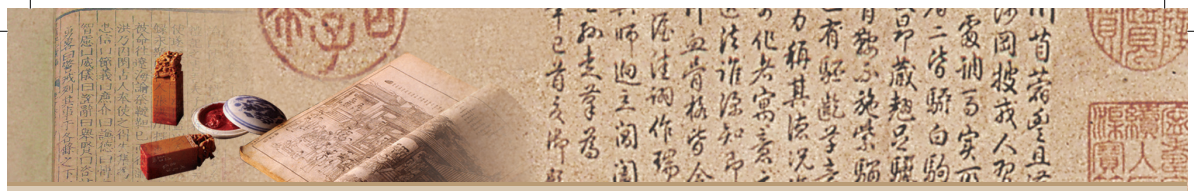
String-Bound Book Digitization Procedures Guideline

Chih-Tung Kao

Taiwan e-Learning and Digital Archives Program
Taiwan Digital Archive Expansion Project



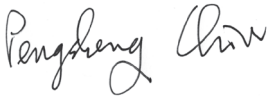
Preface

In January 2002, the National Science Council of the Executive Yuan of the Republic of China launched the National Digital Archives Program under the leadership of Dr. Ovid J.L. Tzeng, then vice-president of Academia Sinica. Over the past five years, the digital achievements attained by the project have been quite remarkable. The majority of these digital achievements may be viewed by the public (for non-commercial, lawful purposes) free of charge on the main portal site of the National Digital Archives Program (http://digitalarchives.tw). Since the beginning of 2008, the project leaders have embarked upon the second stage of development. The name of the project was changed to Taiwan e-Learning and Digital Archives Program (http://www.teldap.tw), and the executive director of this project is the current vice-president of Academia Sinica, Dr. Ts'ui-jung Liu. The Taiwan e-Learning and Digital Archives Program consists of eight core projects, one of which is the Taiwan Digital Archives Expansion Project (http://content.teldap.tw). Under the direction of Dean Fu-shih Lin, the College of Liberal Arts of the National Chung Hsing University, the Taiwan Digital Archives Expansion Project is closely collaborating with the experts and staff of the main TELDAP project and the other core projects. Our aim is to display Taiwan's cultural and natural diversity in digital form.

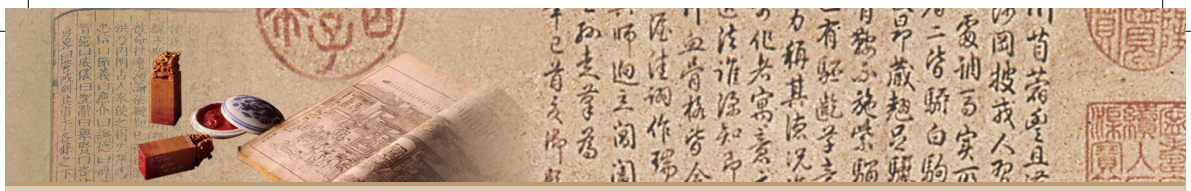
Since the completion of the first stage of the project, we have worked hard to record the digitization procedures of all of the organizations and institutions involved in digitizing their treasures. We have successively published over twenty digitization procedures guidelines related to the digitization of flora and fauna, archival documents, pictures, cultural

relics, and other materials. The purpose of these digitization procedures guidelines is to forge specific guidelines for the digitization of different kinds of archival material, determine priorities in digitization work, establish meta-data, computer programs needed for digitization, and the long-term preservation of digital material. In order to strengthen international exchange in digital archival work, we have now translated three digitization procedures guidelines (on the digitization of ancient Chinese string-bound book, ceramics, painting and calligraphy) into English. By doing so, we hope to share our digitization experiences and spur cooperation in digitization work in future years.

We will continue to work together with the experts and staff of each digitization project to refine the digitization procedures guidelines for different types of digital archival material. We welcome any suggestions or comments on the digital archives by users in Taiwan or abroad.



Peng-Sheng Chiu
Associate Researcher of Institute of History & Philology, Academia Sinica
Co-Principle Investigator of Taiwan Digital Archives Expansion Project
September 7, 2008



Index

2 Preface

4 Index

6 One. Introduction
I. Applicable Targets of this Guideline 9
II. Scope of Application 10
III. References 15

19 Two. Digitization Flowchart
I. Overall Program and Administration Structure 19
II. System Planning and Preparation 20
III. String-bound Book Digital Archives Flowchart 22

24 Three. Preliminary Operations
I. Work Planning 24
II. Digitization Method Selection 29
III. String-bound Book Arrangement 32
IV. Book Condition Inspection 34
V. Noting Book Movement 36
VI. Repairing String-bound Books 36

38 Four. Object Digitization Procedure
I. Work Planning 38
II. String-bound Book Scanning Procedure 40
III. Digital Image Post Processing 44
IV. Digital Image Backup and Management 44
V. Notices on the Digitization Procedure 48

51 Five. Metadata and Database Establishment
I. Metadata 51
II. Database Establishment 52
III. Challenges of Digitization Work 53

57 Six. Equipment and Cost Analysis
I. String-bound Book Scanning Environment 57

II. Main Digitization Scanning Equipment 58
III. Main Digitization Storage Media 65
IV. Cost Analysis 69

72 Seven. Benefits

74 Eight. Future Prospects and Conclusion

76 【Appendix 1】References
I. Specialized Books 76
II. Periodicals 76
III. Websites 78

80 【Appendix 2】Images Production Specification Reference Standards
I. Guideline of the National Library of China 80
II. Guidelines of the Institues in Taiwan 80
III. Cornell University Digital Library Image Scanning Archives Guideline 83

84 【Appendix 3】Comparison of Current Remote Backup Equipment of Different Units
I. Academia Sinica 84
II. National Palace Museum 86
III. National Central Library 88

89 【Appendix 4】Cost Estimation Examples
One. Platform Scanner
I. Platform Scanner 89
Two. Flatbed Scanner
I. Zeutschel Flatbed Scanner 90
II. Book Photographing Platform Flatbed Scanner (Separate purchase of lighting equipment and digital camera required) 92
III. Book Photographing Platform Flatbed Scanner (Unit already has lighting equipment and digital camera) 94

One. Introduction

Examples seen in ancient books of the east and west, such as ancients predicting the trend of times based on their observation of astronomical phenomena, are often veiled in mystery, but also show the reaction of people in nature. In the western world, the Bible's Revelations warns people of the oncoming Day of Judgement. In the eastern world, the "Tuei Bei Tu (Back Pushing Diagrams)", which circulated in non-governmental circles, predicted the fortunes of China's dynasties over the years. Legend has it that the "Tuei Bei Tu" was made by Yuan Tien-Gang and Li Chun-Fong of the Tang Dynasty's Astronomical Bureau after observing astronomy, constellations and referencing "Yi Jing (The Book of Changes)". At that time, Li Chun-Fong became so addicted to calculating the future that he didn't stop until over 2,000 years after the Tang Dynasty, Yuan Tien-Gang was alerted by this fact and pushed his back, saying "fate is a secret that should not be revealed, let us go back and rest!" Yet, the most remarkable thing of the "Tuei Bei Tu" that people take delight in talking about, is that its predictions of different times follow the order of its diagrams and will never skip any single diagram. Although the "will of Heaven" usually has its arrangements, people should still put in effort, observation of the rise of different dynasties all show that "the people's morale" is first roused before the will of Heaven can be followed.

As of today, many disputes remain over the messages that these two ancient books of the east and west wish to warn us of. Not getting into the question of whether you believe it or not, Chinese ancients gained apprehension by observing astronomy, constellations and referencing Yi

Jing; the life of Jesus and his twelve disciples were recorded through astronomical phenomena, signs, prophetic words, and revelations of prophets, and compiled into a book for the convenience of being passed down. The book of Ecclesiastes in the Old Testament says that "people with wisdom will find wisdom of the ancients"; wisdom of the ancients appear in ancient books and we are like Buddhists searching for the inspiring wisdom of Buddha in the mass amount of Buddhist scriptures.

The invention of paper and printing technology resulted in the widespread circulation of ancient books and records. However, different paper materials, printing technologies and binding methods (represented by string sewing) developed by different civilizations have resulted in the distinct physical characteristics of ancient books and records, and it affected circulation patterns and research conducted by later generations.[1] In addition, different understandings of "book collection" by Asian and Western civilizations have also influenced the circulation, inheritance and research of ancient books and records.[2] Still, books and records before 1912[3] had a different circulation pattern, and accompanying the development of computer technology and digital information in the 21st century, the wisdom and book collections of Asian and Western ancients have become transparent through digital archives.

At present, the main units related with digital archives of ancient books include the National Palace Museum, National Central Library,

[1] "We don't have any ancient books stores" edited by Shih Ke-Jung, from the website "pubhistory", Published: November 1st, 2005, Search: December 8th, 2005, http://www.pubhistory.com/show.asp?Id=1201.

[2] Same as the above.

[3] From the definition of "Ancient books cataloging rules", ancient books are hand written or printed traditional Chinese string bound books dated before 1912.

Fu Ssu Nien Library of Academia Sinica's Institute of History and Philology, and the National Taiwan University Library, which are all in the National Repository of Cultural Heritage Project. Computer technology, information digitization and the integration of ancient book related "binding methods", "illustrations in antique books", "name authority", "seal" and "books" have stimulated the acquisition of knowledge to change from passive to active; digitization achievements allow ancient books to accessible in places besides real libraries.

This guideline integrates digitization technology and standards of different libraries for units that also wish to engage in digitization to reference. This is a professional, operable and standardized guideline; in external integration, the establishment of this guideline and its standards allow operators or new staff to see from an overall integrated viewpoint and clearly understand the general situation of work, their own scope of work, detailed work items and execution procedures, reducing their uneasiness and fear of new work processes and procedures; this guideline can be used as training material for new staff. Finally, it can help managers find the cause for deviation after a project introduces a process, propose a way to improve the project, and further assist digitization workers to follow its standards and produce achievements, avoiding loss from human mistakes, saving time cost, while achieving production of digitization products with the same quality. Just like the words of deputy director Chen of the National Library of China: "International cooperation in database construction, avoid repeated or closed databases, and establish a unified international standard" , the final purpose of the Digital Archives Program is to be brought in-line with international standards.

I. Applicable Targets of this Guideline

This guideline was meant for a group and not an individual, for public libraries and not private libraries. In its digitization implementation process, it gives an overall description of the production direction and standards of the digitization work process. In addition, this guideline is based on system management of digitization work, allowing digitization work to become more transparent and more consistent with the concept of library automation by using system control for all digitization processes and personnel management.

Furthermore, ancient book digitization is a new interdisciplinary study that requires communities and organizations of different academic fields, including library and information science, computer science, philology, Chinese literature and history, to communicate, exchange and cooperate with one another. This guideline will describe each academic subject and field that is involved in hopes of giving consideration to both agencies and contracted firms; on one hand providing notices and work standards for agencies, on the other hand helping firms to have a general understanding of matters that require attention during ancient book digitization, allowing both parties to avoid causing further damage to the books and expanding cooperation room between agencies and firms.

Therefore, this guideline can be provided to groups that wish to engage in digitization and plan their digitization projects using a systematic method; digitization can be choosed to be out-sourced or not. Not only will this make minute and complicated procedures transparent, the automated system makes process and quality control easier, achieving the spirit of standardized work process and solving conflicts between store and use of digitization.

II. Scope of Application

(1) Digitization Objects

Applicable objects of this guideline are string-bound books; definitions of the objects are as follows:

a) Ancient books refer to written or printed books dated before 1911.[4]

b) String-bound books

Bookbinding technology is divided into different technologies for pages, scrolls and wooden slices, in which this guideline is based on books with pages sewed together using string. String sewing solved the problem of keeping pages together and books in its form rapidly became popular, turning it into the dominant form of existing ancient books.[5] Bookbinding technology is divided into different technologies for pages, scrolls and wooden slices, in which this guideline is based on books with pages sewed together using string. String sewing solved the problem of keeping pages together and books in its form rapidly became popular, turning it into the dominant form of existing ancient books.[6]

String-bound books are made from paper and string and can easily be damaged by natural disasters or over usage, showing the importance of digitization. This "String-Bound Book Digitization Procedures Guideline" (hereinafter referred to as this "guideline") was established exactly for this reason and is applicable to the digitization of "books with pages bound together using string" (the term "books" in following paragraphs refer to books in this form).

i) Features and development of book binding methods

[4] *General Situation of Ancient Book Verification* by Chen Cheng-Hung and Liang Ying, Shanghai: Shanghai Lexicographical Publishing House, October 2005, first edition.

[5] Same as the above.

[6] Same as 1.

Table 1. Development of Binding Methods

Development of string-bound book binding methods					
Form	Dynasty	Main Material	Features	Advantages	Disadvantages
Scroll Binding	Tang to Song	Mainly paper	Convenient for writing, light and portable, easy to collect, and aesthetic. Wrapped around a single core from left to right.	Convenient for writing, light and portable, easy to collect, and aesthetic. Paper is cheap and popular.	Content of a scroll is hard to find. Requires unrolling while reading.
Whirlwind binding (xuanfeng zhuang)	Tang to Song	Mainly paper	Still in the form of a scroll, can be turned by pages when unrolled, and looks like a tornado when rolled, hence its name.	Easier to look through data than a scroll.	More complicated and time consuming than other binding methods.
Accordion Binding	After the Southern and Northern Dynasties	Paper, silk	Opens up like a scroll, but is in the shape of a rectangle when closed.	Easy to browse, search through and place.	Will easily fall apart. Fold lines will be left on pictures that cross two pages.
Butterfly Fold Binding	Five Dynasties to Song and Yuan	Paper	Pages are turned like a butterfly; the book has a large fore edge.	Easy to browse, search through and place. Won't easily fall apart and the fore edge, top edge and bottom edge can be cut.	The pages are folded inward. Therefore, the reverse sides of pages are blank. Will easily tear where pages are folded and fall apart.
Doubled Leaved Book	Yuan to the middle period of Ming	Paper	Fastened together using paper nails and the back is glued.	Easy to browse, search through and place. Won't easily fall apart.	The fore edge, top edge and bottom edge are easily damaged.
String Sewing	The middle period of Ming	Paper	Fastened together using paper nails and string.	Easy to browse, search through and place. Even harder to fall apart.	The fore edge, top edge and bottom edge are easily damaged.

ii) Names of outer parts of string-bound books

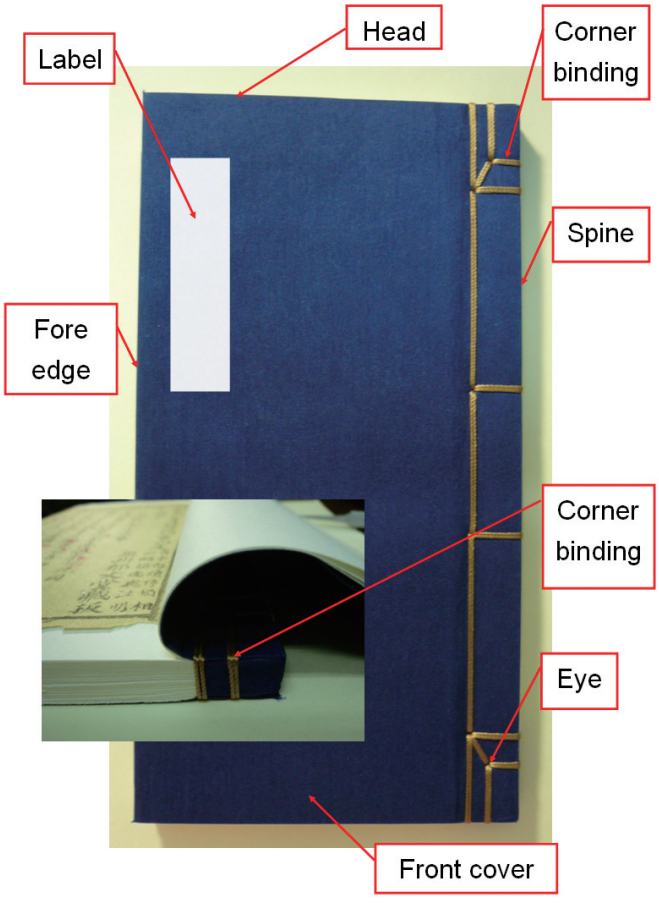


Figure 1. Appearance of a string-bound book

iii) Names of inner parts of string-bound books

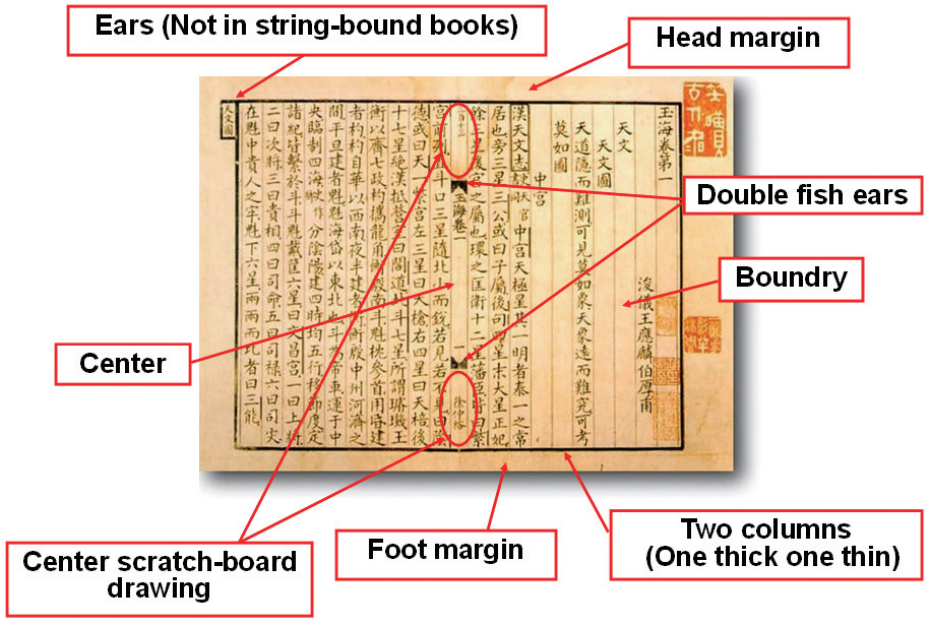


Figure 2. Interior of a string-bound book

(2) Digitization Implementation Method

Scanning is the main means for digitizing string-bound books and is divided into direct scanning and indirect scanning. Direct scanning refers to using a scanner to scan books, such as a flatbed scanner. Indirect scanning refers to turning books into microfilm first and then digitizing the microfilm.

In addition, if the book is in an extremely bad condition, then a digital camera can be used for digitization. This guideline only focuses on the direct scanning process.

(3) Work Items and Scope

The digitization procedure of string-bound books (as shown in Table 2) is divided into three stages in this guideline: "Preliminary Operations" ,"Object Digitization Procedure" and " Metadata and Database Establishment" :

a) "Preliminary Operations" describes the preparation work before digitization work. It includes two parts: 1. Work planning; 2. Ancient books arrangement.

b)"Object Digitization Procedure" describes procedures and notices when using scanning equipment and procedures for post processing of images. It includes four parts: 1. Color management model establishment; 2. String-bound book scanning procedure; 3. Digital image post processing; 4. Digital image backup and management.

c)"Metadata and Database Establishment" describes metadata standards of string-bound books and the establishment of cataloging rules. It includes two parts: 1. Metadata; 2. Database establishment.

Besides the three stages above, this guideline also includes scanning equipment and cost analysis that can be used as reference for planning digitization work; the space planning and equipment required for the scanning process is described and the cost for scanning operations is roughly estimated.

Table 2. Main items of the digitization procedures in different organizations

	Academia Sinica Institute of History and Philology Fu Ssu Nien Library	National Palace Museum Dept. of Rare Books and Documents	National Central Library Special Collection Division
Preliminary Operations	Create a scanning list	Create a scanning list	Check scanning objects
	Book condition inspection (Initial inspection when borrowing the book and book condition inspection before scanning)	Book condition inspection	The responsible staff plans the execution of supervision and assessment
	Trial scan	Trial photograph	Acquire the ancient books or microfilms
	Collection repair		
Object Digitization Procedure	Image scanning	Create image file image inspection (First verification on the digitization site, print the list, backup files and second verfication of the image files)	Original ancient book or ancient book
	Image inspection	Save image	Scan microfilms to the computer
	Create a disc		Collate image files and table of contents files with sub files
			Write each table of contents file into a DVD
Metadata and Database Establishment	Print list and gather statistics	Image inspection- Print list	Create a file listing ancient books
	Create a table of contents file and image links	Establish detailed items of the table of contents	Verify DVD is correct, convert files, save and provide links for searching
		Establish website data	

Table 3. Digitization work process planning

Digitization object	Books with pages sewed together using string	
Digitization implementation method	Digital Scan	Platform scanner
		Flatbed scanner
Work items and scope	Preliminary Operations	Work Planning, Digitization Method Selection
		String-Bound Book Arrangement
	Object Digitization Procedure	Color Management Model Establishment
		String-Bound Book Scanning Procedure
		Digital Image Post Processing
		Digital Image Backup and Management
	Metadata and Database Establishment	Metadata
		Database Establishment

III. References

References of this guideline are first listed out in a table, and then the direction and features of each reference is introduced to compare and explain which parts this guideline wishes to complement. References are divided into two categories: specialized books and periodical papers; they are divided into three directions based on their contents: digitization procedure, string-bound book repair, and research reports and papers.

Table 4. List of references

Category	Content	Booklist
Specialized Books	Digitization procedure	Illustrated Digitization Procedure Description Technology Collection Tasks and Techniques for Digitization of Documents
	String-bound book repair	Book Printing Development History Paper Series Sequel Overview of ancient printed book verification Library ancient book arrangement
Periodical Papers	Research reports and papers	"Object Digitization Work Seminar" Paper Series "2003 Ancient Books Digital Archives Seminar" Conference Paper "2005 Chinese Ancient Books Digitization Cooperation Establishment Seminar" Paper Series "A study on the Quality of Digital Images of the National Archives" Research Report

(1) Specialized books

a) ***Illustrated Digitization Procedure Description***

This was published by the National Digital Archives Program-Content Development Division. Each thematic group is introduced in a separate volume for a total of 15 volumes. Each volume introduces the digitization procedure of agencies in the thematic group, including: basic archives of each agency, project related websites, illustrated descriptions of their digitization procedure, and a survey on the relationship of simplified internal manpower, equipment and difficulties to cost. Although agencies of different thematic groups have different emphases and their description of cost is slightly insufficient, this publication still has reference value to the actual implementation of digitization.

b) ***Technology Collection***

This was published by the National Digital Archives Program- Program Office. It collects technology specifications of the "National Digital Archives Program", "Digital Museum Project" and other related projects, and is divided into metadata that "identifies and describes digital archives", actual digitization operations of "digitization production of archives", and information storage and development of "digital archives technology and system development"; however, it is mainly based on the description of metadata and gathering of technical documents.

c) ***Tasks and Techniques for Digitization of Documents***

This was edited by Ms. Hung Shu-Fen of the National Taiwan University Library Special Collection Division and published by the Training & Promotion Division of the National Digital Archives Program. It analyzes decisions and options, digitization preparation work, methods, specifications, database establishment, and out-sourcing of different steps of the digitization procedure. Although it

covers various digitization aspects, in terms of objects, it is relatively less capable of expressing diverse and complicated objects, such as the main digitization target of this guideline-historical document manuscripts; in terms of agencies, only the National Taiwan University Library is used as an example, relatively lacking information. The main direction of this guideline is to compare and analyze digitization procedures of different agencies for same objects.

d) ***Book Printing Development History Paper Series Sequel***

This was edited by Chiao Yen-Kuan and Chang Chin-Lang and printed by Wenshizhe Publishing Co. It mainly discusses the tools that spread ancient Chinese characters and the forms of Chinese books, in which a significant portion of its content is on the development of printing methods and binding methods in different dynasties, the definition of rare books, and how to repair rare books. This book provides background knowledge for this guideline in the definition, development and maintenance of string-bound books.

e) ***Overview of Ancient Printed Book Verification***

This was compiled by Chen Cheng-Hung and Liang Ying and published by Shanghai Lexicographical Publishing House. It mainly introduces methods used in different times for version verification, and it deepens its content using projects and support tools for verification. This book is special in that it is a colored deluxe edition, besides letting the reader clearly and deeply understands the various binding methods of string-bound books; it is also capable of showing the difference in contents of string-bound books from different dynasties, allowing better understanding of version verification. This book provides background knowledge on bookbinding methods for this guideline.

f) ***Library Ancient Book Arrangement***

This was edited by Wang Shih-Wei and published by the Beijing

Library Press. It mainly introduces the types and forms of ancient books, and includes discussions on the collection, preservation and circulation of ancient books and future developments and applications. This book provides background knowledge on ancient book content for this guideline.

(2) Periodical papers

Periodical papers mainly include papers published in periodicals or seminars by digitization workers and research reports written by related units. Periodical papers, such as the "Planning and Development of Ancient Rare Books Digital Archives System" published by Ms. Lin Miao-Hua of the Academia Sinica Institute of History and Philology Fu Ssu Nien Library, are conclusive plans and reports proposed based on practices of digitization work. Research reports, such as the "Digital library standards and construction-digital work standards and operation guideline" by the National Library of China, analyze the digitization work implementation situation domestically and overseas and interview units and firms engaged in digitization work. Furthermore, recommendations for different stages of digitization proposed in these reports are also valuable references for the establishment of this guideline.

Two. Digitization Flowchart

The digitization flowchart is divided into three parts, including "Overall Program and Administration Structure", "System Planning and Preparation" and "String-bound Book Digital Archives Flowchart". System planning and preparation is further divided into three parts, including "System Preparation and Development Planning", "System Architecture Planning" and "Metadata System Preparation and Planning". Below is an illustrated description of the flowchart:

I. Overall Program and Administration Structure

The overall program is divided two parts: digitization work and system preparation. Digitization work includes physical archives arrangement and management, theme establishment and list creation, digital scanning, and image backup and storage. System preparation includes the preparation of a bibliographic management system, storehouse management system, database search system and digital management system to manage the overall program.

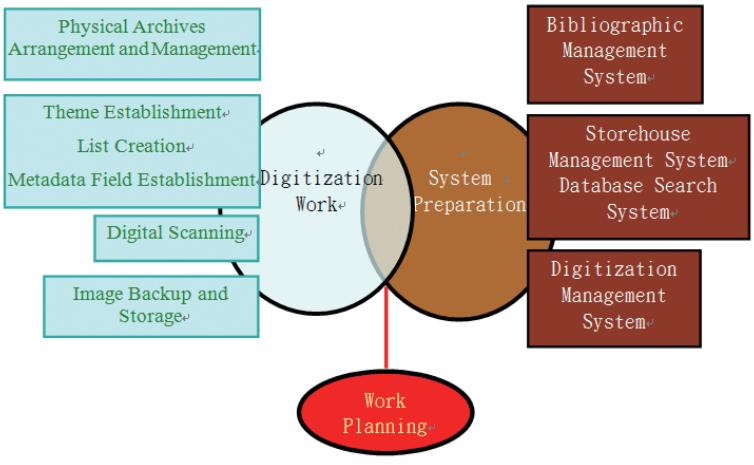


Figure 3. Overall Program Structure

In addition, the overlapped part of digitization work and system preparation is work planning. Work planning requires a detailed understanding of constituent and influencing factors of digitization work and system preparation, and must be based on the agency's actual budget.

II. System Planning and Preparation

(1) System Preparation and Development Planning

In system preparation and development planning, in terms of content, a system must comprise of three parts: "preliminary operations", "object digitization" and "metadata". It is recommended to plan the system for the convenience of unified control over the scope of the digitization procedure.

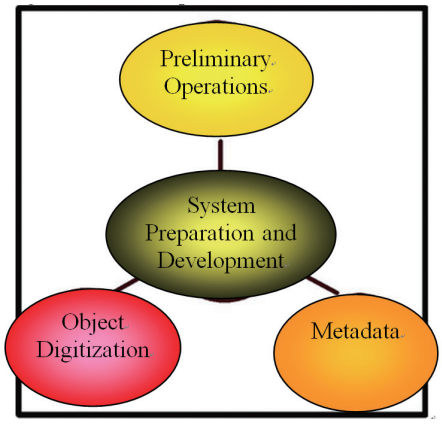


Figure 4. System Preparation and Development

(2) System Architecture Planning

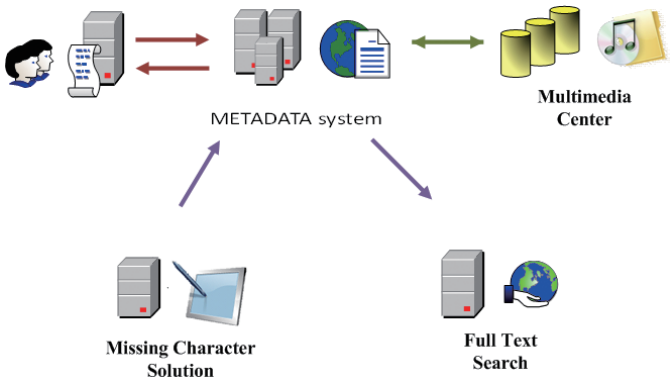


Figure 5. System Architecture Planning

(3) Metadata System Preparation and Planning

After completing metadata requirement interviews and content analysis, formulate a metadata functional requirements specification for the benefit of same or similar project systems; system development can be carried out individually or by collaborating with other organizations (such as universities or the private sector).[7] A metadata system includes three parts: "database system", "data management system" and "data search system".

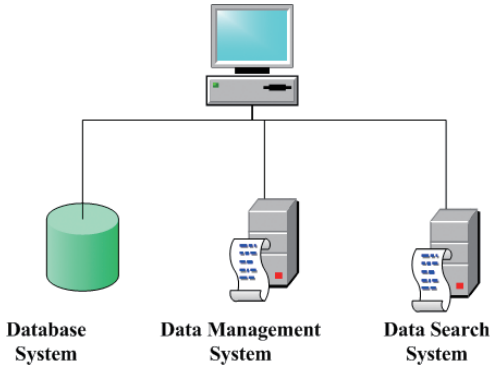


Figure 6. Metadata System Preparation

[7] By the Metadata Group of the National Digital Archives Program, from the website "MAAT", Published: December 20th, 2004, Search: December 1st, 2005, http://www.sinica.edu.tw/~metadata/design/lifecycle_new2.htm.

III. String-bound Book Digital Archives Flowchart

The digital archives flowchart is divided into three parts, including "preliminary operations", "image digitization" and "metadata and database system development". "Preliminary operations" mainly discusses work planning, digitization method selection, ancient book arrangement and list creation; "Image digitization" requires color management work to be carried out first before getting into ancient book scanning, image post processing, image storage and backup and importing images into the database; In "metadata and database system development", fields required by metadata must first be established (the common core fields established by the National Palace Museum, Academia Sinica Institute of History and Philology Fu Ssu Nien Library and National Central Library in the National Digital Archives Program can be referenced), before engaging in database preparation and development.

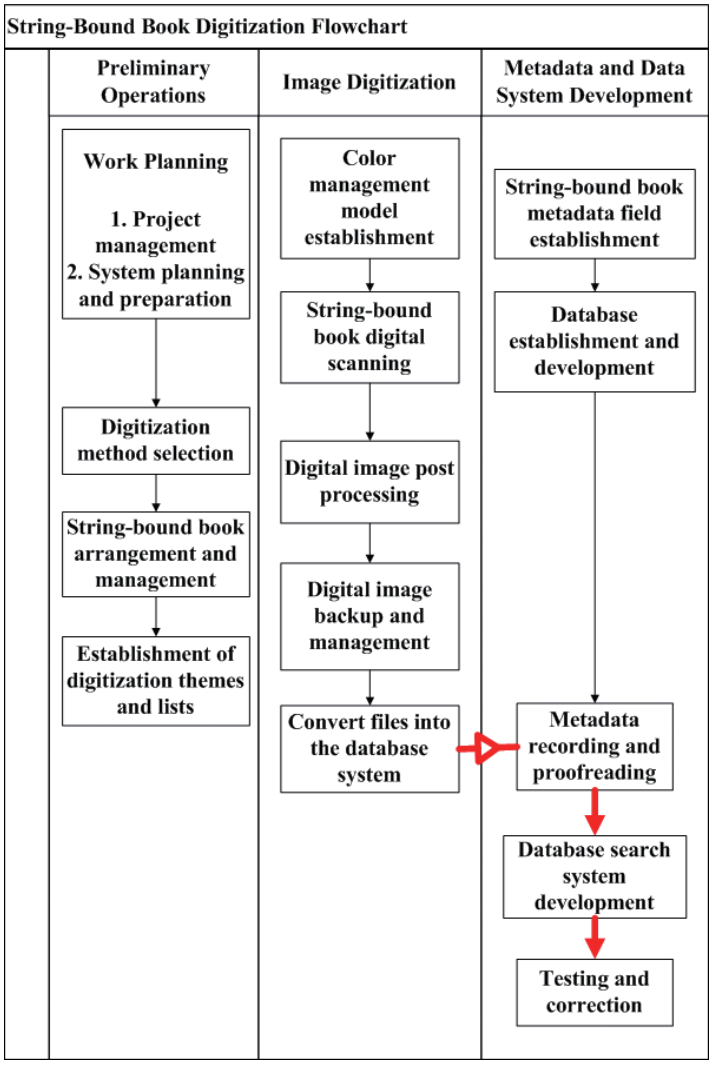


Figure 7. String-bound Book Digital Archives Flowchart

Three. Preliminary Operations

I. Work Planning

As shown in figures 3, 4, 5 and 6, notices on work planning in preliminary operations of digitization work planning and system planning and preparation are gathered and proposed. This chapter can be referred to together with chapters four and five of this guideline.

(1) Notices on Digitization Work Planning:

Digitization work planning must include the purpose of digitization, characteristics and value of collections, organization and personnel accomplishment, fund support and budgeting, project formulation and management, and assessment of results and adjustment. Below are brief descriptions of each aspect and recommendations.

a) Purpose of Digitization

Collections are mostly valuable data on paper and should be appropriately protected. In order to maintain the normal use of collections, modern digitization technology is used to virtualize these collections and allow them to be permanently preserved while providing virtual objects in coordination with regular usage[8], solving the conflict between protection and utilization of traditional paper data collections. It is recommended that the purpose of digitization be planned for holding or using and publishing. In addition, due to the rapid change in software and hardware, infrequently used data is often neglected when transferring old version data, causing regret that it can't be read in the days to come. Therefore, suitable software and hardware should be noticed following the development of technology.

b) Characteristics and Value of Collections

The types and characteristics of collections held by different agencies are not entirely the same. Therefore, as a principle for arranging the order objects are scanned, the characteristics and value of collections

[8] "Service and Management of Modern Library and Information" by Yang Ming-Tsung from the website *NCKU Library Journal*, Published: April 2000, Search: December 21st, 2005, http://www.lib.ncku.edu.tw/journal/journal/5/4.htm.

need to be considered. Using the National Central Library as an example, its digitization investment proportion is determined by the value of the ancient book's version, commonness and its appearance.

c) Organization and Personnel Accomplishment

Digitization work is the integration of numerous academic fields, including library and information science, computer science, philology, Chinese literature and history. From preliminary operations, implementation of digital scanning, to metadata and database establishment, all stages require the integration of different academic fields to complete digitization work and system planning and preparation.

d) Fund Support and Budgeting

The scale of funds that can be invested and how to budget both must be considered. The "Service and Management of Modern Library and Information" by Director Yang Ming-Tsung of the National Cheng Kung University Library mentioned that although technology dismantled the wall between libraries, economic issues seem to have formed an even stronger obstacle. How to budget a project requires careful consideration of the objects condition and relevant factors, such as the overall environment.

e) Project Formulation and Management

Formulate the project plan according to the aspects above and establish operation standards for project implementation in coordination with operation procedures.

f) Assessment of Results and Adjustment

Assess project implementation results and view corrections, adjust the original project plan and then move on to the next stage.

(2) Notices on System Planning and Preparation:[9]

[9] "Information on demand" from the website "Transtech", Search: November 21st, 2005, http://www.transtech.com.tw/kn-share/share_007.htm.

When planning a system, the actual object of digital archives, the agency's background and foundation, and the agency's actual requirements should all be considered. Factors that must be considered during system planning and preparation are described below according to three aspects: network setup, software/hardware selection and database establishment, and recommendations are made for each factor. Furthermore, this section can be referred to together with discussions on "metadata and database establishment" in chapter five.

a) Network Setup

This refers to all lines connecting the media control center and users; which is the network that covers the media control center and its broadcast range. Its core considerations are bandwidth and network connection and switch mode. General network connection and switching modes adopt hubs and switches: Hubs can only process a single data packet within a time frame, its performance is easily affected by data packet collisions when handling the transmission of several data packets; Switches can process the transmission of multiple "one to one" data packets within a single time frame. Therefore, when information on demand service is provided from a fixed location, switches are a more suitable option.

b) Hardware Selection

The utmost consideration of hardware is the support ability of the server and storage system. A server's main functions are managing, accessing and transmitting documents from the database. The transmission of image files does not permit any interruptions in between. Therefore, when selecting a server, one must first consider how many users will be using the information on demand function at the same time before deciding on what kind of server to use. Other considerations include: the network environment supported by the server (such as: Ethernet, ATM, ASDL Modem and Cabal Modem), its

expandability (can system management functions be easily enhanced) and overall investments.

For the storage system, different storage methods can be used for the great amount of image data, such as RAID, Disk Vault, Juke Box and magnetic tapes. These methods can all be used alone or in a combination with other methods. However, take notice of whether or not the storage system can be integrated with the server without affecting the system's performance.

If a unit does not have sufficient funds, it can consider using an image server instead of purchasing a relatively more expensive server. An image server is a disk in the disk vault that is used as a virtual server, providing an archives management system through a disk vault.

c) Software Selection

Considerations of applications begin from functions and interfaces; functions are further divided into user functions and system management functions:

i) In terms of users

Not only should the supported playing functions be noticed, the channels that users use to find required image data should also be kept in mind; this is asking the question of whether or not users can find the information they need through different channels (such as a simple key word search, detailed title or author search, and browse by themes).

ii) Management Functions

Consider whether or not the provided functions are convenient for management operations and are able to enhance services of the archives management system. Basic functions should include database management, user management, digitization process management and statistics gathering, meaning that any functions that help the administrator understand and control the system's usage situation should all be

considered.

iii) User Interface

This includes interfaces for users and administrators. Its evaluation principle is relatively simple: can users easily use the archives management system the first time they see it? This means asking the questions: Are there any unclear terms or phrases preventing the user from understanding a button's true function? Are the positions and size of buttons clear at a glance? Are links between system functions easy and smooth?

d) Database Establishment

Besides software, hardware and network, database establishment is another part of the system that should not be overlooked; consider what contents will be put into the completed structure. The main consideration of this stage is the potential/intended user. Only when the main user group is decided can an organization decide which image files meet requirements. However, deciding the database's contents and attributes are only initial steps, next is deciding the database's data quantity; the storage system's capacity is a basis for deciding the storage capacity for each type of data.

When under a tight budget, consider which objects should have priority for digitization, and digitize them into the database for users to use. After deciding the storage capacity for each type of data, select the collections (software/hardware) that need to be digitized; the following recommendations have been made for deciding on digitization priority:

i) Precious Data

Precious or antique historical documents and non-book materials are either unable to sustain repeated use or are merely collected and not permitted for use. Such data should have priority for digitization to encourage users to utilize it and reduce damage to the original document.

ii) Frequently Used String-bound Books

Frequently used string-bound books within the holding unit should also have priority for digitization. Besides reducing damage to the original document, it can allow more people to access the same data at the same time, avoiding the need to wait and take turns using data or having to buy a duplicate copy. The most important thing is that contents of the database must meet users' requirements.

II. Digitization Method Selection

Factors that influence digitization method selection include a subjective view of the object's condition and the objective factors: organization project budget, manpower allocation, purpose of digitization, image format, object considerations and equipment specifications. Below are descriptions of objective factors and recommendations. In addition, this section doesn't include organization project budget and manpower allocation.

(1) Purpose of Digitization

The digitization method adopted should vary with the purpose of digitization; different methods should be used if the purpose is permanent preservation, printing and publishing, or convenient for online access.

Table 5. Analysis of digitization purposes

Usage Purpose	Analysis	Example
Permanent preservation Printing and publishing	If the holding unit plans to publish or duplicate an object, then it should have even higher requirements on digital image quality.	The purpose of the National Digital Archives Program is permanent preservation and application promotion. Therefore, it requires the storage of high resolution images, using TIFF files with 300dpi and above.
Convenient for circulation and generalized utilization	When establishing a digitization specification, clearness of the digital image and internet transmission speed are both key factors of consideration.	Using the National Archives of Australia as an example, its purpose is to make circulation and generalized utilization convenient, utilizing limited finds to rapidly and on a massive scale provide the public with clear files. Therefore, its digitization specification is 180dpi, using digital cameras to photograph objects.

(2) Image Format Establishment

Below are descriptions of frequently used image formats in digital scanning.

a) TIFF

This file format is generally used on different platforms and different applications; its image printing specification is widely supported. TIFF format files do not cause loss to image quality after compression and can be used on different platforms and software. Therefore, it is suitable for preserving original data and can be used to produce other image formats or for printing high resolution pictures. However, TIFF format has relatively low compression rates and is rarely used in online transmission due to bandwidth considerations.

b) JPEG

This is a file format that provides high compression rates using a destructive compression method. The degree of compression can be adjusted in its saving process. It is mainly used in gray scale and full color image files and can process all color information under the RGB mode. JPEG format is not suitable for the preservation of original data because it causes loss to image quality. However, its high compression rates can shrink files to a size suitable for online browsing.

c) GIF

GIF files can only save 256 colors and have relatively smaller file size than other formats. Therefore, it is suitable for online transmission and is frequently used for previews on webpage. The image mode must be set as black and white, gray scale or 256 colors or else the file cannot be saved. If the original image is colored, then the 256 color mode might not be able to accurately record the original image. Therefore, GIF files are not suitable for preserving original data.

Table 6. Analysis of image formats

	GIF	JPEG	TIFF
Supports RGB Full			✓
Supports 256 Colors	✓		
Supports CMYK	✓		✓
Supports Gray Scale Files			✓
Compression Ability	✓ (LZW Compression)	✓ (High Efficiency)	✓ (LZW Compression)
Decompression Ability		Lossy Compression, Destructive Compression	
Suitable for: Preservation			✓
Preview	✓		
Browse		✓	
Frequently Used Fields	Animations, Transparent Images		Scanning, Drawing, Image Editing, Typesetting
Recommendation	Suitable for online file transmission	Suitable for online browsing, but not suitable for original data preservation	Suitable for original data preservation, but not suitable for downloading

(3) Equipment Selection

Equipment selection requires consideration of a subjective view of the object's condition and the objective factor organization project budget. Both factors are described below.

a) Book Condition

Select equipment based on the object's condition. If the book is in bad condition and fragile, it is recommended to use a scanner that faces the book upwards. In addition, don't use automatic paper feeding devices to protect the original script from further damage caused by inappropriate use of equipment.

b) Organization Project Budget

Besides a subjective view of the object's condition, organization project budget is also a consideration when selecting equipment; this part will be explained in Chapter Six Equipment and Cost Analysis.

(4) Feasible Plan Recommendation

Below are regulations and guidelines on image production specification, CD-ROM backup and naming principles of China, Taiwan and the U.S for the reader to reference.

a) Image Production Specification Reference Standards

Please refer to the image production specification reference standards at the end of this book.

b) Regulations and Guidelines on File Naming

Images should be saved in TIFF format, using one file for each book page. Files names should include the book's registration number. Using the Fu Ssu Nien Library as an example, the book's registration number is used as the folder name; different pages are differentiated using extensions of the file name and saved in the folder.[10]

III. String-bound Book Arrangement

(1) Ancient Book Arrangement

a) Overview

The criticism and research of classic documents was once mentioned in a paper of the third academic conference on cross-strait ancient book arrangement research. The criticism and research of classic documents is a profound branch of knowledge because it concerns textual research of explanatory notes, versions and catalogs and requires a thorough understanding of the whole article, sentences and phrases, and the author's implications. Therefore, books to be digitized are usually selected together by related research personnel and the project director.

i) Collection Work

Investigate and collect in pursuit of completeness. If a unit is to organize string-bound books, then it must first collect all versions of the books for comparison and arrangement. Therefore, the

[10] "Academia Sinica Institute of History and Philology Fu Ssu Nien Library Rare Books Digitization Image Production Specification" by and from Academia Sinica Institute of History and Philology, April 2005 revised version.

collection of anonymous books and the investigation and collection of Chinese books in distant places seem especially important.

ii) Criticism and Research

Compile a complete catalog of ancient books, "the study of catalogs is the first urgent matter, entry into knowledge must take this path". Compilation of string-bound book criticism is the foundation of string-bound book development and utilization. Not only should book catalogs and union catalogs be complied, special subject catalogs should also be compiled in accordance with needs of string-bound book arrangement research.

(2) Create Catalog Lists (System Required Lists)

It is hard to arrange the digitization order when the amount of digitization data is immense and books are in bad condition. Therefore, create digitization lists in batches to make it easier to control the overall digitization order of string-bound books. Carry out digitization in accordance with the digitization list planned by specialist staff. Below are a few recommendations for determining the digitization order:

a) Digitize books of the same topic as the holding unit's research topics.

b) Investigate topics that have undergone relatively less digitization domestically and overseas.

c) Negotiate topics via interlibrary cooperation.

d) Consider the string-bound book version's value, commonness and the book's appearance.

IV. Book Condition Inspection

After initial verification of library book condition, plan the amount of string-bound books to be digitized every week or month based on the library's total collection. Working staff must wear gloves when inspecting book condition to avoid further damage to the string-bound books caused by moisture from their hands. Prepare soft brushes, use paper slips (such as: kozo paper) instead of post-its, and implement digitization work in accordance with the regulations established based on characteristics of the library's collection. Besides needing to read regulations, it is recommended that the system fields include book registration, book movement, scanning calibration, book condition verification and book loan. Below are procedures of book condition inspection:

(1) Obtaining original books

Collecting and returning string-bound books should be done by specialist staff wearing gloves using a book lift. When borrowing books, not only should a data borrow form be filled in, the data must be returned the same day; borrowing hours may differ along with different libraries. In addition, it is recommended that books be borrowed in the order of their binding methods and size.

(2) Initial inspection when obtaining original books

When obtaining string-bound books, first mind the book's condition and determine whether if it requires repair, then verify the book's registration number or code, and then count to see if the total number of books is correct. Initial inspection items are as follows: vermiculation, damaged labels or markings, brown and crisp or chalking, and severeness of ink penetrating to other pages. If the book is in bad condition and cannot be repaired, then for the time being don't borrow it for scanning or photographing.

(3) Book condition inspection before scanning and determination of whether if the book requires repair

Carefully examine each page of each book before digital scanning, determine if pages in the string-bound book are brown and crisp or chipping off, and then start repair work. In addition, if the book is bound to tight, the National Palace Museum Dept. of Rare Books and Documents will first repair the book before commencing digital scanning.[11]

Some books will have missing pages or pieces of paper in between. It is recommended that a label be made for missing pages in the original book. Below are the main items of book condition inspections.[12]

Table 7. Main items of book condition inspections

Item	Content
Dog-ear	A dog-ear is a folded corner of a page. If the side of a page is rolled upward or folded, try to flatten the page as much as possible unless the paper itself is already fragile.
Crease	Creases are marks left from folding or applying pressure to paper. When there are creases, the paper isn't necessarily folded up.
Hole	When there are tears or holes in pages caused by natural or human reasons, determine if it can be repaired and whether if it affects the text on the page.
Mildew	Besides wood dust, pages will also exhibit mildew and vermiculation, the principle at this time is that these do not affect the text on pages.
Covered text	When the text is covered by paper fibers due to natural or human reasons, turn the book to repair staff if the situation can be improved.
Bookmark	Besides book labels or adhesive turning yellow, bookmarks will also be found in books.
Wrong page sequence	If the page sequence was bound wrong in the original book, turn the book to repair staff for correction.

[11] "National Palace Museum Dept. of Rare Books and Documents 2004 Rare Books Digitization Image Production Specification" by and from the National Palace Museum Dept. of Rare Books and Documents, 2004 revised version.

[12] Same as 10.

V. Noting Book Movement

Every digitization step must be noted in the system. This guideline will describe steps of the digitization procedure in the next chapter and will get into details here.

VI. Repairing String-bound Books

After paper documents are preserved for many years, various external factors will cause pages to be stuck together, or worst, cause the whole book to become one hard block with unseparable pages, affecting the books utilization. Pages stuck together are caused by extremely complicated reasons; it is the combination of the string-bound book's material, the environment, organisms and human factors. When repairing books, use paper that is consistent in fiber direction, thickness, color and texture with the book's paper, and then repair the book's holes, missing pieces or folded and worn down parts.

(1) Preliminary Operations

The system's repair management module is used at this time, after specialist staff decides on the repair order, allow them to determine the repair method and items, and make notes in the system on repairs made or conditions that couldn't be repaired to record the books repair history. It is recommended to have book repair and repair progress functions in the system.

(2) Operational Steps

After taking the book apart, first repair pages, and then repair the

cover. When repairing a page, start from the center and then work to the sides; mend large holes first and then mend smaller holes; if the book's mouth is torn, smooth it out first. The time required for repairing a book will vary according to the book's individual condition. Generally, a page will be dry a day after it is repaired. After completing repair work, begin the procedure for the book.

Four. Object Digitization Procedure

I. Color Management Model Establishment[13]

(1) Instrument Color Calibration

The purpose of color calibration is to fully preserve text and image information of string-bound books. Colors and text allow users to acquire the same information from digital archives as the original object, and they are able to understand the book's preservation state when being scanned. Color calibration, however, has always been the toughest problem in computer graphics and printing because many colors displayed on the screen are impossible to print out or exhibit severe color shift. Every step of the digitization procedure is tightly linked, input from the scanner to the screen, printer and output from the printing factory; every step is a potential cause of color shift. Factors that cause color shift are as follows:

a) Scanner

Scanners must use its own designated color calibration card; corrections are made after comparing the theoretical colors with the scanned colors.

b) Screen

Screen calibration requires the use of an optical instrument that is stuck on the screen to read colors of specific areas for correction.

c) Printer, Printing Machine

These also require color calibration to achieve the best output quality within the possible range.

(2) Color Calibration Method

Desktop scanners are based on the CIE Lab model, which was mathematically converted from the CIE Yxy model by the Commission

[13] *Periodical and Newspaper Full Text Input Procedures Guideline* (First Draft) by Cheng Wan Ju and Tseng Hsin-I, Taipei: National Digital Archives Program Content Development Division, September 16th, 2005, pages 14 and 15.

Intornation De'l E' clairage (CIE in short) also known as the International Commission on Illumination (ICI in short) in 1976. The color industrial standard-IT8 is used as a basis for desktop scanner color calibration.

Development of the digital camera in recent years brought an electronic optical device that replaced traditional film-Charge Coupled Device (CCD in short). Following the development of CCD or Complementary Metal-Oxide Semiconductor (CMOS in short) technology, equipment began having their own designated color chart for color calibration, and the device color description file-ICC Profile was produced. This standard image format maintains the output of front/back end equipment as consistent as possible. If any changes are made to equipment, the equipment will require color re-calibration and re-adjustment. This guideline uses a professional multi-purpose scanner (Brand: I2S, Model: DiGiBook10000RGB) as an example for color calibration.

(3) Exceptions

If in the digitization procedure higher brightness is required to show the object's details and characteristics, then consider the purpose and requirements of digitization: should the object's original color have the first priority, or should clearness of details be the utmost consideration? For example, in the scanning of plant specimen, if the emphasis is clearness of veins, then the object's original color will be slightly shifted.

(4) Output Application Mode

a) Printing (Printer)

Generally, individual usage doesn't focus much on printer color calibration. However, for professional color calibration, the only way to ensure consistency of output color quality is to execute color calibration whenever the paper, toner or ink used by the printer is changed.

b) Printing (Printing Machines)

In order to ensure that the printed document's quality is the same as the original document, printing machines also require color

calibration. However, this is difficult in practice due to the fact that most printing machines in Taiwan's market don't support color calibration.

c) Online Browsing

Files that have undergone screen and scanning equipment color calibration can be directly used for online browsing.

II. String-bound Book Scanning Procedure

Before beginning digital scanning operations, first trial scans the books and saves its files as a basis for verification in the future. Additionally, due to the immense amount of string-bound books in domestic and foreign organizations, digital scanning work is mostly out-sourced. The fifth point on notices of digitization work can be referred to by organizations when implementing digitization.

(1) Preliminary Operations of Image Digitization

a) Interleaving Paper

It is recommended to not use interleaving paper for general digital scanning in case it damages the book's center of folio; before trial scanning, if in the original book ink on one page penetrates to other pages, pages are stuck together and cannot be separated, or if the book mouth is split to wide and can't be repaired, then use interleaving paper and scan both pages together. Furthermore, if the book has relatively thinner paper and text on the previous page can be seen from the current page, the National Central Library will add white interleaving paper before scanning[14], but it is recommended that the size of interleaving paper be smaller than the book's page size.

[14] "National Central Library Original Ancient Book, Rubbing, Ordinary String-bound Book Scanning Image Format" by and from the National Central Library, October 25th, 2005 revised version.

b) Not taking apart the binding string

Working staff must wear gloves when carrying out digital scanning to reduce the amount of damage to string-bound books. Generally speaking, digitization should be carried out based on the principle of preserving the book's original appearance and should strive to reduce damage to string-bound books. Therefore, the binding string of string-bound books should not be taken apart. If the book is not suitable for digitization operations, then digitization should be postponed; original books shouldn't be damaged for the sake of digitization. If the binding string needs to be taken apart for book repair, then ask professional repair staff to handle this task.

(2) Color Calibration

Use dedicated reflection or transparent color samples to adjust the scanner's high light, shade and middle Gamma values. When necessary, adjust the R/G/B or C/M/Y/K mono channel values to make the digital image's scale, color and gray balance consistent with the color samples. The following are scanner calibration principles: Before formal digitization scanning operations everyday, it is recommended to follow the color calibration software's standard calibration process. Although this action is somewhat time consuming, it can maintain the accuracy of digital image colors.

(3) Preparing the Original Document

Operation principles need to be established for the handling, maintenance and browsing of original documents. These three steps all require the use of gloves and soft brushes. The holding of books should be based on the principle of balancing it on one's palms, and it is recommended to handle original documents with the whole set as one unit to prevent them from being scattered.

(4) Trial Scanning

First conduct trial scanning of string-bound books; trial scanning

results can be used as calibration samples. Although this action is somewhat time consuming, it can reduce damage to books caused by inappropriate machine operation and accidents and can also reduce the occurrence of digital image problems. When planning or changing daily digitization scanning lists, it is recommended to pick a book from the book set that is most likely to have problems for trial scanning. When equipment malfunctions, contact the manufacturer or software vender for support; if the problem cannot be solved, then suspend digital scanning work.

(5) Printing the Scanned Sample Image

After scanning and verifying string-bound book images, it is recommended to print the sample image in full color and have scanning and verification staff sign the image for future verification. Record trial scanning results in a form and print it out; commence formal digital scanning work after saving the image in a hard drive.

(6) Formal Digital Scanning

The digital scanning method used is determined by the string-bound book's condition; scanning methods are generally divided into original book scanning and parallel scanning. A standard color chart and scale must be added when scanning as a basis for color calibration and determining the original book's size. Clip paper to the sides of the scanning panel to make it clear and clean.

Run the scanning color calibration program before formal scanning everyday to check light and scanning effects; browse the image when running the program to avoid image color shift from occurring. In general, if binding strings of the original book are not taken apart when scanning, do not close the scanner's cover to avoid damaging the book; closing the cover will achieve better results if the binding strings are taken apart for scanning. Finally, make notes in the system.

In addition, staff are required to take training courses before operating equipment, which is why it is not recommended to let part-time workers handle this task; besides shift issues, part-time workers are not from this field and need longer training because the information will be hard for them to absorb.

a) General Digital Scanning

The different equipment is divided into flatbed scanners that face books upwards and platform scanners that face books downwards. Equipment selection requires consideration of book size, book condition, unit budget and manpower allocation. Therefore, the compatibility of equipment functions and objects should be noticed during scanning.

b) Cross Page Image Scanning

When pictures in string-bound books stretch across pages, the National Palace Museum and Academia Sinica Institute of History and Philology Fu Ssu Nien Library have different considerations. The National Palace Museum divides the picture into several regions based on its size, and the overlapping area between two regions is at least 2 to 4 cm to benefit piecing the picture together; the scanned picture is pieced together after image files are verified to be correct.[15] Academia Sinica Institute of History and Philology Fu Ssu Nien Library, on the other hand, tries to maintain the book's original appearance and doesn't use this method.

[15] "National Palace Museum Dept. of Rare Books and Documents 2004 Rare Books Digitization Image Production Specification" by and from the National Palace Museum Dept. of Rare Books and Documents, 2004 revised version.

III. Digital Image Post Processing

Image files from digital scanning may only be backed up on a DVD disc or magnetic tape after verification. Normally the project's research assistant is responsible for verifying images, but if the research assistant already has too many tasks to perform, then a college or above part-time worker may be asked to complete verification. Furthermore, it is recommended that the system fields include convert to DVD progress, verification quality statistics, repair progress statistics and scanning type.

Image verification is generally conducted twice, the first time is on the digital scanning site and the second time is carried out by verification personnel. Image files are usually checked to see if they are straight, have black edges, blurred or inappropriately cut, and then the problem is noted in the verification list. A third verification is only needed when pages with problems are corrected, but original book preservation is still the main consideration.

IV. Digital Image Backup and Management

The suitable storage media for data storage and backup is based on actual requirements and environment. Therefore, planning personnel's understanding of the overall environment and requirements should be depended on to avoid evaluation errors and purchase shortage. Below includes two aspects: digital image backup and digital image management; backup reasons and backup structure are briefly described. In addition, chapter six will provide detailed descriptions of storage media types, their

storage capacity and transmission speed, and introductions to industrial application products.

(1) Digital Image Backup

a) Backup Reason[16]

i) Component Fault

Computer systems are made of many components, the malfunction of any component will result in data being damaged or lost. For example, a hard drive malfunction will damage data, and an unstable system crashing will result in data write errors or the original file being replacing.

ii) Unpredictable disasters (natural disasters, war, and artificial damage) may cause hardware malfunctions, which will make the system unusable.

iii) Invasive Damage (system being hacked, causing data to be deleted, modified, invalid or become malignant).

iv) Human Factors (most commonly the deletion of files by mistake).

b) Backup Structure

Data backup is generally done in a distributed structure or centralized structure, the former is more suitable for personal data backup. This guideline was made out for groups and will focus on centralized data backup. Centralized data backup normally uses backup software (such as ARCServeIT, Veritas Backup Exec... etc.), to achieve scheduling, backup records, restoration and disaster recovery, and the addition of a client backup agent allows backup programs to have cross-platform real-time backup ability.

[16] "Storage Equipment and Backup" the collaborative work of several authors, from the website "study-area", search December 22nd, 2005, http://www.study-area.org/storage/storage.html.

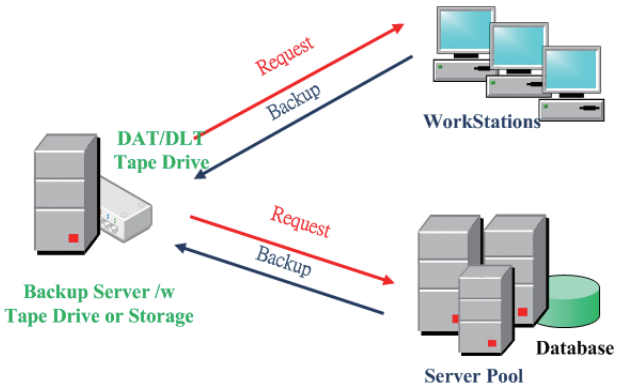


Figure 8. Centralized data backup structure

c) Other Relevant Matters

i) Storage Media and Writing to Disc

Optical resolution 300dpi images are usually saved in the standard uncompressed TIFF format, it is recommended to make two backups on DVD and one backup on magnetic tape for safety. Storage media should be chosen based on the unit's actual situation and requirements; image file size, unit budget, current equipment and space are all factors that need to be considered.

ii) Storage Media Labels

CD permanent markers are generally used for writing labels, but hand writing labels is relatively time consuming. Thermal transfer printers are recommended for labeling data, they are neat, provide complete information, fast and can save manpower. Labels should have the disc number and book registration number.

iii) Storage Media Preservation Environment

It is recommended to preserve storage media with the temperature at 20℃ and humidity at 45%-55%, especially avoid high temperatures.

Data on discs are stored near the printed side (writing side); scratches will prevent data from being read, please pay special attention.

(2) Digital Image Management

Computer technology allows real objects to be saved in an alternative method, but computer technology can be affected by operational mistakes, crashes, viruses, natural disasters (fire, flood, and earthquake), blackouts and hackers, resulting in abnormal information system operation or damaged data. Below are in-depth discussions of backup saving and remote backup. Comparisons of storage media will be dwelled on in chapter six.

a) Backup Saving[17]

Backing up personal data should be made a regular habit. After work everyday, data must be saved and backed up at the same time to avoid viruses or crashes causing data loss. Backup can be divided into three methods based on access characteristics: 1. On-Line Backup: uses hard drives as the storage media, provides fast reading and writing, emphasizes on data protection using RAID, but is the most expensive. 2. Near Online Backup: uses a jukebox as the storage media, can be accessed at any time, convenient for searching, high data guarantee, longer retention duration, and relatively cheaper. 3. Off Line Backup: uses magnetic tape for storage, is the most economic but provides slow access, and is a relatively less safe storage method.

A complete backup system should integrate the three storage methods above to make sure that data is perfectly safe. This not only requires RAID, jukebox and magnetic tape, but also needs operable storage management software.

[17] "Constructing Enterprise Storage Environments Making Money Go Further" by Lo Tai-Shan, from the website "Taiwan Telecom", Published: September 2005, Search: December 22nd, 2005, http://www.arconet.com.tw/taiwantelecom/tech/tech_1.asp?idxid=6683.

b) Remote Backup[18]

The key to remote backup is backing up required data in different locations sot that a second backup mechanism remains when the first set of data or system is damaged. Other factors that affect overall establishment cost include server grade, capacity of RAID, interface (SAN or SCSI), brand, and single tape drive and automatic jukebox. NT or Unix version software, different internet bandwidth, establishment cost should be in accordance with an organization's requirements.

Based on the project infrastructure of each unit, units must first have storage equipment that satisfy its digital archiving requirements, or seek a nearby environment that is capable of providing better service quality as a foundation for permanent archive preservation. The main bottleneck of the remote backup structure is the bandwidth between each unit to the remote backup location. Therefore, there should be a main backup location beside the remote backup location, and the infrastructure should be gradually improved year by year. Based on the infrastructure between holding units and remote backup locations, access speeds from will be different compared with the main backup location; this difference is a basis for choosing remote backup locations. Based on the main backup location's situation, data should be automatically accessed from either the main backup location or the remote backup location, which ever the faster, and save the burden of manual operation.

V. Notices on the Digitization Procedure

(1) Preliminary Operations

Concerning book condition inspection and book borrowing, only borrow books that are about to be digitized to prevent temperatures and humidity from damaging data. If a book is found to be in extremely poor

[18] Same as the above.

condition during book inspections, use other methods for digitization, such as digital photographing. In addition, for preliminary operations of string-bound book arrangement and picture mounting, it is recommended to ask specialists of this field to assist digitization operations, which will save manpower cost and reduce damage to books.

(2) Digitization Procedure-Field Choice

The sidebars of Chinese string-bound book fields may have special designs, the most common design is the left and right sidebars and the center of folio (where book pages are folded) have black mouth; the left and right sidebars are symmetrical. If one image file is used for one page of this design, the image will not be complete. Block-printed string-bound books have larger font size and will show clearly even if it is reprinted in a reduced format. However, books come in different sizes, using *Tong Shu* as an example, the book's width is 1/3 of A4 size and has small text, which is why string-bound book digitization requires special attention. Most Western books don't add sidebars to its text, which is why one image file is used for one page.

Furthermore, compared with block-printed string-bound books, modern books use relatively smaller font size and will be hard to read if they are reprinted in a reduced format. Therefore, field choice should be based on book type, page layout design, font size, and shouldn't be fixed. Examples below illustrate the effects of different image selection methods adopted by different string-bound books. In addition, Western books use different typesetting, so notice if the scanning order is consistent with the book's direction. Therefore, holding units should evaluate requirements, usage convenience and equipment limitations.

(3) Digital Image Backup and Storage[19]

Immense data quantity requires immense storage room, considering current computer technology development, storage room is not an issue,

[19] "Collection Development: Digital Library ppt Files" by Yeh Ching-Ling, May 30th, 2005.

but the retention duration of storage media is a topic that requires attention; for example, magnetic tapes have retention duration of 15 to 20 years, and the most durable CD-ROM has a lifespan of but 30 years. Therefore, continuous reproduction is the only way to maintain data stability. It is somewhat ironic to say that the magnetic tapes and CD-ROMs that people use in hopes of permanently preserving data have a far shorter lifespan than specially treated paper and microfilm. Moreover, these storage media are easily affected and damaged by external force, such as compute viruses, magnetic fields and electromagnetic pulses, and the degree of damage is no less than burning down a real library. Therefore, many issues remain in the path of long-term preservation.

Due the rapid development of technology, media transformation should be planned as a regular task while searching for storage media with longer lifespans. Although past research shows that DVDs can be preserved for 50 years, recent research has found that 2 to 5 years is its effective retention period. Therefore, it should be reasonable to transfer or backup digital image files every two years. Advance the scheduled backup time if DVDs become foggy within two years.[20]

[20] "A Study on the Quality of Digital Images of the National Archives" by Hsiang Chieh, Chen Hsueh-Hua and Wu Hai-Ju, Taipei: National Taiwan University Department of Computer Science and Information Engineering, October 30th, 2004.

Five. Metadata and Database Establishment[21]

I. Metadata

(1) Meaning of Metadata

Metadata is data about data. The establishment of metadata is to record characteristics of digital data, for example: a photo taken by a digital camera can have simple fields like photo name, theme, name of photographer, time, location, description...etc., these data fields are metadata of the photo. Metadata should achieve three purposes:

a) Describe the contents and features of digital archives.

b) Allow users to easily find data when searching online or in a computer.

c) Allow easy interflow and exchange of data with other databases. Of course, it should achieve different purposes based on different considerations, such as the database's purpose of establishment and users.

(2) String-bound Book Metadata Field Recommendations

Core metadata elements of rare books are divided into two categories based on their function: descriptive metadata and administrative metadata elements. Descriptions of the two categories are as follows:

Table 8. Metadata field recommendations of the National Digital Archives Program

Descriptive Metadata Elements (12 Elements in Total)	Type	
	Identification No.	Registration No., Shelf No., Call No.
	Title	Title Proper, Alternative, Roman
	Author	Name, Dynasty and Reign Title or Nationality, Role
	Contributor	Name, Dynasty and Reign Title or Nationality, Role
	Publication Information	Responsible Person, Location, Time, Responsibility
	Quantity	
	Description	Edition, Binding, Mount, Decoration, Lines Per Page, Tabooed Word, etc.
	Subject	Subject, Key Words
	Language	
	Relation	Series, Subtitle, Combined Issue, Original Document Copy, Image File
	Owner	

[21] "How to establish metadata for digitization databases: Using rare books digitization work as an example" by Chiang Jen-Chieh, Taipei: National Digital Archives Program Content Development Division, September 2005.

Administrative Metadata Elements (1 Element in Total)	File Records	File Staff, File Time, Revision Staff, Revision Time

II. Database Establishment

(1) Database Establishment

Database metadata can be roughly divided into two situations: 1) If there aren't many digitization objects, or if only simple noting and search functions are required, metadata can be very simple. 2) More exact and complete metadata can be established for academic and research purposes, using tens or even hundreds of fields to describe a single piece of data. Metadata established by units involved in the National Digital Archives Program Content Development Division are usually the second situation.

When establishing metadata, units usually consider their requirements, refer to a commonly used metadata standard, and revise the standard for their own use. That's why when exchanging data, although the corresponding field in different databases still needs to be found, the process is simpler.

Using the Rare Books Thematic Group of the National Digital Archives Program Content Development Division as an example, the "Academia Sinica Institute of History and Philology Fu Ssu Nien Library Rare Books" project referenced the Machine Readable Cataloging Record (MARC) Format, which is the most commonly used standard by libraries, and then added string-bound book characteristics to form its database metadata; the "Digitization of NCL Rare Books Collection" project referenced another frequently used metadata standard the Dublin Core.[22]

[22] See the official website http://dublincore.org/ for an introduction on Dublin Core, or refer to the traditional Chinese translation by Wu Cheng-Jui at http://dimes.lins.fju.edu.tw/dublin/, the simplified Chinese translation by the Shanghai Library at http://dc.library.sh.cn/, or "Reference Standard: Dublin Core" translated by the National Digital Archives Program Metadata Architecture and Application Team (MAAT), 2002, http://www.sinica.edu.tw/~metadata/standard/standard-frame.html.

(2) Database Establishment Process

Basically, file establishment is a daily task; files are proofread after being established. Two recommendations are proposed below. Use the original booklist system to create a list, transfer the main booklists into the metadata database at this time for more detailed image description and display. If fundamental fields have not been created in the original database, create corresponding fields of pages and image files. Due to the large table of contents in string-bound books, creation of detailed contents lists is required for the reader to acquire more detailed information. Establishment of booklist files and verification can be carried out after image files are inspected.

III. Challenges of Digitization Work

(1) Rare Chinese Characters

When establishing table of contents files, Chinese characters that rarely appear in modern life cause problems. The "Missing Chinese Characters Input Technology" was developed by the National Digital Archives Program Research & Development of Technology Division to handle rare Chinese characters. The character construction system developed by Academia Sinica Institute of Information Science Historical Documents Laboratory is adopted as the interchange code. Its structure follows Chinese character construction principles, breaking characters down into basic parts and then using these basic parts to combine tens of thousands of Chinese characters, solving the problem of insufficient interchange codes; the system currently provides basic parts of tubercle, regular script and variants.

(2) Time to Image File Output

Pay special attention to the verification of image files because of the large number of files. It is recommended to complete image files the day they are established in case too many waiting files are accumulated.

(3) High Employee Turnover Rate[23]

Due to the fact that part-time workers are hired to perform digital scanning, holding units should provide them with suitable training to ensure work quality because of their high turnover rate. Training should cover the purpose of digitization work, notices on original book arrangement and digital scanning, methods for screen calibration, and principles and procedures of quality verification. Utilize short-term condensed training to help new staff understand the importance of their work and rapidly get into the swing of things. Furthermore, when workflows or digitization equipment are changed, also pay attention to safety control and inspections of working staff. Verify a consistent standard for every step through teaching materials and practice operations. When a unit has high employee turnover rates, it is recommended to have special staff responsible for the overall planning of training matters and list string-bound book digitization as regular tasks.

(4) Immediacy of Standard Revision

No matter how well thought out workflows and standards are, or how many regular reviews and coordinations are carried out, standards and procedures may still be out of line with practice. Therefore, a unit must understand the situation of quality inspections and be in control of its digitization progress.

(5) Implementation of Project Management[24]

Systematic workflows benefit the improvement of digitization work efficiency and digital image quality. Preliminary preparation and planning requires careful consideration, and first conduct tests on small quantities to ensure nothing is left out.

a) Digitization Equipment

Based on the original book's size and preservation condition, the

[23] Same as 21.

[24] Same as 21.

digitization equipment used may be different. Notices and the digitization procedure may be slightly different for different digitization equipment.

b) Quality Verification

The digital image quality verification process should comprise of an initial verification, a second verification by special staff, and a program verification of whether or not the digital image can be normally opened and used. The first and second verification should compare all files with original documents one at a time to ensure digital image quality and avoid neglecting any detail. After verifying the digital image is correct, write the files to disc to avoid waste of manpower and resources.

c) Workplace Rules

Besides notices on digitization work and digital image quality verification, all workplace rules and original document usage rules influence work efficiency and protection of original documents.

d) Computerized System Control over Digitization Process

At present, many units still use paper forms for procedure control. Studies have found that computerized system control over scanning/digital photographing, quality verification, and CD-ROM/magnetic tape production makes the digitization process clearer, easier to analyze, and utilizes resources more effectively.

(6) Control over Related Technology

Digitization related scanning and data storage equipment have high elimination rates because of the rapid development of technology. Not only should units notice the retention duration of equipment, but also understand the current technology development situation to benefit equipment compatibility.

(7) Digitization Intellectual Property Rights Issues and Data Security Problems[25]

a) Difficulties Faced by Invisible Watermarks

i) Can't be added to black and white digital images. Even if it is added, it will disappear once the data is transferred.

ii) Some invisible watermarks can't process TIFF files.

iii)Can't be added to files that are too large (the invisible watermark developed by Academia Sinica can't be added files larger than 80MB).

iv) Adding invisible watermarks to the original digital image will damage it; invisible watermarks have been found to produce ripple effects.

b) Data Falsification

To prevent digital images from being falsified, add a date and time to the digital archive, or make two copies of the original digital image, one for permanent preservation and not to be borrowed, and the other to be provided for data transfer and further processing and use. Using this method, the copy that is not to be borrowed can be used for comparison when suspicions are raised over a digital image.

c) Intellectual Property Rights Issues

Before carrying out digitization work, make sure the image file is open for use and check if it has intellectual property rights. Even when dealing with historical data, many intellectual property rights, copyright and privacy issues remain to be handled.

[25] Same as 21.

Six. Equipment and Cost Analysis

This chapter roughly estimates the space, equipment (software and hardware) and cost required for digital scanning, and can be used as reference for planning digitization work.

I. String-bound Book Scanning Environment[26]

(1) Location

The scanning location should be near the center of a building and far from the influence of motor rooms, pipelines, water gates and light coming in from outside. The scanning location is recommended to be near the library, cargo elevator, string-bound book repair room, and research fellow's office. This way operation routes can be planned to benefit the convenience of the digitization work procedure.

(2) Space

In order for scanning work to be effectively carried out, prepare work space for working staff and safety space for management. Besides being used for placing computer equipment and scanning equipment, it is recommended to make room on the working table that is 7 times the size of the book about to be scanned to ensure there is enough room to place books that haven't been digitized and books that have completed digitization.

(3) Temporary Storage Space Temperature and Humidity Control

Standard temporary storage room temperature and relative humidity: 47%-55% relative humidity and temperature between 18-21℃. Scanning work requires sufficient air cooling and circulation devices to cool down the high temperatures produced from scanning. If no centralized control system exists, the cooling system of air conditioners can be used to filter

[26] *Text Archives Digitization Procedures Guideline* (First Draft) by Lin Yen-Hung, Taipei: National Digital Archives Program Content Development Division, September 16th, 2005, page 38.

impurities in the air.

(4) Lighting

Requirements on lighting are relatively high, besides suitable brightness, a comfortable and pleasurable atmosphere is even more important. Therefore, when considering good lighting, one must first understand color temperature, color rendering and economic efficiency. Using ultra violet rays as an example, its brightness, nature and wave length will damage string-bound books or other collections, which is why putting books under direct sunlight should be avoided; close the windows and use dark colored curtains to block sunlight. Scanning locations mostly use florescent tubes and light bulbs as the main light source, so it is recommended to cover books with materials that absorb ultra violet rays. Avoid the use of white light bulbs, they produce heat and cause partial humidity fluctuations, affecting temperature control.

(5) Other Requirements

In order to prevent pests from damaging string-bound books, maintain cleanness of environment, bringing food and drink into the workplace is strictly prohibited. To prevent books from dropping, or being damaged by natural disasters like floods, placing books on the floor is prohibited. In addition, books can be placed in shelves, but must be 18cm from the ground. Placing books above 18cm from the ground protects them from potential damage from water, dust and dirt, and avoids books being misplaced in inappropriate places by cleaning staff.

II. Main Digitization Scanning Equipment

(1) Scanning Equipment

Most string-bound books are digitized using platform scanners and flatbed scanners.

a) Platform Scanner

This scanner faces books downwards when scanning. Not covering

the blackout panel is recommended when digitizing string-bound books. Every page scanned using this type of scanner means turning the whole book once. Therefore, this type of scanner is not suitable for books in bad condition; this type of scanner is suitable for data of identical size and books in good condition.

b) Flatbed Scanner

This scanner faces books upwards when scanning. Therefore, it doesn't require turning the whole book when finished with one page, only the pages need to be turned. The light source is placed high above and the scanner scans images from the upper side; therefore, this type of scanner is suitable for string-bound book digitization. However, be careful of the books bindings on the back to avoid inappropriate machine operation from damaging the books.

Table 9. Comparison of digital scanning equipment

Scanner	Definition and Features	
Platform Scanner	Definition	Generally scans A4 or A3 sized pages.
	Features	Cheap and small
Flatbed Scanner	Definition	Generally scans A1 or A0 sized pages.
	Features	Reduces contact with books and images that aren't straight Relatively expensive, rarely purchased by units, and more often out-sourced.

(2) Comparison of Flatbed Scanners

Looking back on equipment and techniques of the past, digitization output was the utmost consideration and less attention was paid to artifact protection. This resulted in original documents being damaged or destroyed by equipment (the scanner's centrifugal force was too strong or light was too hot), storage space (too dark and humid) or human factors (not using suitable gloves when scanning). Today, consideration can be given to both digitization output and book preservation, thanks to the development of mechanical equipment (such as book scanners,

professional multi-purpose scanners) specially designed for string-bound book digitization, reducing the amount of damage to original books during digitization.

a) Equipment Type

Scanners currently in the market support string-bound book digitization without damaging the original book. Its process doesn't require touching books or removing its bindings. Scanners use a 180 degree book cradle (Figure 9) or 120 degree book cradle (Figure 10) to support and balance both sides of the book.

Figure 9. 180 degree book cradle
(Image: Fu Ssu Nien Library's Book Photographing Platform)

Figure 10. 120 degree book cradle
(Image: Zeutschel (Germany) Flatbed Scanner)

b) Scanning Method

To solve the issue of left and right pages being symmetrical, the current method used for scanning is to scan both pages into one image file. At present, most scanners use 180 degree book cradles, in which the German brand Zeutschel is most frequently used. This scanner uses glass instead of hands to flatten books, achieving extremely clear images after digitization, but when books are bound too close to its text and are flattened, they will be damaged at their weakest point the bindings. Therefore, consider the book's condition before deciding on the equipment to use.

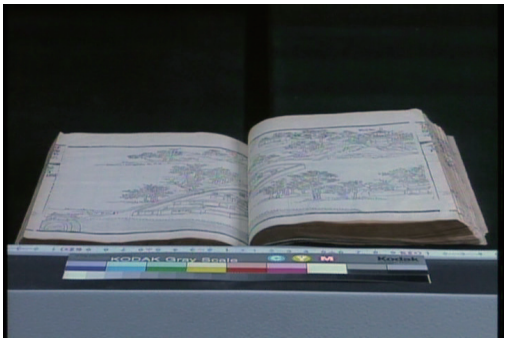


Figure 11. 180 degree book cradle (no glass)
(Image: Fu Ssu Nien Library's Book Photographing Platform)

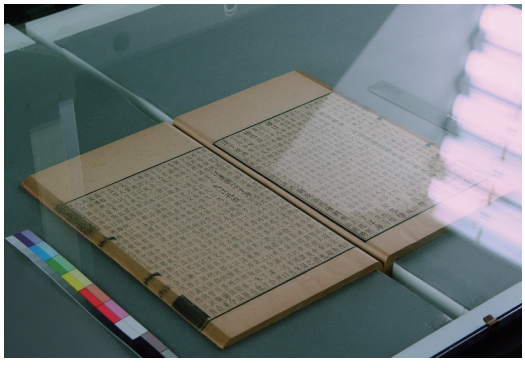


Figure 12. 180 degree book cradle (with glass)
(Image: Fu Ssu Nien Library's Book Photographing Platform)

c) Book Positioning

With the premise of not dividing ancient books according to language, Western books and string-bound books all have different thickness and size. Below is a comparison of positioning two different kinds of books. The picture shows that the Western book has a thicker book back, the platform has been moved to provide a greater height difference, and pages are harder to turn, showing how difficult it is to keep Western books in place. String-bound books are thinner and their pages are easier to turn, which is why the platform uses glass instead of hands.

Figure 13. 180 degree book cradle (no glass)
(Image: Fu Ssu Nien Library's Book Photographing Platform)

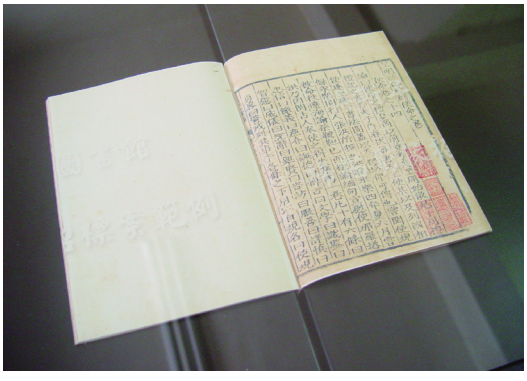


Figure 14. 180 degree book cradle (with glass)
(Image: Fu Ssu Nien Library's Book Photographing Platform)

d) Platform Comparison

The performance of book scanning equipment common in Taiwan is compared below.

Table 10. Comparison of digital scanning platforms

	Book Photographing Platform	Zeutschel (Germany)	Digi Book professional multi-purpose
Type Comparison	Domestically developed	6000	2000LC
Image			
Price	NT$300 thousand	NT$1.6-3.5 million	NT$4.5-6 million
Underlay color	Black lint (to prevent light reflection)	White	Green
Movable Panel	Magnetic panel that moves up and down	Moves up and down	Moves up and down
Micro Adjustment Method	Step motor	Electronic	Manual
Book Fixing	Liftable glass	Fixed glass	None
Book positioning and distance to glass	No contact	Tight	Hand pressed
Largest Photograph Range (On the scale of 1:1)	86cm×105cm (A0 and above)	A1 size	A1 size
Platform center of folio width	>5cm	<5cm	<5cm
Center of folio movement	More Flexible	Standard	Standard
Objects	Books, artifacts, sculptures, periodicals, newspapers, files	Books, periodicals, newspapers, files	Books, periodicals, newspapers, files, maps
Rating	Semi-automatic	Fully-automatic	Semi-automatic
Notes	Separate purchase of lights and digital camera is required	No separate purchase required	No separate purchase required

(3) Book Condition and Equipment Selection

Besides considering project budget, book condition should also be considered for equipment selection. Books are divided into three states below, including "Good condition identical size books", "Bad condition different size books" and "Extremely bad condition different size books".

Table 11. Relationship between book condition and equipment selection

Recommended Digitization Equipment Original Book Condition	Function	Explanation
Good condition Identical size Books	Scanner	Platform Scanner, Flatbed Scanner
	Features	Relatively cheap, small, numerous scanners can be purchased at once for use, and provides good scanning quality.
	Disadvantages	Will easily produce images that aren't straight. Books frequently come in contact with equipment.
	Notes	These types of scanners are used by many units, including Academia Sinica's Institute of History and Philology and Institute of Modern History.
Bad condition Different size Books	Scanner	Flatbed Scanner
	Features	Reduces contact with books and images not being straight.
	Disadvantages	Relatively expensive and rarely purchased by units, most digitization work is out-sourced.
	Notes	The Tan-Hsin-Tang-An completed by the National Taiwan University Library and digitization work carried out by other units, such as the National Central Library and the National Palace Museum.
Extremely bad condition Different size Books	Scanner	Book photographing platform.
	Features	Requires minimum contact with books, the platform can be expanded to hold objects of different size, and reduces the chances of images not being straight.
	Disadvantages	This equipment currently doesn't exist in the market
	Notes	Academia Sinica Institute of History and Philology Fu Ssu Nien Library developed its own platform in collaboration with firms.

III. Main Digitization Storage Media

Main digitization storage media is discussed in three sections, including an introduction of "hardware storage equipment types", "other factors that affect storage media" and "comparison of frequently used hardware storage media".

(1) Hardware Storage Equipment Types[27]

According to a prior discussion on access characteristics, storage media can be divided into three types: 1. On-Line Backup: uses hard drives as the storage media, provides fast reading and writing, emphasizes on data protection using RAID, but is the most expensive. 2. Near Online Backup: uses a jukebox as the storage media, can be accessed at any time, convenient for searching, high data guarantee, longer retention duration, and relatively cheap. 3. Off Line Backup: uses magnetic tape for storage, is the most economic but provides slow access, and is a relatively less safe storage method.

Table 12. Storage media divided according to access characteristics

Type (Based on storage characteristic)	Function	Advantage	Disadvantag
Magnetic Tape Disc (CD, DVD, MO)	Data is saved in a safe place after being written, suitable for long-term or permanent preservation, this type of storage media is cheap, but has relatively worse data immediacy and is hard to modify.	Storage media has low prices, high data preservation ability, can be used for long-term data backup.	Requires manual change, data is not immediate and relatively slow to access, not suitable for use by servers.
Jukebox (CD-ROM, DVD-ROM, DVD-RW, MO, DVD-RAM, DVD+RW, etc.)	Automatically changes storage media, data is less immediate because of its many mechanical structure and problems are easier to occur, but saves large amount of data and is frequently use for the preservation of historical data.	Uses Off-Line Storage and automatic changing mechanism, storage media has low prices and high data preservation ability.	Many mechanical actions for automatic changing, high fault rate, the jukebox is relatively expensive, and data transfer is not immediate.

[27] "Technical Articles: High Capacity Storage Equipment" from the website "ASSC", Search: December 22nd, 2005, http://www.assc.com.tw/tech/article.asp.

Type (Based on storage characteristic)	Function	Advantage	Disadvantag
Hard Drive or RAID	Provides immediate and rapid data access, most are collocated with RAID controllers to achieve data security, but equipment is relatively more expensive.	Provides immediate, rapid and stable data access.	Relatively higher price, capacity is wasted when using RAID.

(2) Other Factors that Affect Storage Media[28]

a) Data storage equipment

b) Capacity

Data storage capacity is calculated using Byte (which is 8 bits) as the unit; 1,024 Bytes = 1M bytes, 1,024M bytes = 1G bytes, 1,024G bytes = 1T bytes.

c) Transfer Speed

Data transfer speed of storage equipment uses data quantity transferred "per second" as its unit; an example of transfer speed is 80MB/Sec, which means "80M bytes can be transferred per second". The specification on labels is the highest transfer speed of the interface and not the actual transfer speed. For example, a hard drive that is labeled ATA-100 means that it uses a 100MB/Sec transfer interface, but the real access speed is around 50MB/Sec. In order to fully use the bandwidth provided by an interface, multiple hard drives must be connected to the same channel.

[28] Same as 16.

Table 13. Information on other factors that influence storage media

Storage Media	Data Storage Equipment	Common Interfaces
Tape Storage	DAT, DLT, SuperDLT, Ultrium	SCSI
Compact Disk	LTO CD-ROM, CD-R, CD-R/W, DVD-ROM, DVD-RAM, DVD-R/W, DVD+R/W	IDE / SCSI IDE / SCSI
Magneto-optica	MO 540, 640, 1.3GB....	IDE / SCSI
Juke-box	CD / DVD / MO Juke box	
Disk Array	SCSI / Fiber / IEEE-1394 Disk Array / SAN / NAS	
Hard Disk	IDE HDD, SCSI HDD, Fiber HDD	IDE / SCSI / Fiber
Flash Memory	Compact Flash / Multi Media Card / Memory Stick / Smart Media Card	Specialized Interface
Removable Media	Floppy / LS-120 / ZIP	FDD / IDE / SCSI / USB

(3) Comparison of Frequently Used Hardware Storage Media[29]

Storage equipment is a place that computers use to store or access large numbers of files, programs and data. Different storage equipment has different capacity, access speed, access method, function and service life. Below is a comparison of common storage media for digital files provided by a research report on digital image quality of national archives.

[29] Same as 21.

Table 14. Comparison of common storage media for digital files

Storage Media	Capacity	Type Suitable to Become Storage Disc	Advantages	Disadvantages	Service Life
CD-R	650MB 700MB	CD-R CD-RW	1. High storage density 2. Possesses liquidity 3. Easy for searches and long-distance transfer 4. Easy to duplicate 5. Good restoration effects 6. Wide range of application, can save pictures and text	1. Recording speed isn't as fast as microfilm 2. Higher fault rate than microfilm when recording 3. Uncertain service life that is far shorter than microfilm 4. Unified international standard on disc technology yet to be established	1. CD-R: 2 years 2. CD-ROM: 5 years (requires normal room temperature and relative humidity under 90%)
DVD	Roughly the capacity of 7 CDs	1. DVD-ROM (Read-only Digital Multi-function Disc) 2. DVD-R (Writable Digital Multi-function Disc) 3. DVD-RW / DVD+RW (Re-writable Digital Multi-function Disc)	Same as CD-R	Same as CD-R	2-5 years
Magnetic Tape	200GB/ 400GB		Suitable for saving large quantities of digital data	1. Longer read time, each read must start from the beginning 2. If the tape is old or has poor quality, damaged sectors will not only affect backup speed, but also reduce usable capacity	2-5 years
Hard Drive	40-200G		The simplest way to preserve electronic files, usability of electronic files can be maintained as long as the computer system is preserved	1. Easily damaged; data easily lost 2. Requires large storage space 3. High maintenance cost	2-5 years

IV. Cost Analysis

(1) Cost Composition

Cost of digitization by scanning composes of three aspects: material expenses, labor expenses, and expenses:

a) Material expenses are mainly expenses on consumables used during work.

b) Labor expenses are mainly the salaries of staff.

c) Expenses can be divided into direct expenses and indirect expenses:

i) Direct expenses: Purchase fee and depreciation of information equipment and scanners and fees for information software.

ii) Indirect expenses: This can be divided into scanning space and system space. Expenses on scanning space include depreciation, rental, repair, insurance, utilities and other. Expenses on system space include system preparation and maintenance.

(2) Cost Estimation

Due to limited data, the cost analysis of this guideline is only a rough estimate of image scanning cost based on material expenses, labor expenses, direct expenses, and system preparation and maintenance expenses.

a) Calculation Method: This can be divided into two methods based on equipment amortization.

i) Service Life:

Table 15. Cost Estimation–Service Life

Definition	Material Expenses	DVDR
	Labor Expenses	Salary
	Equipment Amortization Expenses	(Scanning Equipment Software/ Hardware + System Preparation Software/ Hardware)/Service Life
Formula	[Material Expenses (NT$)/Digital Output Quantity (pages)] + [Labor Expenses (NT$) + Equipment Amortization Expenses (NT$)/Digital Output Quantity (pages)]=Cost per page (NT$/page)	

ii) Digital Output:

Table 16. Cost Estimation–Digital Output

Definition	Material Expenses	DVDR
	Labor Expenses	Salary
	Equipment Amortization Expenses	(Scanning Equipment Software/ Hardware + System Preparation Software/ Hardware)/Service Life
Formula	Material Expenses (NT$)/Digital Output Quantity (pages) + [Labor Expenses (NT$)/Digital Output Quantity (pages) + (Scanning Equipment Software/ Hardware + System Preparation Software/Hardware)/ Digital Output Quantity (pages)] = Cost per page (NT$/page)	

b) Calculation Example:

The cost for each page will vary with the equipment used. From the description above we know that basic cost estimation includes information on computer equipment, software/hardware scanning equipment, related products, system development and preparation, manpower, files size and equipment service life. Cost required for different scanning platforms is estimated in the four appendixes below, showing the different costs resulting from use of different equipment.

c) Recommendations

The cost calculation of this guideline is only an initial evaluation, and main considerations are limited to equipment and human resources. However, this simple formula shows major factors of cost control:

i) Human Resources:

In order to reduce labor expenses, professional part-time workers can be trained to complete scanning operations because they have fixed procedures. Image verification, however, requires image processing specialists. It is recommended that special staff be hired for image verification to ensure image quality.

ii) Software/Hardware:

Although hardware is expensive, if it increases digitization output speed, which means reducing labor expenses, it doesn't necessarily increase the total cost. From another point of view, if using low performance equipment results in reduced digitization output and increased labor expenses, the total cost won't necessarily be reduced. Furthermore, although developing software is expensive, it has the following advantages:

A. Can control cost: Reduces labor expenses.

B. Guarantees production: Enhances productivity.

C. Unified quality control: The process is easy to control because of its transparency.

D. Software can be integrated for utilization.

Seven. Benefits

A comparison of digitization achievements in the Taiwan area and China area are listed below for readers to view another kind of benefits analysis.

Table 17. Current achievements of different organizations

Comparison of Domestic and Foreign Digitization Achievements					
	Taiwan			China	
	Digitization System	Digitization Items		Digitization System	Digitization Items
Fu Ssu Nien Library	1. Image Scanning 2. Text Identification 3. Full Text Search 4. Geographic Information System 5. Cross Database Integration	1.Rare books bibliographic database 2.Reorganize file catalog of the Institute of History and Philology 3.Collection seal database 4.Rare books image search system 5.Name authority search system 6.Spatial information system 7.Taiwan Public and Private Ancient Archival document Collection Image Database 8.Digitization management system (scanning and proofreading system)	National Library of China	1. Features of Local Chronicles 2. Image Scanning 3.Text Identification 4. Full Text Search and Standard Control 5. Special Topic Database 6. Geographic Information System 7. Cross Database Search, Resource Integration	1.Digital local chronicles 2.Bei Tie Jing Hua 3.Dun Huang Yi Jhen 4.Si Sia Suei Jin 5.Jia Gu Shih Jie 6.Yong Le Da Dian 7.China Research Resource Database 8.Bibliographic Database
National Central Library	1. Image Scanning 2. Text Identification 3. Full Text Search 4. Bibliographic Management System	1.Ancient book image search system 2.Online special collection exhibition hall - e-books, Introduction to Chinese books...etc. 3.Taiwan Local Documents Database			

Comparison of Domestic and Foreign Digitization Achievements					
	Taiwan			China	
	Digitization System	Digitization Items		Digitization System	Digitization Items
National Place Museum	1. Rare Books Bibliographic Management System 2. Full Text Search 3. Image Scanning 4. Text Identification	1.Rare books catalog database 2."Han Cyuan" Classic Document Full Text Search Database 3.Genealogical tree document database 4.National Palace Museum Buddhist Scripture Illustrations Index 5.Overseas treasures - search of non-collection relic images			

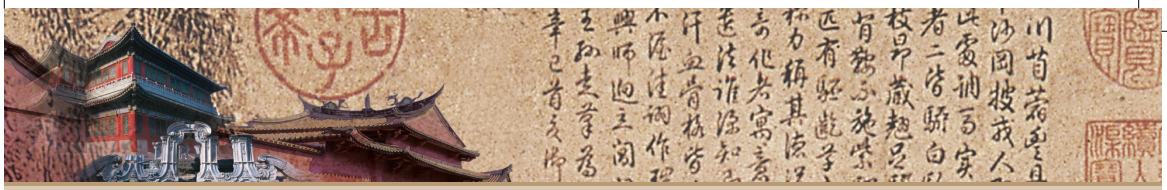
Eight. Future Prospects and Conclusion

Virtual libraries are no longer just a slogan, thanks to development of the internet. Google's Content Partnerships Director Jim Gerber once proclaimed that "In the future, the only thing that will get read is something that will be online. If it isn't online, it doesn't exist", showing the urgent need of modern people for a virtual environment of the future. Collectors and libraries have always existed in Chinese traditional culture, but they all emphasized on "collecting", collecting books into their libraries, and thus creating the contradiction between collecting and using string-bound books. In concepts of the Western culture, Western collectors and scholars collect books not only for their individual collection, but also commercial benefits, donations, research and serving others, which boosted a far wider spread of books than in China.

Book preservation is difficult, but the contradiction between collecting and using was solved with the concept of digital archives and development of information technology. Ms. Lin Miao-Hua of the Academia Sinica Institute of History and Philology Fu Ssu Nien Library indicated that digitization and network communication will effectively put an end to the world's information gap, digital gap and knowledge gap. Since scholars insist on seeing with their own eyes before they are willing to believe, Ms. Wu Pi-Yung of the National Palace Museum Dept. of Rare Books and Document recommends combining digital images and text to provide mass quantities of research resources and allow the special publication of real books through database establishment, such as illustrations in antique books and artwork, integrating digitization achievements into real life. Chen Li, Deputy Director of the National Library of China once said

that international cooperation in database construction, avoid repeated or closed databases, and establish a unified international standard are tasks that need to be aggressively carried out at the current stage.

In addition, in consideration of scanning platform size limitations during the digitization procedure, the National Palace Museum Dept. of Rare Books and Documents gives priority to books in good condition and smaller than A0 size, which is why many books remain in its storehouse yet to be digitized; the National Palace Museum's collection is mostly books read by emperors of China. Academia Sinica Institute of History and Philology developed a specialized book photographing platform in response to the scanning size problem, solving a fundamental difficulty of the digitization procedure. Although digitization related technology still requires working on, the purpose of digital archives is to provide value-added knowledge of archives, and in a world that will be led by knowledge, it will be like the words of an internal manager of MSN "physical contents are believable, and they are where answers to people's questions exist". Therefore, book digitization is only the first step.



【Appendix 1】References

I. Specialized Books

1. *Digitization Procedures* book series by National Digital Archives Program Content Development Division et al, Taipei: Content Development Division, January 2005, Third Edition.
2. *Tasks and Techniques for Digitization of Documents*. by Hung Shu-Fen, Taipei: National Digital Archives Program Training & Promotion Division, February 2004, First Edition.
3. *Digital Archives Technology Collection*. by Hsieh Ying-Chun et al, Taipei: National Digital Archives Program Program Office, November 2004.
4. *21st Century Library of Congress Digital Strategy.* by Chiang Chuan-Ming, translated by Yuen Ke-Li, China: National Library of China, February 2004, First Edition.
5. *Conservation Concerns.* by Konstanze Bachmann, translated by Liu Lan-Yu, Taipei: Five Senses Art Management, 2001, First Edition.
6. *Digital Image Color Managemen.* by Hsu Fa-Kun, Taipei: Lung Chen, July 1999, First Edition.
7. *Mastering Digital Printing.* by Harald Johnson, translated by Chiu Fang-Shen, Taipei: Flag Publishing Co., February 2005, First Edition.

II. Periodicals

1. "New Impression of Digital Content of Rare Books" by Lin Miao-Hua, collected in the *Collection Digitization Work Conference Paper Serie*, July 2005, pages 57-70.

2. Huang Kuan-Chung et al, collected in the *2003 Ancient Books Digital Archives Seminar Paper Serie*, September 2003.
3. Chen Kuo-Tung et al, collected in the *2005 Chinese Ancient Books Digitization Cooperation Establishment Seminar Paper Series* , September 2005.
4. "Exploration of Digital Archives Establishment SOP" by Huang Ju-Chu and Mei Shih-Chieh, collected in *The Second Digital Archives Technology Seminar Paper Series*, 2003.
5. "Digitization Process and Future Development Direction of Chinese Books, Documents and Data in Taiwan" by Lo Feng-Chu, 2000.
6. "Retrospection and Foresight of Taiwanese Literature Digitization" by Lo Feng-Chu et al, collected in the *2004 International Academic Conference on Chinese Studies Paper Serie*, October 2004.
7. "Current Status and Prospects of Ancient Books Digitization in Taiwan" by Lo Feng-Chu, collected in *The Third Cross-strait Ancient Books Collation Research Conference Paper Series*, April 2001.
8. "An Exploration of the Digital Library and the Organization of Ancient Books" by Ku Li-Jen, *Journal of Library and Information Studies*, Volume 14, December 1999, pages 115-128.
9. "Summary of the International Conference on Sinological Resources in the Digital Era" by Chen Chao-Chen, collected in the *Newsletter for Research in Chinese Studies*, February 2005, pages 24-30.
10. "Applications of Digitization Technologies of American Libraries (A)" by Yang Hsiao-Wen, collected in the *Bulletin of NCL Taiwan Branch*, Volume 6 Issue 4, May 2005, pages 58-67.
11. "Applications of Digitization Technologies of American Libraries

(B)" by Yang Hsiao-Wen, collected in the *Bulletin of NCL Taiwan Branch*, Volume 6 Issue 5 , May 2005, pages 40-56.

12. "Newsletter for Research in Chinese Studies 20 : 3", collected in the *Newsletter for Research in Chinese Studies* Issue 79, August 2001, page 97.

III. Websites

(1) String-bound book definition and digitization procedure websites

1.SOP Definition: http://atj.org.tw/newscon1.asp?page=prev255

2.Academia Sinica Computing Center: http://www.ascc.sinica.edu.tw/

3.Academia Sinica Institute of History and Philology: http://www.ihp.sinica.edu.tw/index_page.htm

4.National Palace Museum: http://www.npm.gov.tw/ch/a040101

5.Fu Ssu Nien Library: http://lib.ihp.sinica.edu.tw/

6.National Digital Archives Program: http://content.ndap.org.tw/main/index.php

7.National Central Library: http://www.ncl.edu.tw/

8.National Library of China: http://www.nlc.gov.cn/old/index.htm

9.Taiwan Archives Online: http://ithda.ith.sinica.edu.tw/

10.Hong Kong Public Libraries: http://www.hkpl.gov.hk/

11.China Digital Library: http://www.d-library.com.cn/index.jsp

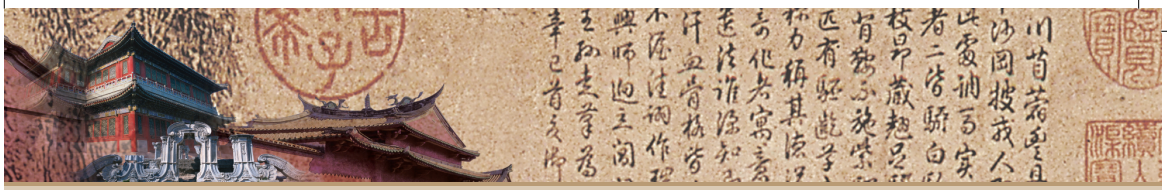
12.Wun Jin Liou Shang: http://www.nlc.gov.cn/publish/wjls/html/index.htm

13.Bibliography related website: http://catweb.ncl.edu.tw/sect-2.htm

(2) Related equipment websites

1.Avision scanner related website: http://www.591pc.com/avision.htm

2.Taiwan Imaging Systems, Inc.: http://www.tisi.com.tw/scanners.htm
3.Flatbed scanner related website: http://www.digital-center.pl/omni6000.htm



【Appendix 2】Images Production Specification Reference Standards

I. Guideline of the National Library of China

This guideline defines the digitization class, color requirement, preservation format and resolution for each type of resource, below is an example of the digitization standard for rare books.

Table 1. The National Library of China

Category			Rare Books		
Type			Original Books		
Purpose			Collection, Detailed Printing, Online Browse		
Color Requirement			24-bit color		
Class	A (Archives)	P (Duplication)	L (Browse-Limited)	MBrowse-	S (Reduction)
Format	TIFF	JPEG	JPEG	Limited) JPEG	GIF
Minimum Resolution	600DPI	600DPI	300DPI	150DPI	72DPI
Notes	Pages are scanned facing upwards, which means flatbed scanners are used. Requirements on general ancient books are referred to if OCR is needed.				

II. Guidelines of the Institues in Taiwan

(1) National Digital Archives Program Guideline

Table 2. National Digital Archives Program

National Digital Archives Program			
Digitization Purpose	Archives Class	Business Class	Browse Class
Scale	1 : 1		
File Format	TIFF	JPEG	JPEG
Color Mode	RGB(24bit/piexl)	RGB(24bit/piexl)	RGB(24bit/piexl)
Resolution and Size	Original size 300dpi and above	Original size 300dpi	72 dpi

National Palace Museum Dept. of Rare Books and Documents[30]			
Book Status		Book < A3	Book > A3 (including A3)
Scale		1 : 1	
Field Choice		The cover is one image Each pair of left and right bifolios is one image	The cover is one image Images are connected (left and right bifolios are photographed individually)
File Format		TIFF	TIFF
Color Mode		RGB (24bit/pixel)	RGB (24bit/pixel)
Resolution and Size		Original size 300dpi	Original size 300dpi
Data Backup		DVD, DAT magnetic tape	

Academia Sinica Institute of History and Philology Fu Ssu Nien Library[31]			
Book Status		Content is illustrations	Content is text
Scale		1 : 1	
Field Choice		The cover is one image. Each pair of left and right bifolios is one image.	
File Format		TIFF	TIFF
Color Mode		RGB (24bit/pixel)	RGB (24bit/pixel)
Resolution and Size		Original size 600dpi	Original size 300dpi and above
Compressed Files	File Format	JPEG	JPEG
	Color Mode	RGB (24bit/pixel)	RGB (24bit/pixel)
	Resolution and Size	144 dpi	72 dpi
Data Backup		DVD	DVD

[30] "National Palace Museum Dept. of Rare Books and Documents 2004 Rare Books Digitization Image Production Specification" by and from the National Palace Museum, 2004 Revised Edition.

[31] "Academia Sinica Institute of History and Philology Fu Ssu Nien Library Rare Books Digitization Image Production Specification" by and from the Academia Sinica Institute of History and Philology, April 2005 Revised Edition.

National Central Library				
Book Status		If the text is too small, lines are too thin, or other special requirements	Chinese Rare Book Size B4-A3	General String-bound Book Size B4-A3
Scale		1 : 1		
Field Choice		The cover is one image. Each pair of left and right bifolios is one image.		
File Format		TIFF	TIFF	TIFF
Color Mode		RGB(24bit/piexl)	RGB(24bit/piexl)	Black and White
Resolution and Size		Original size 600dpi and above	Original size 300dpi and above	Original size 300dpi and above
Compressed File	File Format		JPEG	TIFF
	Color Mode		RGB (24bit/pixel)	Black and White
	Resdution and size		Original size A4 72dpi (85%)	72dpi (G4)
Data Backup		DVD		

(2) National Repository of Cultural Heritage Archives Guideline

Table 3. National Repository of Cultural Heritage Archive

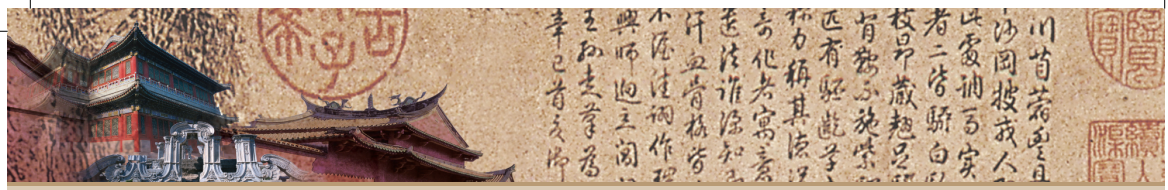
National Repository of Cultural Heritage Archives		
Digitization Purpose	Permanent Preservation	Online Browse Format
File Format	TIFF	JPEG
Color Mode	RGB	RGB
Color Depth	Color 24bit/pixel	Color 24bit/pixel
Resolution and Size	600dpi and above	300dpi 500×400 to 1000×700 pixels
Compression Quality	Uncompressed	75%

National Taiwan University Library				
Digitization Purpose	Archives Class	E-commerce Class	Public Information Class (Display, Print)	Public Information Class (Display)
File Format	TIFF	JPEG	JPEG	JPEG
Color Mode	RGB			
Color Depth	Color 24bit/pixel			
Resolution and Size	300dpi	300dpi	150dpi	75dpi
Data Backup	DVD, RAID			

III. Cornell University Digital Library Image Scanning Archives Guideline

Table 4. Cornell University Digital Library Image Scanning Archives

Document Type		Resolution	Color Depth	Permitted Editing and Processing	File Format
Recommended Standard	Printed document	600dpi	Black and White	Sharpen, Cut,	TIFF5/6
	Image files with artwork	400dpi	8-bit grayscale 24-bit color/black and white	Slant Fixed ratio expansion, minimum color and hues adjustment	TIFF5/6
	Precious or damaged printed documents	400/600dpi			
	Manuscript	300/500dpi			
	Art works expressed on paper	400dpi	24-bit color		
Minimum Standard	Printed document	300-400dpi	Black and White	Sharpen, Cut, Slant	TIFF4/5/6 JPEG,JFIF
		200dpi	8-bit grayscale	Minimum color and hues adjustment	
	Image files with artwork	300dpi	8-bit grayscale	Fixed ratio expansion, minimum color and hues adjustment	TIFF4/5/6 JPEG,JFIF
		600dpi	Black	Sharpen, Cut, Slant	
	Precious or damaged printed documents	300dpi	8-bit grayscale 24-bit color/black and white	Fixed ratio expansion, minimum color and hues adjustment Sharpen, Cut, Slant	TIFF4/5/6 JPEG,JFIF KPCD
		600dpi			
	Manuscript	200dpi	8-bit grayscale 24-bit color/black and white	Fixed ratio expansion, minimum color and hues adjustment	TIFF4/5/6 JPEG,JFIF
		400/600dpi			
	Art works expressed on paper	300dpi	8-bit grayscale 24-bit color	Fixed ratio expansion, color correction, sharpen	TIFF4/5/6 JPEG,JFIF KPCD



【Appendix 3】Comparison of Current Remote Backup Equipment of Different Units

I. Academia Sinica[32]

(1) Storage Equipment

Each node is a storage system constructed from storage bricks that are disk arrays 10TB in capacity, storing data produced in the recent year and digitization data that is required for online real-time access services. All data is saved in duplicates online (one copy belongs to the local node and the other copy belongs to another node online, achieving remote backup within the institute); the Computing Center produces a separate backup copy and saves it in a magnetic disc for permanent preservation. Data that is one year or older is saved on magnetic tape in the Computing Center's mass data storage system.

(2) Remote Backup

Each node must help provide the room and environment for placing related equipment, all equipment are in 19" machine cabinets (Width x Depth x Height is approximately 60x100x206cm, the front and back of machine cabinets should be at least 100cm from other equipment or walls for heat ventilation and work maintenance), machine cabinets include the following equipment and have left room for expansions.

a) Unix server * 1

b) Gigabit Ethernet Network

c) 5TB Disk Array * 1; Disks are 250GB SATA Disk

d) UPS * 1; Capacity 3000 VA

In addition, whether or not different units should use the same architecture and same technology/system should be taken into account. Using the same system allows direct federation and resource sharing and backup is also the easiest, but is not absolutely necessary.

[32] "Academia Sinica Remote Backup Storage Data Network Preparation ppt file" by Academia Sinica, from Academia Sinica's 2003 midyear auditing conference, October 2005.

(3) Structure

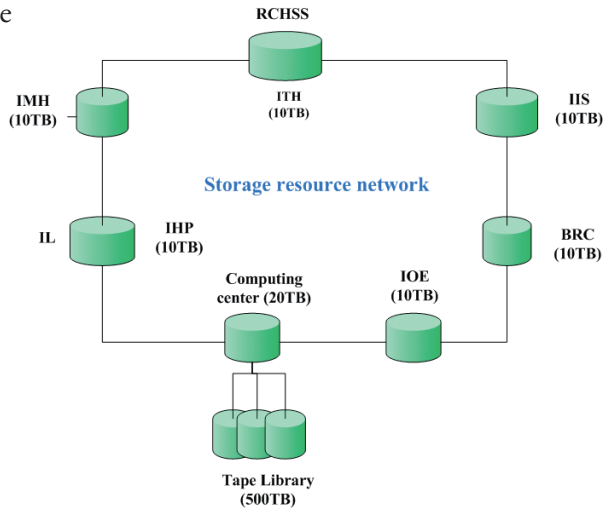


Figure 1. Storage Resources Network Structure

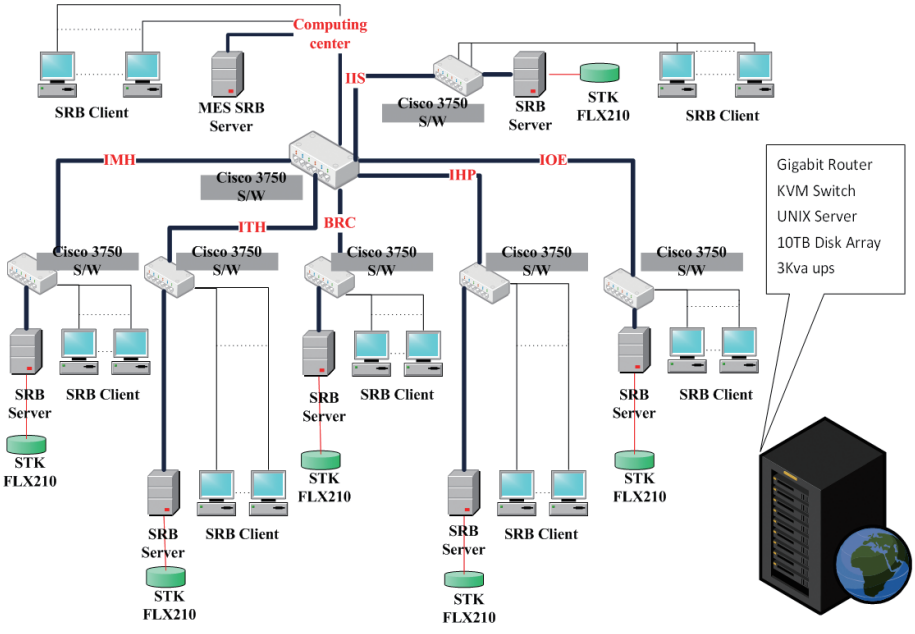


Figure 2. Deployment of DataGrid for Digital Archives in Academia Sinica

II. National Palace Museum[33]

Lin Kuo-Ping, NPM researcher and Information Center Director, indicated that the National Palace Museum has currently completed the establishment of NAS (Network Attached Storage) and SAN (Storage Area Network), and it has completed the purchase of Fiber-Channel Hard Disks with a total capacity of 4TB for on-line storage; the National Palace Museum will continue to purchase SATA hard disks with a total capacity of 6TB this year for near-line storage. In other words, by the end of this year, the National Palace Museum will have 10TB of storage space for saving digital archives.

NAS mainly adopts the Microsoft WSS 2003 Architecture, supports open systems, and is used for sharing external files. The establishment of SAN allows users to integrate the complicated storage environment, and achieves automated management of storage equipment via user strategies and settings. Lin Kuo-Ping pointed out that digital files are transferred via a dedicated optic fiber line, which doesn't interfere with the bandwidth of the regular system and achieves transfer security. SAN and NAS coexist in the current storage architecture. The cluster architecture is adopted for hosts and servers, achieving high system availability and can be used for backup; two sets of servers and operating systems are used, when one of the digital archives systems shuts down randomly, the other system is ready to take over.

(1) Storage Equipment

Digitization is a process that requires strenuous efforts to complete. In

[33] "Artifact Digitization, NPM from A to A+" by Huang Yen-Fen, from the website "iThom", Published : August 23rd, 2005, Search: December 22nd, 2005, http://www.ithome.com.tw/itadm/news/news.php?c=32667.

order to let every digital file be preserved in a detailed and complete state, the National Palace Museum makes regular backups and uses multiple storage methods for its digital archives. Lin Kuo-Ping, NPM researcher and Information Center Director, said that the earliest backup method was to first save data on a hard disk, and then backup the data on a magnetic tape and CD. Saving the data on a CD was for passing data for business affairs, but now many image files are easily larger than 4.7GB and using magnetic tapes in replace has been considered as an option.

(2) Remote Backup

When digitization reaches a certain degree, each digital file will require backup. At first, the National Palace Museum used CDs as its main storage media for backup, it then added a LTO-1 magnetic tape machine and collocated it with a 400GB RAID; afterwards, it upgraded to a LTO-2 magnetic tape machine and added a 2TB RAID, and then expanded it to 4TB. Servers were mainly IBM's R6 during early periods, they were then upgraded to E450, E650 and V880; some servers used Microsoft operating systems and were aided by Tivoli and Veritas software. These software and hardware ensure the backup security of digital artifacts.

Digital archives start from digital photographing, to the preservation of RAW files, file fragmentation, conversion to TIFF files, naming management, saving digital files to RAID, and then either writing to disc on a jukebox or using the LTO-2 magnetic tape machine for backup. Moving data is a tormenting task. Therefore, after NPM digital files are uploaded to a fixed partition, they are directly written to disc or magnetic tape for backup. The National Palace Museum currently has 2 jukeboxes, 2 LTO-2 magnetic tape machines and 1 LTO-1 magnetic tape machine to support these backup operations. NPM has also planned backup systems in general facilities, but in consideration to the security and usability of

digital files, Lin Kuo-Ping believes that the true essence of digital archives should be the assured preservation and backup of artifacts. Giving the storehouse the highest anti-shock and anti-disaster security is probably providing the best location for remote backup.

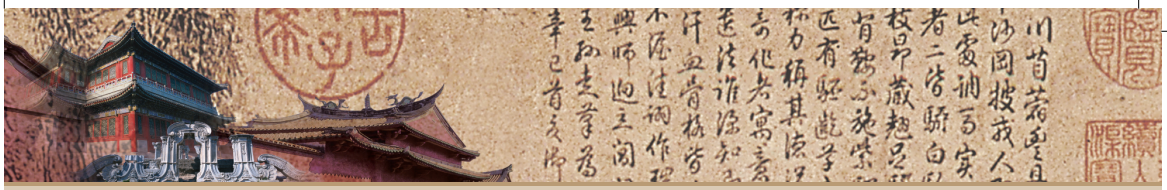
III. National Central Library

(1) Storage Equipment

Before the National Digital Archives Program, the National Central Library used microfilms as its storage media, and it wasn't until 2001 when it joined the Program did it start using digital scanning and DVD and RAID as storage media.

(2) Remote Backup

Editor Sung Mei-Chen explained that two backup copies are made for all digitization achievements and are stored on different floors of the Library, "We are considering on storing one backup copy in the Information and Computing Library or elsewhere".



One. Platform Scanner

I. Platform Scanner

(1) Basic Settings:

Arrange for three working staff, in which two are responsible for scanning, arrangement and verification, and the other is responsible for system development, preparation and maintenance. The equipment used is two computers and one platform scanner.

Table 1. Basic settings of platform scanner cost estimation

	Category		Description	Quantity	Price
1	Scanning equipment	Software	Computer	2	NT$60,000
			Platform Scanner	1	NT$130,000
		Hardware	Adobe Photoshop	1	NT$20,000
2	System development and preparation (software + hardware)		String-bound book digitization management system preparation expenses	1	NT$600,000
3	Manpower		Salary	3	NT$90,000 per month
4	File Size		A3 Full color 300dpi scanning	1	Takes roughly 90 seconds
6	Output		Digitization output per month (20 workdays)	4000 pages	
7	Service Life		Equipment lifespan (software + hardware)	4 years	
8	Consumables		DVDR	1 disc	NT$15

(2) Calculation Example: Calculation of equipment amortization expenses based on service life.

Table 2. Calculation example of platform scanner cost estimation

	Scanning Cost Estimation (Platform Scanner)		
Definition	Material	DVDR (4.7GB=4700MB)	
	Expenses	Salaries	
	Labor Expenses Equipment Amortization Expenses	(Scanning Equipment Software/Hardware + System Preparation Software/ Hardware)/Service Life	
Formula	[Material Expenses (NT$)/Digital Output Quantity (pages)] + [Labor Expenses (NT$) + Equipment Amortization Expenses (NT$)/Digital Output Quantity (pages)=Cost per page (NT$/page)		
Calculation Example	Material Expenses	Pages Scanned	4700/50 (MB)=94 (pages)
		DVD cost per page	15/94=NT$0.15 per page
		Quantity×2 sets	0.15×2=NT$0.3 per page
	Labor Expenses	NT$30,000*3=NT$90,000	
	Equipment Amortization Expenses	(60,000+130,000+20,000)/4=NT$52,500 per year Or, NT$1,375 per month	
	Cost per page	(90,000+1,375)/4000=NT$22.8 per page 0.15 + 22.8 = NT$22.95 per page	

Two. Flatbed Scanner

I. Zeutschel Flatbed Scanner

(1) Basic Settings:

Arrange for three working staff, in which two are responsible for scanning, arrangement and verification, and the other is responsible for system development, preparation and maintenance. The equipment used is two computers and one flatbed scanner.

Table 3. Basic settings of flatbed scanner (Zeutschel) cost estimation

	Category	Description		Quantity	Price
1	Scanning equipment	Software	Computer	2	NT$60,000
			Platform Scanner	1	NT$1,000,000
		Hardware	Adobe Photoshop	1	NT$20,000
2	System development and preparation (software + hardware)	String-bound book digitization management system preparation expenses		1	NT$600,000
3	Manpower	Salary		3	NT$90,000 per month
4	File Size	A3 Full color 300dpi scanning		1	Takes roughly 90 seconds
6	Output	Digitization output per month (20 workdays)		4000 pages	
7	Service Life	Equipment lifespan (software + hardware)		4 years	
8	Consumables	DVDR		1 disc	NT$15

(2) Calculation Example: Calculation of equipment amortization expenses based on service life.

Table 4. Calculation example of platform scanner cost estimation

	Scanning Cost Estimation (Platform Scanner)		
Definition	Material Expenses	DVDR (4.7GB=4700MB)	
	Labor Expenses	Salaries	
	Equipment Amortization Expenses	(Scanning Equipment Software/ Hardware + System Preparation Software/Hardware)/Service Life	
Formula	[Material Expenses (NT$)/Digital Output Quantity (pages)] + [Labor Expenses (NT$) + Equipment Amortization Expenses (NT$)/Digital Output Quantity (pages)]= Cost per page (NT$/page)		
Calculation Example	Material Expenses	Pages Scanned	4700/50 (MB)=94 (pages)
		DVD cost per page	15/94=NT$0.15 per page
		Quantity×2 sets	0.15×2=NT$0.3 per page

	Labor Expenses	NT$30,000*3=NT$90,000
	Equipment Amortization Expenses	(60,000+1,000,000+20,000)/4=NT$ 270,000 per year Or, NT$22,500 per month
	Cost per page	(90,000+22,500)/4000=NT$28.13 per page 0.15 + 28.13 = NT$28.28 per page

II. Book Photographing Platform–Flatbed Scanner (Separate purchase of lighting equipment and digital camera required)

(1) Basic Settings:

Arrange for three working staff, in which two are responsible for scanning, arrangement and verification, and the other is responsible for system development, preparation and maintenance. The equipment used is two computers and one flatbed scanner.

Table 5. Basic settings of flatbed scanner (book photographing platform) cost estimation

Category			Description	Quantity	Price
1	Scanning equipment	Hardware	Computer	2	NT$60,000
			Book photographing platform	1	NT$350,000
			Lighting equipment		NT$1,200,000
			Digital camera		NT$150,000
		Software	Adobe Photoshop	1	NT$20,000
2	System development and preparation (software + hardware)		String-bound book digitization management system preparation expenses	1	NT$600,000
3	Manpower		Salary	3	NT$90,000 per month
4	File Size		A3 Full color 300dpi scanning	1	Takes roughly 90 seconds
6	Output		Digitization output per month (20 workdays)	4000 pages	
7	Service Life		Equipment lifespan (software + hardware)	4 years	
8	Consumables		DVDR	1 disc	NT$15

(2) Calculation Example: Calculation of equipment amortization expenses based on service life.

Table 6. Calculation example of flatbed scanner (book photographing platform) cost estimation

	Scanning Cost Estimation (Book Photographing Platform)		
Definition	Material Expenses	DVDR (4.7GB=4700MB)	
	Labor Expenses	Salaries	
	Equipment Amortization Expenses	(Scanning Equipment Software/ Hardware + System Preparation Software/Hardware)/Service Life	
Formula	[Material Expenses (NT$)/Digital Output Quantity (pages)] + [Labor Expenses (NT$) + Equipment Amortization Expenses (NT$)/Digital Output Quantity (pages)]=Cost per page (NT$/page)		
Calculation Example	Material Expenses	Pages Scanned	4700/50 (MB)=94 (pages)
		DVD cost per page	15/94=NT$0.15 per page
		Quantity×2 sets	0.15×2=NT$0.3 per page
	Labor Expenses	NT$30,000*3=NT$90,000	
	Equipment Amortization Expenses	(60,000+1,700,000+20,000)/4=NT$445,000 per year Or, NT$37,083 per month	
	Cost per page	(90,000+37,083)/4000=NT$31.77 per page 0.15 + 31.77 = NT$31.92 per page	

III. Book Photographing Platform–Flatbed Scanner (Unit already has lighting equipment and digital camera)

(1) Basic Settings:

Arrange for three working staff, in which two are responsible for scanning, arrangement and verification, and the other is responsible for system development, preparation and maintenance. The equipment used is two computers and one flatbed scanner.

Table 7. Basic settings of flatbed scanner (book photographing platform) cost estimation

Category			Description	Quantity	Price
1	Scanning equipment	Hardware	Computer	2	NT$60,000
			Book photographing platform	1	NT$350,000
		Software	Adobe Photoshop	1	NT$20,000
2	System development and preparation (software + hardware)		String-bound book digitization management system preparation expenses	1	NT$600,000
3	Manpower		Salary	3	NT$90,000 per month
4	File Size		A3 Full color 300dpi scanning	1	Takes roughly 90 seconds
6	Output		Digitization output per month (20 workdays)	4000 pages	
7	Service Life		Equipment lifespan (software + hardware)	4 years	
8	Consumables		DVDR	1 disc	NT$15

(2) Calculation Example: Calculation of equipment amortization expenses based on service life.

Table 8. Calculation example of flatbed scanner (book photographing platform) cost estimation

	Scanning Cost Estimation (Unit already has lighting equipment and digital camera)		
Definition	Material Expenses	DVDR (4.7GB=4700MB)	
	Labor Expenses	Salaries	
	Equipment Amortization Expenses	(Scanning Equipment Software/ Hardware + System Preparation Software/Hardware)/Service Life	
Formula	[Material Expenses (NT$)/Digital Output Quantity (pages)] + [Labor Expenses (NT$) + Equipment Amortization Expenses (NT$)/Digital Output Quantity (pages)] = Cost per page (NT$/page)		
Calculation Example	Material Expenses	Pages Scanned	4700/50 (MB)=94 (pages)
		DVD cost per page	15/94=NT$0.15 per page
		Quantity×2 sets	0.15×2=NT$0.3 per page
	Labor Expenses	NT$30,000*3=NT$90,000	
	Equipment Amortization Expenses	(60,000+350,000+20,000)/4=NT$107,500 per year Or, NT$8,958 per month	
	Cost per page	(90,000+8,958)/4000=NT$24.74 per page 0.15 + 24.74 = NT$24.89 per page	

String-Bound Book Digitization Procedures Guideline

Publisher: Taiwan Digital Archives Expansion Project
Editor-in-Chief: Peng-Sheng Chiu
Executive Editor: Huei-Jing Lin
Author: Chih-Tung Kao
Copy-Editor: Wei Peng
Transfer: International Collaboration and Promotion of Taiwan e-Learning and Digital Archives Project
Typesetting: Ya-Chen Tsai
Printing Co: Hegu Exquisite Printing Co., LTD.

Taiwan e-Learning and Digital Archives Program
Taiwan Digital Archives Expansion Project of Taiwan e-Learning and Digital Archives Program
Tel: +886-2-27829555#288
Fax: +886-2-27868834
Website: http://content.teldap.tw
Email: content@gate.sinica.edu.tw
Address: Taiwan Digital Archives Expansion Project Office, Museum of the Institute of History & Philology, Academia Sinica, NanKang, Taipei 11529, Taiwan
Publish Date: 1st Edition November 2008
ISBN: 978-986-01-5450-4（平裝）