

檔案數位化之相關技術

Relevant Technologies about Archival Digitalization

作者/黎冠成 高素華
Li, Guan-Cheng ; Kao, Su-Hwa

述說歷史 – 檔案

Archive in support of history

歷史述說一切，歷史也將一再重演。

History bespeaks. History repeats.

歷史事件記述著人類文明的進步過程，即為某某人做了某某事的紀錄。該細節可以是主觀的、客觀的、關於時間、關於空間、關於過程、關於結果、以及後續影響，以及某人做了某事的事件的原因，某人目睹或被影響的過程，其背景是如何，以及事件的連鎖反應。聽起來很複雜嗎？有一件事是確定的：我們可以藉由檔案索引到他們。

The advancement of the civilization is narrated by the collection of historical events which were essentially about who does what. The details can be subjective, objective, epoch, spatiality, procedure, results, and following consequences. Namely, who did what for what reason, who witnessed or was influenced by the ripple caused by the actor or actress, the surrounding background pushed the actor, and the chain reaction which involved other incidences. Sounds complicated? One thing is for certain: They were indexed by the archives.

檔案儲存是一種瞭解歷史的方法，且提供一個情境足以抓住事件發展的方向。然而，單就檔案儲存而言，自以前到現在，這都不是一門容易的學問。在早期，書面保存的技術還很陽春。書寫在文件上的筆跡將會隨著時間褪色，影響褪色速度的原因包含有環境的溫濕度，或是遭遇到不可避色的天災人禍。時至今日，因應知識爆炸，快速累積

的檔案數量與日俱增，對有限的數位儲存空間造成了不小的壓力。然而，檔案儲存的困難度卻成了數位儲存媒體進步的主要原因。

Archival storage is a mean of apprehending history, or at least grasping a picture of how the circumstances would evolve. However, the storage itself is by no means easy, from antecedence to nowadays. In the earlier ages, the quality of hardcopy techniques was primitive. The calligraphy on the document tended to fade away, physically or chronically, from time to time, with dispatch contingent on the local proximity such as temperature and humidity or the global habitat such as inevitable destruction caused by natural and man-made calamities. Nowadays, the increasingly accumulated amount of archiving puts heavy pressure on the limited digital storage in response to the knowledge explosion. Notwithstanding, the impediment of keeping archive safe and sound is always the main reason that facilitates the advancement of the digital Restoration media.

當今，儲存媒體變得更多樣化，包含有書面紙本，歷史圖片，聲音紀錄，和動態攝影。所有的媒體都可以被數位化。在這一篇文章裡，我們探討一個用以取代書面作業的理想化檔案架構，並且利於電子化儲存，資料損毀後的重建，相關搜尋及有效的存取，機密文件管制以及逾期檔案銷毀。在另外一方面，數位儲存必需要有一個完整的檔案系統和可靠的數位資料庫。重要的技術包容有使用磁性裝置的重複備份、異地備份、資料壓縮以及傳輸。

Up to present, the storage media becomes much more manifold including hardcopy, historical photograph, acoustical recording, and moving pictures. All of the means can now be facilitated digitally. In this paper, we investigate an ideal archiving structure mainly on hardcopy which is in favor of the electronic storage, catastrophic restoration, relevant searching and expedient accessing, secrecy classification and out-of-date destruction. On the other hand, the digital storage requires a

decent archiving system and reliable digital database in support of the storage. The inherent techniques may include redundancy backlog with the use of magnetic devices, absent backup, data compression and transmission.

檔案擷取系統包含有兩個部分。一個是前置的索引服務，而另外一個是擷取服務。索引服務包容有關鍵字搜尋。對於歷史事件的完整瞭解包含有人、事、時、地、如何發生、以及為何發生。使用者可以針對事件使用數個關鍵字搜尋。更具體地說，輸入一個歷史人物，系統會輸出數個相關的事件；輸入一個特定的時間，系統會輸出數個在該時間內發生的事件，輸入一個地點，系統會輸出在該地點發生的所有事件。

The archival retrieval system includes two parts. One is the pre-stage indexing service while the other is the retrieval service. The indexing service includes the keyword search. A complete understanding into a historical event includes who, what, when, where, how, and why. An user can search for several keywords subject to an event. More specifically, she can recognize events relevant to a historical figure. He can designate a certain time period when events occurred. She can identify a place where events were located.

文件的掃描和存取服務是以幫助保存現有資訊和簡化作業程序為目標。該系統被設計於加快文件搜尋和取得的速度。並且，系統應搭配各式各樣的資料格式。我們也應提供整合性的服務。如果技術上可行，我們應建立一個可以自動存取、轉換、和引入現行資料的電腦程式。

The document scanning and retrieval services are aimed to help preserve current data and simplify operation processes. Such service system is designed to make the search-and-retrieval speed optimal. The system should work well with a variety of current data formats. We should

also provide integration services. If technically applicable, we can create a program to automatically retrieve, convert, and import legacy data into the new database.

考慮到電子檔案的格式，我們可以利用目前的技術水準來完成數位化儲存，不論原來的資料表是試算表、文字排版、桌面排版、或是影像處理。對於僅有格式為印刷品之文件，我們使用掃描軟體並且指派專責的人員來完成掃描的動作，最後將電子檔案統一儲存為「可攜式文件檔案(PDF)」。此標準的 PDF 閱覽和編輯軟體為 Adobe 公司的 Acrobat。我們也必須責成一個完整的編輯團隊，專職於資料輸入、校稿、編輯、和確保輸入資料的一致性。此團隊負責確保數位內容和印刷品內容是一模一樣的，如此才能確立高品質的電腦螢幕顯示。

As far as the electronic formatting is concerned, we can utilize state-of-the-art tools to convert electronic data from its native format, whether it's currently in spreadsheet, word processing, desktop publishing, or image processing applications. For data that exists only in printed form, we use scanning facilities and expert technicians to scan documents or images for consistent conversion to the portable document file (PDF). The standardized PDF reading and editing software is the Adobe Acrobat. There is also a need to maintain a full editorial staff proficient in manual data entry, as well as proofing and editing, ensuring the consistency of the input data. They are responsible for ensuring that the digital edition is an exact replica of the print publication, yielding a high quality page display.

有關於資料庫設計和整合服務的部分，我們應建立一個可靠的資料儲存空間，使其不止可以與現有的資料特性、輸出需求、和企業的規則配合，更需要有可以因應未來需求而擴充的彈性。一個設計良好的資料庫儲存空間提供有效率的搜尋和存取，強化在各個企業系統之間有效的轉換，並且促進多重資料輸出的功能性。該空間也必須支援諸如交叉查詢和相關資料結構等等進階的功能，軍事檔案系統尤其著重保

密，機密以及極機密檔案的取得應被嚴格地規範且遵守，以確保整個系統的安全性。

As far as the database design and integration services is concerned, we should have reliable data repositories that not only suit our specific data characteristics, output requirements, and business rules, but also are scalable enough to accommodate your current and future initiatives. A well-designed data repository allows for effective search and retrieval, facilitates efficient porting of information amongst business systems, and enables output to multiple media. The repository also supports advanced features such as cross-referencing and relational data structures. The restriction on the military archiving is mainly about secrecy. The access to the classified and highly classified files should be strictly regulated and followed to ensure the security of the whole system.

資料導引和索引是一種讀取資料庫的方法。這可以藉由原始的方法如目錄查詢，或是較高階的方法如關鍵字索引來達成。全文檢索將會是最終的目標，這期間功能強大的文字辨識軟體(OCR)將會扮演重要的角色。

Navigation and indexing is a mean for the access to the database. They can be accomplished through the use of naive methods such as table of contents or of more advanced methods such as keyword search. Full text search will be the final goal as a complete reference, which includes all text where powerful OCR (optical character recognition) finds significance.

關於過去的經驗

Regarding the past experience

從民國二十四年至三十六年，先總統蔣公曾兼任陸軍工兵學校的校長，這段期間的輝煌歷史已被記錄為各式各樣的媒體諸如書法、相片

和攝影。現在，我們享受了毫不費力卻能瞭解歷史的便利性。自民國五十七年以來，陸軍工兵學校已建立了一個健全的數位資料庫，用以乘載上述的檔案系統。除了密級的公文只能被保存為紙本且被鎖在保密櫃之外，我們計劃完成大約十萬件檔案的掃描。在計劃執行初期我們僅使用傳統的掃描器，然而，我們很快地發現到傳統掃描器的效率低落，並且改用了數位相機。拍照的邊界可能不容易控制，但是速度卻快了許多。時至今日，我們已完成了進度百分之十五的數位化，並且期待在不久的將來能完成到百分之百。我們珍惜祖先們所遺留下來的文化遺產。這也是我們重視檔案整合應用的主要原因，讓我們可以好好地瞭解到歷史檔案的好處。在現任校長蔡振義將軍的領導之下，我們組織了這樣一個團隊，用以因應如此神聖的任務－檔案數位化，我們並且期盼著，珍貴的文化遺產可以被完整地數位化和完善地被組織於資料庫中，讓我們的下一代從歷史中受益！

From 1935 to 1947, Ex-president General Jiang Kai Shi served as the commander of R.O.C Army Engineer School which had been faithfully recorded into a variety of media including the calligraphy, photocopy, and videotapes. Nowadays, we enjoy the convenience of easy access to the historical treasury. The R.O.C Army Engineer School has established a robust digital database in support of the archive system from 1968 till now. In particular, the file indexing service plays a crucial role in the application of the archives. Except the classified files that have to be kept in hardcopy and latched in the security depository, we plan to finish the scanning of approximately a hundred thousand pieces of official archives into the computer. In the beginning, we work with the use of traditional scanners. Nevertheless, we soon found the traditional scanners inefficient and changed the strategy with the use of digital cameras. The boundary of photo taking is not easily controlled during each time of taking, but the process is a much faster method than the old one. Up to present, we have completed fifteen percent of the digitization of archives, and are hoping to carry out the task in the near future. We cherish the historical legacy left by

our previous ancestors and commanders. This is also the main reason why we put the emphasis on the integration of the different resources that maximize the advantages of historical files. Under the guidance of Major General Tsai, Zhen-Yi, we organized a team in support of such a remarkable task, and we hope that the cultural treasury can be completely digitalized and well-organized in our database for our next generation to learn from the past!