRATION.....	- 3 -
CHAPTER2.	MIGRATION OF DISCS AND DISKETTES	- 3 -
SECTION1.	MIGRATION	- 3 -
SECTION2.	VERIFICATION AND VALIDATION	- 3 -
SECTION3.	PAST EXPERIENCE	- 4 -
SECTION4.	NOTES ON MEDIA STORAGE.....	- 4 -
CHAPTER3.	AUDIO/VIDEO MIGRATION.....	- 5 -
SECTION1.	MIGRATION OF VIDEO TAPES	- 5 -
SECTION2.	MIGRATION OF AUDIO TAPES	- 6 -
SECTION3.	MIGRATION OF LP RECORDS	- 7 -
SECTION4.	VERIFICATION AND VALIDATION	- 7 -
CHAPTER4.	MIGRATION OF MICROFILMS	- 8 -
SECTION1.	EQUIPMENT.....	- 8 -
SECTION2.	VERIFICATION	- 8 -
MEDIA REPAIR TUTORIAL.....		- 9 -
CHAPTER1.	SOFTWARE AND HARDWARE.....	- 9 -
CHAPTER2.	DISC SCRATCH REMOVER.....	- 11 -
SECTION1.	EQUIPMENT.....	- 11 -
SECTION2.	MATERIALS	- 12 -
SECTION3.	ATTENTION	- 12 -
CHAPTER3.	RESTORATION SOFTWARE.....	- 13 -
SECTION1.	DATA RESCUE AND FILES DELETED BY MISTAKE.....	- 13 -

SECTION2.	HOW DATA RECOVERY WORKS	- 15 -
CHAPTER4.	VHS CLEANER.....	- 15 -
CHAPTER5.	FORENSIC TALON	- 17 -
	MEDIA DESTRUCTION TUTORIAL	- 18 -
CHAPTER1.	EQUIPMENT USED FOR DESTROYING MEDIA	- 18 -
CHAPTER2.	MEDIA SHREDDING MACHINE	- 19 -
CHAPTER3.	DEGAUSSER	- 20 -
CHAPTER4.	ERASER	- 21 -
	PEARS TUTORIAL	- 22 -
CHAPTER1.	INTRODUCTION TO PRESERVING ELECTRONIC ARCHIVES & RECORDS SUITE (PEARS).....	- 22 -
SECTION1.	FILE MIGRATION	- 22 -
Item1	Document formats	- 22 -
Item2	Image formats	- 22 -
Item3	Video formats	- 23 -
Item4	Audio formats.....	- 23 -
CHAPTER2.	VERIFICATION OF MIGRATION QUALITY.....	- 24 -
SECTION1.	QUALITY VERIFICATION COVERS	- 24 -
SECTION2.	IMAGE QUALITY	- 24 -
SECTION3.	VIDEO QUALITY	- 25 -
SECTION4.	OTHERS.....	- 25 -
CHAPTER3.	OCR / BAR CODE RECOGNITION	- 26 -
SECTION1.	OCR	- 26 -
SECTION2.	BAR CODE RECOGNITION.....	- 27 -
CHAPTER4.	RESTORATION OF AGED PHOTOS	- 28 -

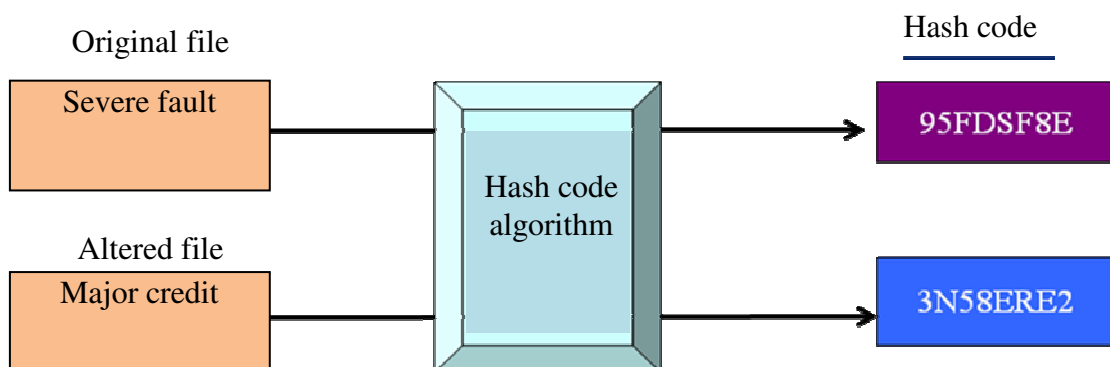
Chapter1. Hash codes and encapsulation files

Section1. Explanation on hash codes

In computer science, the hash function is a method of processing data. By applying a special algorithm, the searchable items are linked to the index for easy reference, and the value used in the data structure is referred to as the “hash code”.

Hash codes are also used for data security, as it can identify whether files and data have been altered, in order to make sure that the contents are indeed produced by the originator.

Today, the hash code algorithm is also used to encrypt passwords stored on database. Due to the irreversibility of hash codes, passwords can stay protected.



Please see the above figure:

In the original file, there was a severe fault keyed into the personal record, but was later altered to a major credit instead. Because a hash code was already calculated at the time the fault was registered, it is easy to tell by looking at the hash code that the file content has been altered.

Section2. **Explanation on encapsulation files**

An encapsulation file is like an onion, with its core covered in layers. The core consists of electronic records, attachments, sign-off procedures, signatory's opinions, signature data, hash code, metadata, media encapsulation files, and transfer (handover) information.



Media Migration Tutorial

Chapter1. Form of Migration

- CD to DVD
- Video tape (VHS, Beta, Betacam) to DVD
- 3.5" diskette to DVD
- Audio tape to mp3 or WAV
- LP record to mp3 or WAV
- Microfilm to JPEG, TIFF, or PDF/A

Chapter2. Migration of Discs and Diskettes

Section1. Migration

1. Use a reader to copy data from the disc/diskette into the computer.
2. Calculate hash codes for the copied data and apply digital signature.
3. Burn the copied data and the encapsulation file onto the DVD disc.
4. Proceed with verification

The Electronic Records Technical Service Center currently only has 3.5" floppy drives, and therefore migration services are available only for 3.5" diskettes, not 8" or 5.25".

Section2. Verification and Validation

Data migrated from discs and diskettes are validated using a proprietary media encapsulation and verification tool developed by the Electronic Records Technical Service Center. This tool compares the digital data stored on the original media against the data and encapsulation file burnt onto the new disc, and thereby ensuring data congruency.

Section3. **Past Experience**

The Electronic Records Technical Service Center had previously completed the National Scanned Image Disc Migration project back in 2011.

- Storage media: CD
- Volume: 1538 discs.
- Optical drives used: SONY & BUFFALO
- Disc-recording software: NERO

Over the course of media migration, some discs could not be read or were found to contain corrupt files. We have identified the following reasons for the above failure:

- Scratches and fungus on disc
- Disc was not recognized by particular drives
- Degradation of disc label
- Degradation of disc material

Section4. **Notes on Media Storage**

Storage media should be kept in a cool, dry place with controlled temperature (recommended at $18^{\circ}\text{C} \pm 2^{\circ}\text{C}$) and humidity (recommended at $35\% \pm 5\%$).

Avoid using sticky labels for disc production; write with water-based marker pens instead. Avoid stacking discs vertically in spindle containers, because the bottom disc may be damaged due to excessive weight over a long time. It is recommendable to store discs in individual cases so that they are not overly pressed.

Magnetic media such as video tapes, audio tapes, and microfilms are best kept in dry cabinets. To minimize oxidation and fungus on video/audio tapes and microfilms, these media should be cleaned regularly and rewound after use. If you do not have the equipment to retrieve data stored on these media, you may apply for a data migration service.

Avoid placing hard drives in places of excessive humidity, temperature, and dust. When in use, avoid shutting off power abruptly, and move only after the hard drive has stopped spinning.

Chapter3. Audio/video migration

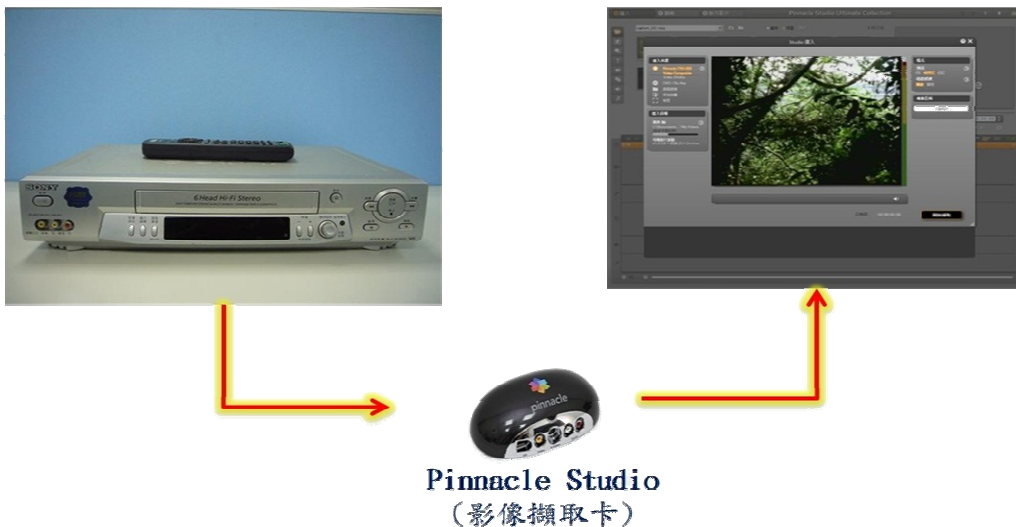
Audio/video migration includes

- Migration of video tapes
- Migration of audio tapes
- Migration of LP records

Section1. Migration of video

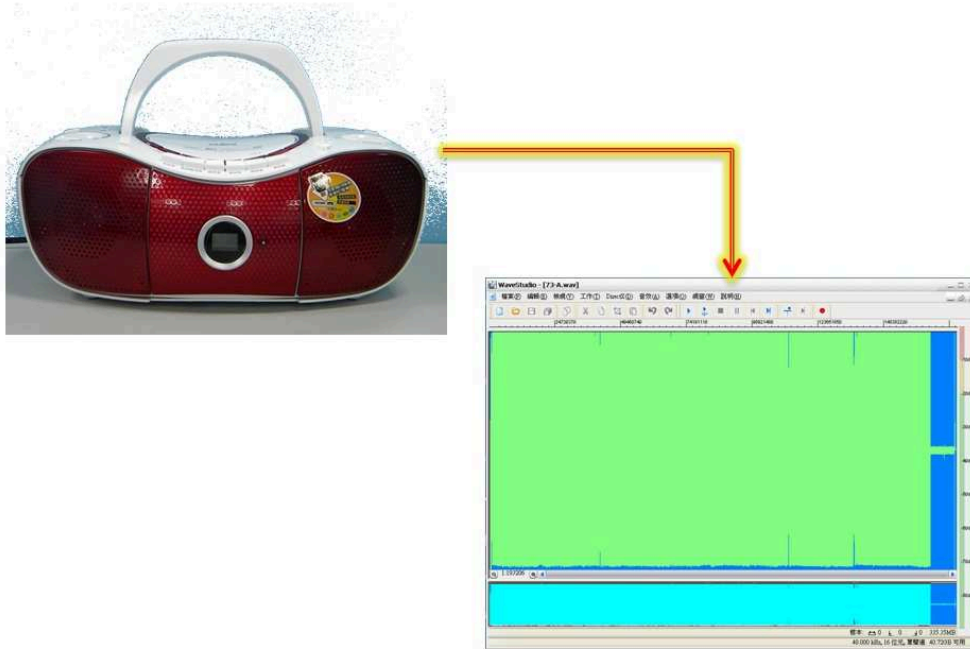
tapes

VHS、Beta(cam) 轉置數位影像



1. Playback the video tape and convert analog signals into digital signals using a video capture interface, which can then be saved into the computer.
2. Use the video-recording software developed by Pinnacle Studio to convert digital signals into MPEG-2 or AVI format.

Section2. Migration of audio tapes



1. Use an audio cable to connect analog output from the tape player (i.e., the headphone jack) to the computer (the microphone jack).
2. Convert analog audio into MP3 and WAV formats using sound-recording software (such as: Creative WaveStudio7).

Section3. Migration of LP records



1. Connect a NUMARK LP player to the computer using the USB interface, and send analog audio into the computer.
2. Convert audio signals into MP3 using an audio-recording software.

Section4. Verification and Validation

We verify video/audio migrations manually by playing back the converted video/audio files simultaneously with the original media. We then see and listen to ensure consistent quality, playback speed, and length.

Chapter4. **Migration of microfilms**

Section1. **Equipment**

Name of equipment: digital microfilm scanner

Supported microfilm types: positives, negatives, microfiche, cassettes, punch cards, 16mm and 35mm microfilms.

Migrated format: JPEG, PDF, TIFF



Section2. **Verification**

We adopt the following methods to verify microfilm migrations:

1. Manual verification: The microfilm is played back simultaneously with the migrated video file, which we then observe with naked eyes to discern difference in quality.
2. Verification by encapsulation files: The encapsulation file produced after the microfilm migration is verified using our proprietary media encapsulation and verification tool to confirm congruency of the encapsulation file and the external files.

Media Repair Tutorial

Chapter1. Software and hardware

We possess the following tools:

- Disc scratch remover



- Restoration software - R-Studio



- VHS cleaner



- Forensic Talon



Chapter2. Disc scratch remover

Section1. Equipment



The SIMO Tech scratch remover is believed to be the smallest and fastest model in the world. It can remove scratches of any depth and restore the disc's shiny surface for proper playback. However, scratches to the printed side are permanent and cannot be removed.

- It takes 10 seconds to complete one fast removal (world's fastest); a shallow scratch takes about two runs, while deeper scratches need more cycles
- 3-second fast cleaning (world's fastest)
- Accepts all disc types:
CD/CD-R/VCD/DVD/DVD+R/DVD-R/GAME-DISC/PS2/X-BOX etc
- Accepts 12CM & 8CM sized discs

- Dimension: 15X16.5X18cm
- Weight: 3kg
- Different remover fluids are used to treat different levels of scratch

Section2. **Materials**

1. Disc cleaning: Blue fluid and blue disc pad
2. Light scratch removal: Yellow fluid and yellow disc pad
3. Deep scratch removal: Red fluid and yellow disc pad
4. Extensive scratch removal: Red fluid and black square pad
5. Disc polish: White fluid and white disc pad
6. Plastic disc: For added thickness during disc repair

Section3. **Attention**

1. Please use disc-cleaning liquid
2. After applying repair materials, please use disc polish to remove scars.
3. If you still cannot remove a light scratch after 6 cycles, please try

deep/extensive scratch removal instead.

Chapter3. Restoration software

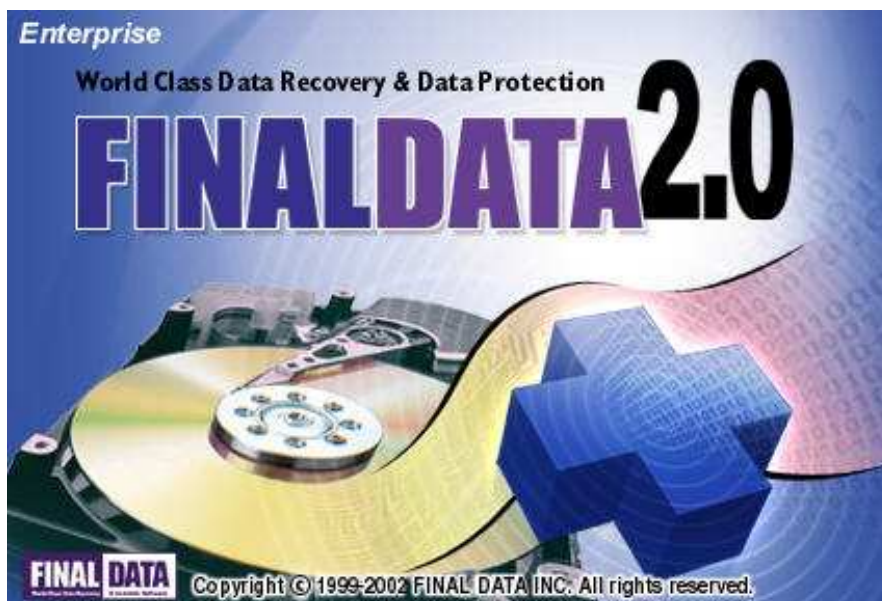
Section1. Data rescue and files deleted by mistake

We use the following recovery packages:

- R-studio

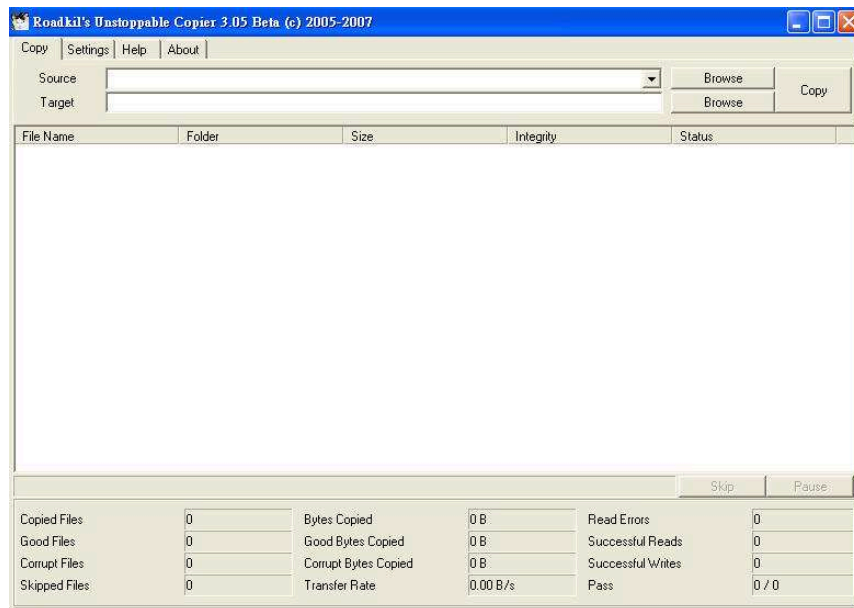


- FinaData: performs slower than R-studio



We use the following repair packages:

- Unstoppable Copier: This is a data rescue tool that applies a special method to repair corrupted data segments, and is not 100% accurate.



- CD Recovery toolbox: for rescuing data on disc



Accepted media

1. R-studio: Hard drives, thumb drives, CD, DVD
2. FinalData: Hard drives, thumb drives, CD, DVD, memory cards
3. Unstoppable Copier: Hard drives, thumb drives, CD, DVD
4. CD Recovery Toolbox: CD, DVD

Section2. **How data recovery works**

Let's compare a storage media to a book, where the first few pages contain an index, and the main context of the book is printed after the index.

The files that users see laid out before them are like the index of a book; opening a file is like flipping to the page where the content is stored. When users delete files from their computers, they are actually deleting the index, and not the real contents.

What recovery software does is recover real contents stored on the hard drive, whether or not they are indexed. However, files that have been corrupted or overwritten cannot be recovered.

How disc recovery works:

Data can be recovered from a corrupted disc by repeatedly reading data segments that cause errors, or by reading at a lower speed for better accuracy, or by ignoring the error and salvage whatever information possible from the disc.

Chapter4. **VHS cleaner**



We possess an RTI TapeChek 470 VHS cleaner. By applying a specially formulated wipe, dusts, oxidants, and small particles can be removed from the tape surface, and in the meantime smoothening the tape surface. The cleaner automatically detects the extent of degradation in the magnetic tape, and determines their quality grades. It takes an average of only 5 minutes to clean one VHS tape, which is hundred times faster than manual cleaning.

Chapter5. Forensic Talon



Forensic Talon is a forensic tool used to extract, save, analyze, and present data in ways that comply with law, in the event of hack, vandalism, fraud, or other forms of attack against computers.

When used in file rescue, it is capable of copying segment by segment and allows complete hard drive replication. Data should be backed up before using Forensic Talon on hard drive recovery; this is to prevent damage to the data during the recovery process.

Media destruction Tutorial

Chapter1. Equipment used for destroying media

We possess the following tools for destroying storage media

- Media Shredding Machine



- Degausser



- ERASER



Chapter2. Media Shredding Machine

The Media Shredding Machine is capable of destroying CD, DVD, audio tape, tape reel, diskettes etc.

The storage media is physically shredded into pieces, making it permanently unusable.



Chapter3. Degausser



The degausser removes all magnetic characteristics of the storage media by using a permanent magnet (13000 gauss). It takes only 20 seconds to complete.

Chapter4. ERASER

How ERASER works



ERASER is software that completely deletes files stored on media. It works by randomly generating dummy data and overwriting itself 65535 times. The contents that existed on the data segment are permanently erased after so many overwrites.

PEARS Tutorial

Chapter1. Introduction to Preserving Electronic

Archives & Records Suite (PEARS)

PEARS features the following functionalities:

1. Electronic record migration and verification
2. Video and image file repair
3. OCR / bar code recognition
4. Restoration and migration of aged photos

Section1. File migration

Performs migration on documents, images, videos, and audios.

Item1 Document formats

WDL to PDF/A,
DOC to PDF/A,
DOCX to PDF/A,
POSTSCRIPT to PDF/A
DOC to ODT

Item2 Image formats

TIFF to JPEG
JPEG to TIFF

TIFF to PNG

GIF to PNG

XLS to PDF/A

PPT to PDF/A

EML to PDF/A

Item3 Video formats

WMV to MPEG-2

MPEG-2 to H.264

Item4 Audio formats

MP3 to WAV

Chapter2. Verification of migration quality

Section1. Quality verification covers

1. Image quality
2. Video quality
3. Others

Section2. Image quality

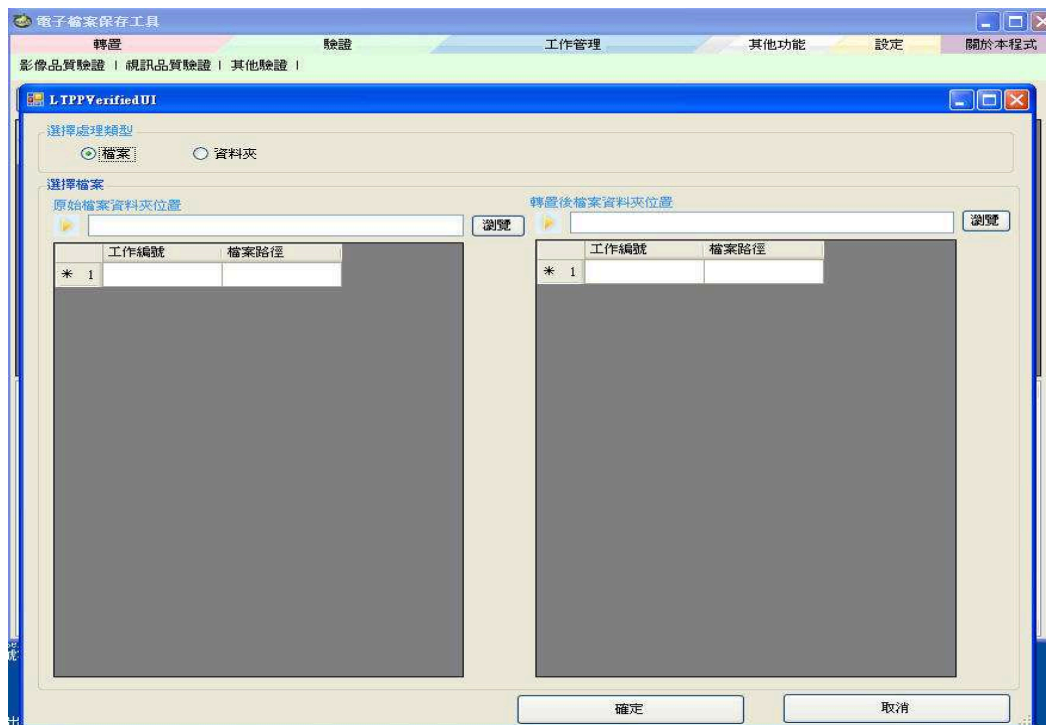
1. Methods for assessing image quality: MSE, PSNR, SSIM, UQI, HVS-PSNR.
2. Scores are calculated as measures for image quality.
3. Scores are calculated on the image before and after the migration using a pre-defined formula; by comparing these scores, we are able to determine whether the migrated image has maintained a certain level of quality from the original.

Section3. Video quality

There are six ways of verifying video quality, which involve a combination of trait identification, visual psychology, linear color differential, non-linear color differential, and overall evaluation.

Quality assessments may be carried out in three variation: Full-Reference, Reduced Reference, and No-Reference.

Section4. Others



For text files (e.g., WDL or DOC) migrated into images, the image quality verification approach is applied.

Chapter3. OCR / bar code recognition

Section1. OCR

"OCR" stands for "Optical Character Recognition", which is the process of capturing text contained in image files.

This technology is often used to capture words from scanned text images and save them in .txt format, therefore saving time on retyping while making the text readily available for editing.

Before OCR

檔 號： /100602/	
保存年限：5年	
裝	一、本局於98年11月11日召開「本局99年度委託研究及自行研究案研商會議」，並通過本組提列之99年度自行研究計畫「政府機關網站內容保存方式之研究」。
二、惟因本局「99年度電子檔案生命週期管理機制委託服務案」已先行以本局網站試作網頁內容之保存，且考量網站內容保存事涉本會職掌之政府網站建置相關規範，刻值組織改造業務移交之際，為避免造成事權混淆，擬暫緩進行本項自行研究案。	
三、本案如奉 核可，擬請企劃組調整本局99年度自行研究項目，以上所擬，當否？	
敬請	
核示	
訂	會辦單位：企劃組 承辦單位：檔案資訊組
後會	企劃組
決行	

After OCR

0990008395-P-0001.txt - 記事本

檔案(E) 編輯(E) 格式(O) 檢視(V) 說明(H)

檔號：/100602/
保存年限：5年

一、本局於98年11月11日召開「本局99年度委託研究及自行研究案研商會議」，並通過本組提列之99年度自行研究計畫「政府機關網站內容保存方式之研究」

二、惟因本局「99年度電子檔案生命週期管理機制委託服務案」已先行以本局網站試作網頁內容之保存，且考量網站內容保存事涉本會職掌之政府網站建置相關規範，刻值組織改造業務移交之際，為避免造成事權混淆，擬暫緩進行本項自行研究案。

三、本案如奉核可，擬請企劃組調整本局99年度自行研究項目，以上所擬，當否？

敬請
核示

會辦單位：企劃組
承辦單位：復會
檔案資訊組企劃組
決行

Section2. Bar code recognition

Bar code recognition is the process of converting bar codes into numbers. It is useful for capturing the correspondence number, allowing easier access by users.



After bar code recognition



Chapter4. Restoration of aged photos

Documents in early days were preserved on paper, which may be prone to wear and tear depending on the place and time preserved. We can scan worn documents into image files, then apply photo restoration software to recover the damaged part. Although the restoration is somewhat limited in its effectiveness, the file contents may remain intact nevertheless.

Before restoration



After restoration

