Archaeological Data Digitization Procedures Guideline

Wang Ya-Ping
Lu Chun-Yi

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

Acknowledgements

For the completion of this book we thank the Life and Culture Thematic Group under TELDAP Taiwan Digital Archives Expansion Project, including “The Project of Historical and Cultural Heritages Developed in the Institute of History and Philology, Institute Academia Sinica,” the National Museum of Natural Science Archaeology Department, National Taiwan University’s “Digitizing Project of Anthropological Collections,” the National Museum of History: National Historical and Cultural Artifacts Digital Collection Project, and Chang Jung Christian University. Data used by this guideline are the results and experiences of digitization projects over the years, and without their strenuous efforts, this guideline could not have been completed. We hope to aid even more digitization projects and digital archive workers with this digitization procedures guideline. We would like to specially thank Lin Yu-Yun of Academia Sinica Institute of History and Philology for acting as the reviewer of this book, as well as Miss Huang Ting Yu, Mr. Wen Tzu-Chun, and fellow colleagues of this project for providing suggestions on the contents and writing method of this book.

Publisher’s Preface

After the “National Digital Archives Program” was initiated in 2002, members of numerous institutional projects and request-for-proposals projects joined our team to engage in digital work that covered countless categories and massive amounts of content. The first phase of the five year project was successfully completed in 2006. The following year, the “National Digital Archives Program” and “National Science and Technology Program for e-Learning” were integrated into the “Taiwan e-Learning and Digital Archives Program (TELDAP, http://teldap.tw/)”, striving to achieve the ultimate goal of “presenting Taiwan’s cultural and natural diversity” as it continued to expand digital resources in various fields, and systemically promoted digital achievements in education, research and industries. TELDAP is preparing to actively collaborate with the private sector to drive growth in related industries, not only preserving important cultural assets, but also accelerating the development of a new culture in the digital age of today.

Originally named the “Content Development Division” during the first phase, we were renamed “Taiwan Digital Archives Expansion Project” (http://content.teldap.tw) as a subproject of TELDAP, and took more active measures to expand the sources of digital content, extending our reach to the collections of private institutions and even individuals. We have widely requested proposals for digitization projects related to archives, archaeology, philology, geography, ethnicity, art, daily life, animals and plants, and hope to better integrate digital content with different characteristics, to develop them into fun and inspiring digital materials, and to provide them free of charge to the public for education and research; this will also help firms and public or private holding institutions to find cooperation opportunities in value-added applications. Collaboration between the “Taiwan Digital Archives Expansion Project” and other projects under the “Taiwan e-Learning and Digital Archives Program” will help speed up development of educational, research and commercial value-added applications of digital content, which will benefit the presentation of Taiwan’s cultural and natural diversity, and allow people everywhere around to understand and appreciate the richness of our history and culture, as well as the beauty of our natural ecology.

While collecting and developing value-added applications of digital content, whether it may be during the “Content Development Division” or “Taiwan Digital Archives Expansion Project” period, members of this project have continuously followed up on digital workflow related technologies used by public and private institutions and open request-for-proposals projects, and compiled a series of “Digitization Procedures Guideline Books” that introduce various international standards and provide information on digitization technologies and workflows. Since 2005, we have written 21 digitization procedures guidelines on different themes (the full text of all of the 21 books can be downloaded from the “Taiwan Digital Archives Expansion Project” website under “Virtual Library: Digitization Books”), selecting exquisite digital objects, such as ceramics, paintings, calligraphy, and string-bound books, combining the experiences of different institutional projects, and supporting them with domestic and foreign theories and practice results.

Since 2008, we have continuously revised and expanded our “Digitization Procedures Guideline” book series, hoping to expand distribution channels so that they may be provided to even more museums, libraries, institutions and individuals for reference. Our preparations are mainly divided into revising existing guidelines for “selected objects” and compiling new guidelines on “common principles”. The former refers to revising the existing 21 guidelines with a focus on introducing new digitization technologies and specifications, more practical software and hardware, and digital content protection mechanisms; we expect to revise seven books per year and complete all 21 books within three years. As for compiling guidelines on “common principles,” our emphasis will be on the introduction of key concepts, such as the “life cycle” of digital information and quality control, studying multiple types of

objects instead of a single type of object, and adopting common principles as the guideline framework. The so called common principles refer to project planning, integrated workflow, audiovisual data, text data, color management, outsourcing management, and digital content protection and authorization. These eight common principles are topics of which we will investigate, study and write guidelines for; we expect to publish eight guidelines in three years.

Guidelines for selected objects and guidelines on common principles in fact complement each another. Guidelines on common principles emphasize on the analysis of important topics in digitization work, guiding readers to thoroughly consider the advantages and disadvantages of digitization. Guidelines on selected objects describe practices and techniques for digitizing specific objects, helping readers to select the most suitable, most effective digitization workflow. By publishing this "Digitization Procedures Guideline" book series, we believe that we are providing institutions and individuals with the intention to engage in digitization work with a series of practical guidelines that provide an overall view, while guiding them step by step through the digital workflow. Here we must stress that the theoretical foundation of this book series is the precious experiences of institutional and request-for-proposal project teams accumulated throughout the years. These experiences allow higher quality digital content to be produced, presented and maintained with less cost, further enriching our digital archives and e-learning content. As we continue to publish our "Digitization Procedures Guideline" book series, we must give special thanks to working partners who were interviewed and colleagues who were a part of writing the guidelines, and are grateful to the scholars and specialists that reviewed and provided their advice on the book series. Finally, we hope that readers will not be reluctant to correct any mistakes or make recommendations that will help us be even better.

Taiwan e-Learning and Digital Archives Program
Taiwan Digital Archives Expansion Project
Digital Archives Sub-project of Project Integration

Project Director

February 10th, 2010

Contents

ONE. Introduction

Every day you open your eyes to an environment that couldn't be more familiar, but have you ever thought about the traces left by the ancients hidden beneath? Have you ever found a piece of shattered ceramic when you were taking a walk in the hill near where you live? These objects could very likely be what we call in archeology as "archeological data." So why do these "archeological data" appear in our surroundings? What meaning does traces of how past people led their lives bring to us? This is an interesting exploration process that has great research value.

I. What is archeology?

Whenever archeology becomes the topic of discussion, people generally think of movie scenes with Indiana Jones or Lara Croft going deep into the wild or deserts, and finding great treasures after surviving unthinkable dangers. But these are mostly just fantasies created by novels or movies. In the real world, archeologists go into the field and use scientific methods and techniques to excavate remains, relics and sites left by early people; they do not go on treasure hunts or adventures.

The first site to be found in Taiwan was the Chih Shan Yen Prehistoric Site found in 1896 during the Japanese Colonial Rule Period. Following the discovery of the Yuanshan Site, Japanese scholars began systematic excavations, gradually entering Taiwan's east coast and south west plains, and proposed and established Taiwan's prehistoric culture system. Archeological data shows that human beings fished, hunted and gathered in Taiwan as early as 30~50 thousand years ago; their cultural characteristics can be found from the human bones excavated from the Changbin Culture of Eastern Taiwan and Tainan Zuozhen. This shows that research work of archeological excavations proves different cultural appearances during different periods, and also helps illustrate a complete image of Taiwan's history and culture.[1]

A further explanation of the purpose of archeology is "to utilize scientific field work to understand the meaning of sites, remains and relics, and further understand different aspects of human behavior and culture."[2] To achieve this purpose, everywhere around the world there are archeologists searching for prehistoric sites, and excavating relics and remains of the ancients.[3] So what exactly are the sites, relics, and remains mentioned in archeological data? In simple words, a "site" is where ancient people dwelled or engaged in various activities; "relics" are the small portion of utensils used in everyday life that survived through history; and "remains" are the houses, graves, wells, fortresses, and irrigation ditches built by ancient people.[4] Here, we believe everyone has a fundamental understanding of the purpose of archeology and contents of archeological data, and that the life, behavior or culture of ancient people are inferred or explained based on the sites, remains and relics found in field investigations.

However, after these precious archeological data are unearthed, they still face several issues. First, due to their history and material, relics will rust or weather; next, besides relatively complete specimen that are published in formal academic reports or displayed in museums, most archeological data are hidden away from the general public in public and private institutions. Following the development of modern technology, we have adopted a new concept and approach to the preservation and application of archeological data, this new approach is the so called "digitization" of archeological data.

II. Why digitization?

"Digitization" is an often heard term that few people ever truly understand. "Digitization" can simply be defined as the use of a digital scanner or camera to convert text or images into digital data that can be processed by a computer.[5] As a matter of fact, many collections are not limited to text and images, and may include specimen, implements, video data, and audio data. These collections can also be converted into digital data with the right digitization procedures using digital scanners or cameras (this is further divided into 2D and 3D photography), and become "digital archives". Therefore, converting precious archeological data into "digital archives" not only ensures the availability, preservation and integration of data, but also provides the following benefits:

1. Increase utilization value

In the past, archeological data that were excavated or gathered were always stored in repositories for protection and preservation considerations. The digitization of archeological data provides an alternative to using physical

[1] Taiwan Network-based Research – Taiwan Archeology, http://twstudy.iis.sinica.edu.tw/archeotw/index.html, Search: January 2010.

[2] Chia Nai, Wang Chung-Shu, "Encyclopedia of China – Archeology", Taipei: Encyclopedia of China Publishing House, 1986, pages 1-3.

[3] "Hsuan Hsuan Learns Archeology", written by Liu Ke-Hung, illustations by Li Chin Lun, Taipei City: Council for Cultural Affairs, 1997, page 55.

[4] Same as 3.

[5] Chen Yu-Wen, "Introduction to data digital conversion", National Central Library Bulletin Volume 28 Issue 2, 1995, pages 3-12.

objects. Besides researchers and scholars, the general public can also utilize digital databases to search through catalogs and gain additional information on physical objects displayed in museums. High quality digital images are not only convenient for display and publishing, but also enhance archive management.

2. Facilitate academic exchanges

   Digitization allows the preservation of precious archeological data, and establishes a digital knowledge base for access. The far stretching internet allows archeological data stored in knowledge bases to be studied and browsed by archeologists in remote areas, which is sure to provide certain benefits to academic exchanges and development.

3. Promote archeology education

   Digital archeological data can be utilized in archeology related courses or as additional materials in general education courses, allowing archeology knowledge to strike roots and benefit the propagation and promotion of cultural assets.

From the above we can clearly understand that applying digital archives to archeology brings numerous benefits. In fact, many institutions in Taiwan involved with archeology related research or teaching, e.g. Academia Sinica Institute of History and Philology, National Taiwan University Department of Anthropology, National Museum of Natural Science Anthropology Department Archeology Division, and National Museum of Prehistory, have been engaged in the digitization of archeological data for numerous years.

However, due to different characteristics and management methods of institutions, different digitization procedures are adopted for the same type of collection (for example, cultural relics may either be categorized according to material or function). Under this circumstance, how can we provide institutions or individuals with the intention to become engaged in the field of digital archives with good reference material? Thus, compiling a "digitization procedures guideline" for archeological data is a task that can not wait.

Archeological data comes in a great variety, each with its own digitization method. After considering the actual situation of various digitization projects, we have generalized a few main procedures. In the following chapter we will briefly introduce the "digitization flowchart" for archeological data, which will mainly focus on digital photography procedures.

TWO. Digitization Flowchart

Archaeological data can be divided into two major categories “field data” and “archaeological remains.” Field data includes data collected from different work stages (artifacts, remains, sites) and by the smallest work units (excavation unit). The two main digitization methods are “scanning” and “photography.” This guidelines focuses on the “photography” digitization method, please refer to “Archives Digitization Procedures Guideline” and “String-bound Book Digitization Procedures Guideline” for more details on the “scanning” digitization method.

The digital workflow may vary along with characteristics and management method of the institution, but can generally be divided into the following procedures (as shown in Fig.2-1):

1. “Preliminary procedures” are preparations before digitization, including image file specification establishment , formulating digital file nomenclature, arrangements and file creation.
2. “Digitization procedures” explains working procedures and notices of using photographic equipments during digitalization, and introduces working procedures of 2D and 3D photography.
3. “Metadata and database develepment” explains the format of metadata for archaeological data, working procedures of cagaloging metadata and workflow of database develepment.

Besides the above contents, this guideline also includes discussions on digital content protection mechanisms, equipment and cost analysis, and outsourcing for reference by future digitization projects.

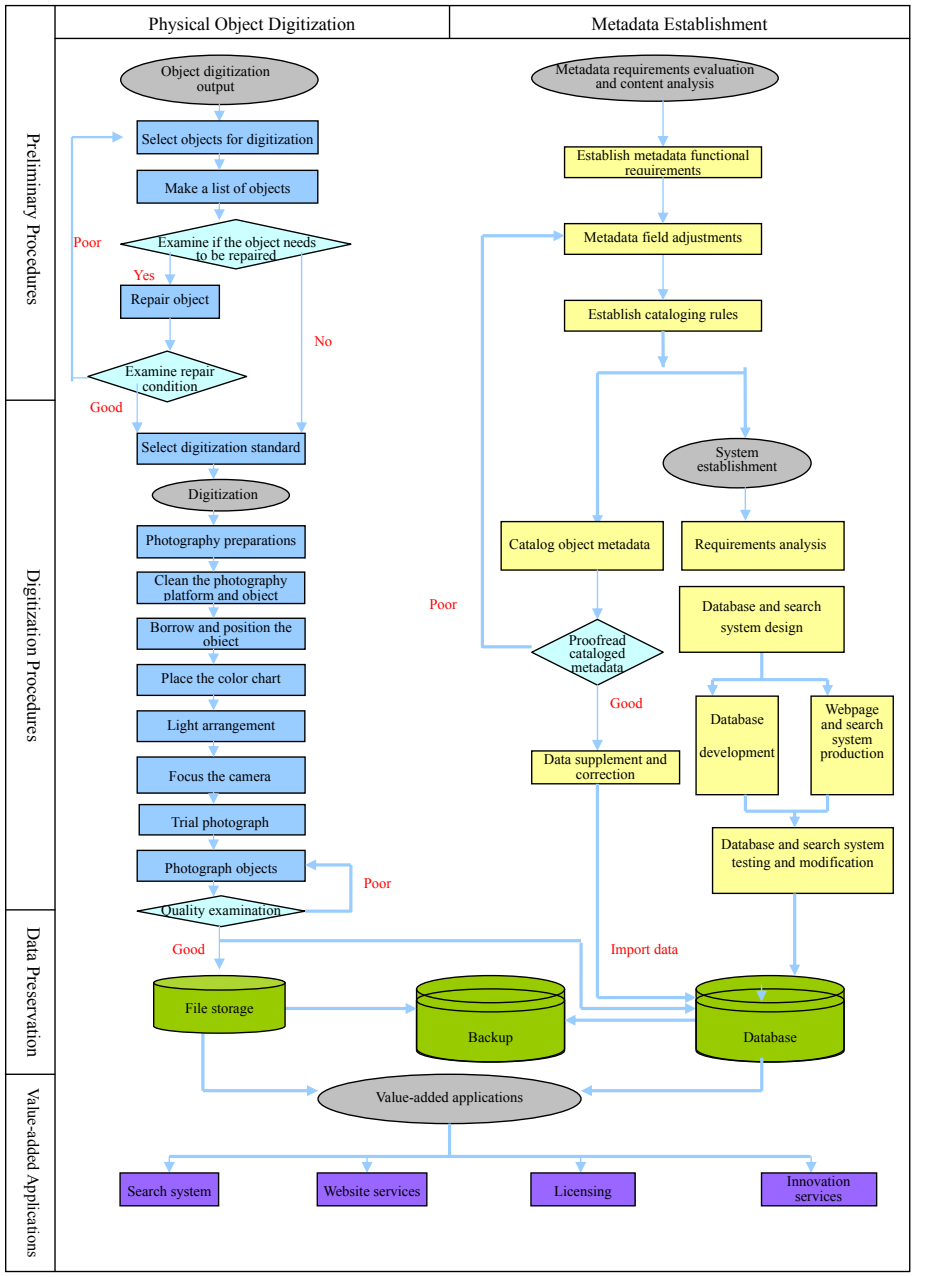


Fig.2-1 Digitalization Flowchart for Archaeological Data

THREE. Preliminary Procedures

Telling from the digitization flowchart for archeological data, the preliminary procedures stage is the foundation of all procedures. Preliminary procedures are relatively time consuming and laborious because they plan following digitization procedures, and develop approaches to achieve project goals and quality under limited cost and human resources. Besides make a list of objects to be digitized, preliminary procedures also include image file specification establishment, image file name coding principles, and arrangements and file creation.

I. Image File Specification Establishment

The image file format must be established based on considerations of future usage requirements and characteristics before initiating digitization.

1. Image file specifications

Generally speaking, image file specifications include file format, color mode, and resolution. Understand each file characteristic and establish a specification based on requirements of the execution unit.

i. File format

The main file formats that are adopted in image digitization are RAW, TIFF, and JPEG. In which RAW files are the original image files produced by digital cameras; RAW files are not compressed, can show rich colors and layers, and can undergo white balance and brightness adjustments. In terms of various applications, RAW is currently a good storage format because it stores an unmodified image. RAW has gradually been adopted by institutions engaged in digitization as a means for permanent preservation. RAW files are generally converted into TIFF or JPEG files for applications. TIFF files are the easiest to circulate because its technology achieves lossless compression with relatively higher image quality, making it commonly used for professional applications, such as printing. JPEG is a file format that uses lossy compression, but it is the most popular on the internet because of its small file size. Please see the table below for characteristics of digital image file formats (Table 3-1).

Table 3-1 Characteristics of digital image file formats[6]

File format	RAW	TIFF	EPS	JPEG	GIF	BMP	PICT	PSD	PNG
Supports RGB full color	●	●	●	●		●	●	●	●
Supports 256 colors	●	●	●		●	●	●	●	●
Supports CMYK	●	●	●	●				●	
Image compression		●	●	●	●		●		●
Supports layers								●	
Supports masks			◎		●			●	●
Supports online display				●	●				●
Suitable for typical image storage		●		●				●	●
Suitable for long-term image storage		●							
Suitable for printing output		●	●						

● Indicates the function is supported ◎ The new TIFF format supports masks

Source: "Digital Photography Technology" by Hsu Ming-Ching, 2001

ii. Color mode

The recommended color mode is RGB[7] because it contains more colors (color gamut) than the CMYK[8] color mode, and photographed images are always first observed on a monitor. A simple conversion can be carried out when digital images are used for other purposes, such as printing.

iii. Color depth

Color depth is related to the color mode chosen, in the RGB color mode R, G and B each has 8 bits (1 Byte), so the color depth of typical computer monitors is 24 bits, also known as 24 bits full color mode. Today, digital products in the market have CCDs (Charge Couple Device)[9] that can sample 16 bits or higher for R, G and B, but this means that equipment with higher specifications are required to process such image files.[10]

iv. Resolution and image size

DPI (Dot Per Inch) expresses an image by the number of dots that are in an inch. The dots that form images are what we call pixels. Typical digital cameras use the total number of pixels to express their resolution; professional digital cameras can have 60 megapixels or even higher. Most people believe that higher resolution is better, but resolutions should in fact be selected based on properties of the object being photographed and the application purpose (e.g. publishing or online browsing); a camera with the highest resolution may not

[6] If you have any questions on image file characteristics, file formats and selection, please refer to chapter five "Data Storage and File Formats" of "Digital Photography Technology" by Hsu Ming-Ching.

[7] RGB is the most commonly used color mode. RGB stands for the three primary colors Red, Green and Blue.

[8] CMYK refers to the four colors used in printing, C for Cyan, M for Magenta, Y for Yellow and K for Skeleton or Black.

[9] CCD (Charged Couple Device) is the light sensing device used in most digital cameras. It can convert light signals of different intensities into electronic signals.

[10] Please refer to chapter one "Basic Concepts: of the "Digitization Procedures Guideline: Color Management" issued by the Taiwan Digital Archives Expansion Project for more information on basic concepts of color management, April, 2009.

necessarily be the best for you. Higher resolution means longer digitization time, larger file size, and relatively higher cost, including future maintenance cost (e.g. permanent storage and remote backup). Therefore, whether or not the highest quality is adopted during tender invitation should be based on the institution's requirements.

2. Image file usage requirements

Digital image files can be divided into three categories: permanent preservation, commercial purposes, and online browsing. Digital image files for "permanent preservation" have the highest quality, and are provided for digital archives and research purposes; hence, such image files must meet basic requirements of certain digitization parameters to ensure their quality. Digital image files for "commercial purposes" are mainly used for commercial value-added applications; while files under the "online browsing" category are provided for download or online browsing, and have lower requirements on image quality. Institutions can establish a suitable digitization format (as shown in Table 3-2) after considering their requirements and actual condition.

For the abovementioned image usage requirements and file format characteristics, the "Taiwan e-Learning and Digital Archives Program" provided a table on digital image file formats (please refer to Table 3-2) to use as a standard or reference for digitization work. If the digital images will be used in publications or value-added applications, also refer to Table 3-3 for image file size and common output size.

Table 3-2 Digital image file formats for different purposes

Purpose	Resolution and size	Color mode	File format
Permanent preservation	Original size, 300dpi and above	RGB (24bit/pixel) or above	RAW or TIFF (Uncompressed)
Commercial purpose	Original size, 300dpi	RGB (24bit/pixel)	TIFF (Uncompressed)
Online browsing	Based on website requirements, 72dpi	RGB (24bit/pixel)	JPEG (Compressed)

Table 3-3 Image file size and common output size

Size (inch)	Size (centimeter)	Pixel Height x Width	Resolution (dot per inch)	File format	File size (MB)
4x6	10.16x15.24	1200x1800	300dpi	TIFF	6.18
8x10	20.32x25.4	2400x3000	300dpi	TIFF	20.6
A4	21x29.7	2480x3508	300dpi	TIFF	24.9
10x12	25.4x30.48	3000x3600	300dpi	TIFF	30.9
A3	29.7x42	3508x4961	300dpi	TIFF	49.8

Table redrawn from: Chapter four "File Size and Image Size" of "Digital Photography Technology" by Hsu Ming-Ching, page 44.

When using the 2 tables above, for example, a file is categorized under "permanent preservation" or "commercial purposes" based on the scanning condition of the archive or negative. At present, most 2D images of archeological data are digital born, meaning that files are created with pixels and not resolution (Dot Per Inch). Current professional digital cameras have up to 60 megapixels or even higher, institutions can change the file size according to future requirements; for example, if the image file will be printed in A3 size, with the resolution requirement of printing at 300dpi, then a comparison with Table 3-3 will tell you that the image must be photographed with at least 3508x4961 pixels; if the image file is in TIFF format and a typical color depth, then the file size will be roughly 49.8MB.

In the case of Academia Sinica Institute of History and Philology, their archeological data is further divided into typical relics, important specimen, and articles of fine quality for digital photography. Each category uses different resolutions and different files sizes; the digital images may be in 2D or 3D (as shown in Table 3-4).

Table 3-4 Archeological data digitization specifications of Academia Sinica Institute of History and Philology

Type	Format
2D images	Articles of fine quality: At least six images per article, 48bits/RGB/tiff; 10050x12600 pixels, ca150MB. (Phase One)
	Important specimen (artifacts): At least 2-3 images per article, 8bit/RGB/tiff; 10050x7000-8000 pixels, ca150MB. (Phase One)
	Typical artifacts: 1-2 image per article, 8bit/RGB/jpg: 1MB
3D images	168 images of 512 by 512 pixels

Source: Academia Sinica Institute of History and Philology Archeology Division

II. Image File Name Coding Principles

File names to images are like ID numbers to us, so file name coding principles must be thoroughly thought out.[11] Image file names aim to provide convenience of

file management and link requirements during cataloging. When establishing coding principles, there are two matters that require special attention: First, coding principles must be established when image files are first created; this is to prevent waste of time and human resources on arranging images again and again when image files accumulate to a certain quantity. Second, most digital files are named using numbers only, which makes management convenient, but prevents recognition of image contents based on the file name alone; therefore, file names should aim to make content recognition easy.[12] Using Academia Sinica Institute of History and Philology as an example, its coding principles for archeological data images give consideration to the site, pit, level, archive number, image type, full/partial, and direction/position. The image file name coding principles for archeological data of Academia Sinica Institute of History and Philology are provided in appendix 2, institutions can refer to the example to plan image file name coding principles that best meet their requirements.

III. Arrangements and File Creation

1. Arrangement and repair

   Before digitization, archeologists should arrange relics excavated from the field, select pieces that meet standards for repair, and take the relics from the repository out for repair.

2. Categorization and selection

   After repair the relics, categorize them according to material, color, decorations, and physical appearance, and then select relics that need to be digitized based on different types and characteristics.

3. File creation

   Researchers should collect data on relics, including the holding institution, site name, culture type, function or purpose, era, level, and where the relic is stored. This information should be stored in a Microsoft Excel form, creating a preliminary data list.

[11] "Digital image file establishment procedures and current status of value-added applications – Using the National Palace Museum as an example", Chien Sung-Tsun, Chen Yao-Tung, Taipei: International Symposium on Digital Museums – Antiquities and Paintings and Calligraphy, 2003.

[12] Academia Sinica Institute of History and Philology "Academia Sinica Institute of History and Philology Archeological Data Digitization Workflow", "Digitization Workflow: Archeology Thematic Group", Taipei City: Contents Development Division, 2004, page 146.

FOUR. Object Digitization Procedures

In the following section digitization of archeological data is divided into two methods, 2D and 3D photography; procedures include equipment preparation, object photography, color management, post processing and file storage. However, this guideline merely serves as reference for operating procedures, and does not include discussions on artistic rendering, please consult professionals in related fields for details if you have such requirements.

I. Photography Site Planning

Before formally photographing objects, select a suitable photography location to benefit equipment arrangements and photography work. Site arrangements aim to eliminate any factors that might affect image quality and facilitate the digitization process. Therefore, take notice of the following principles:

1. The photography site should not be too small, and should at least be able to hold a photography platform, lights, camera, computer, and desk, as well as have space for personnel to move around. Larger space not only allows personnel to move around easier, but also reduces risk of damaging objects. If large objects are being photographed, then view depth should also be put into consideration; larger space will benefit moving photography equipment and placing objects.
2. To prevent stray light from affecting image quality, the photography site is recommended to be completely dark with only lights from lighting equipment. If the work place has windows, put up blackout fabric or thick paperboards to block the light.
3. Selection of the photography location should give consideration to the safety of objects, reduce transportation distance, and carefully plan the line of movement for moving objects. The photography studio is best away from vibration sources to avoid blurry images, including the street, railway and crowds.
4. Photography studio arrangements: The best wall color is gray, followed by white or black. The main consideration of wall color selection is to avoid affecting color management; lights used for photographing objects will reflect off the wall and ground and affect image quality.

II. Photography Equipment Preparation

"One must have good tools in order to do a good job," to gain good and stable digital output quality, equipment preparation before photographing objects is also an important link. Preparations for the photography platform, lighting equipment and camera are briefly described below.

1. Photography platform

Main considerations of the photography platform include object size and background paper. Before placing objects, use a brush to swipe the platform free of any dust. Replace the background paper if it cannot be cleaned. Typically, gray, white or black background paper is used, in which gray is a neutral color; the recommended color is 18% standard gray because it is suitable for objects of any color, and facilitate color management and future publishing applications. However, avoid using background paper that is too similar in color to the object being photographed; for example, if the object is light colored, then select black or gray background paper; if the object is dark colored, then select white background paper.

2. Lighting equipment

Photography lights are divided into cold lights and flash lights. Although cold lights can maintain a standard color (5000K-6000K) and does little harm to objects, photography results are not as good as flash lights. Typically, 2D photography starts roughly 20~30 minutes after warming up the lighting system and computer.

3. Camera

i. Attach the camera to a stand to ensure that it is level and stable. Use a level to measure and adjust the camera's angle to the object being photographed. Make sure the object is level with the camera so that it is completely captured in the image.

ii. A lens hood can be added to prevent light from reflecting into the lens and producing spots.

iii. While setting up the camera, link it to the computer for software configurations. (Please refer to chapter seven for information on camera related equipment)

4. Color calibration

For image color to be correctly displayed, the computer monitor, camera and printer all need to complete color calibration.

i. Monitor color calibration: Turn on the monitor thirty minutes before color calibration. A monitor hood is recommended to eliminate influence from light sources other than the monitor, which will reduce the accuracy of color calibration. For color calibration you can use Adobe Photoshop's Adobe Gamma

or Apple MAC's ColorSync together with a color meter to generate an ICC Profile.

ii. Digital camera color calibration: Place the color calibration card in a photography environment under a stable light source. After photographing the color calibration cord, use your computer's color calibration software to generate a color profile. A new color profile must be generated whenever the light source is adjusted.

5. Other: Computer desk, object cleaning tools, and lists of objects to be photographed.

III. 2D Photography

1. Borrow objects

Select objects according to digitization requirements, and select suitable specimens based on different digitization methods. Institutions have varying regulations on borrowing objects, so staff members are usually asked to fill out a loan form; loan forms are used as proof of incoming and outgoing objects, and help institutions manage their collections.

2. Position the object

Personnel should place the object on the photography platform and verify the photography angle and number of pictures to be taken. Use a wind blower to remove any dust on the relic, if necessary.

3. Place the color chart and scale

Place a color chart in a suitable place on the photography platform for the purpose of color calibration. Color charts serve two purposes, one is white balance or gray balance, and the other is for creating an ICC Profile; in an ideal situation, any change the angle, number or type of lights requires recalibration.

4. Light arrangements

Adjust the lights on both sides and the light on the ceiling to gain suitable lighting. Normally, not all three lights are used; depending on the photographer's lighting requirements, one or two lights might be used.

5. Focus the camera

Adjust the focal length to gain the clearest picture.

6. Photograph objects on a trail basis

After adjusting the lights, use the camera to take a picture on a trial basis. The purpose is to check if the object is at the center of the image, and whether or not light arrangements gain results desired by the execution unit.

7. Photograph objects

Start photographing objects once adjustments are complete.

8. Check image quality

After photographing objects, use Photoshop to open image files and check the image to see if it meets quality requirements; re-photograph the object is there is a problem with the image. Adjust the image when necessary; for example, adjust digital images to make them clearer or sharper.

9. File storage and backup

Store digital image files in a disk array, and upload the files to another data storage center for backup once they reach a certain quantity to obtain remote backup effects. Compare the size and name of image files on both ends once upload is complete, test a few images to make sure the upload was successful and images are correct.

IV. 3D Photography

1. The digitization method described above is mainly for 2D photography. Considering that quite a few relics are solid objects, 3D photography can be used to create images that can fully render entire objects. Using the digital archives program of Academia Sinica Institute of History and Philology as an example, in its collaboration with professor Hung-Yi-Ping, Taiwan Axis 3D Technology Co. and Japan's TEXNAI in 3D imaging, partial images are self-photographed, for which procedures include two major parts: 1. Object photography 2. Post processing – image synthesis. Basic equipment requirements include:

i. Object photography 1

(1)AutoQTVR 3D photography equipment: Object rotating platform (360 degree horizontal object rotation), rotating arm (controls the object rotating device and vertical camera rotating arm), and motor (as shown in Fig.4-1).

(2)Digital camera: High-end DSLR camera.

(3)Lighting equipment: Florescent lights and hood.

(4)Backdrop: Black background paper and background stand.

(5)Computer equipment: Windows 2000/XP, Petium4 2.0G HZ or above, 512MB RAM or above, 64M RAM 32bit graphics card, 80GB hard drive or above, Firewire (IEEE1394) port, USB port, 15" or above monitor.

(6)Software: AutoQTVR automatic control software.

Table 4-1 AutoQTVR 3D photography device

Rotating platform	Rotating arm	Size range	Hardware control interface	Weight limit	Weight	Rotator	Size	Motor	Entire system
Extendable pillar height (50~100cm)	Vertical rotation radius (90~110cm)	20~75(H) x 20~80(W)	RS-232C	Approx. 60kg	Approx. 80kg	110Vx2A	180(D) x 250(H) x 60(W)	Stepping motor	Approx. 5KW including the lighting device
Rotation range (0~360°)	Length of the rotating arm (80~100cm)								
Diameter of the rotating platform {5~15 (standard) ~ 60cm}	Rotation range (+90° ~ -30°)								

Source: The Project of Historical and Cultural Heritages Developed in the Institute of History and Philology, Institute Academia Sinica

ii. Post processing – image synthesis

(1)Computer equipment

Windows 2000/XP, Petium4 2.0G HZ or above, 512MB RAM or above, 64M RAM 32bit graphics card, 80GB hard drive or above, Firewire (IEEE1394) port, USB port, 15” monitor or above.

(2)Software

VR Tool Box (software for synthesizing the final 3D image), Quick Time (software for producing Quick Time VR files), Photoshop (for post processing, such as background removal), Internet browser (for examining results online; installation of Quick Time required).

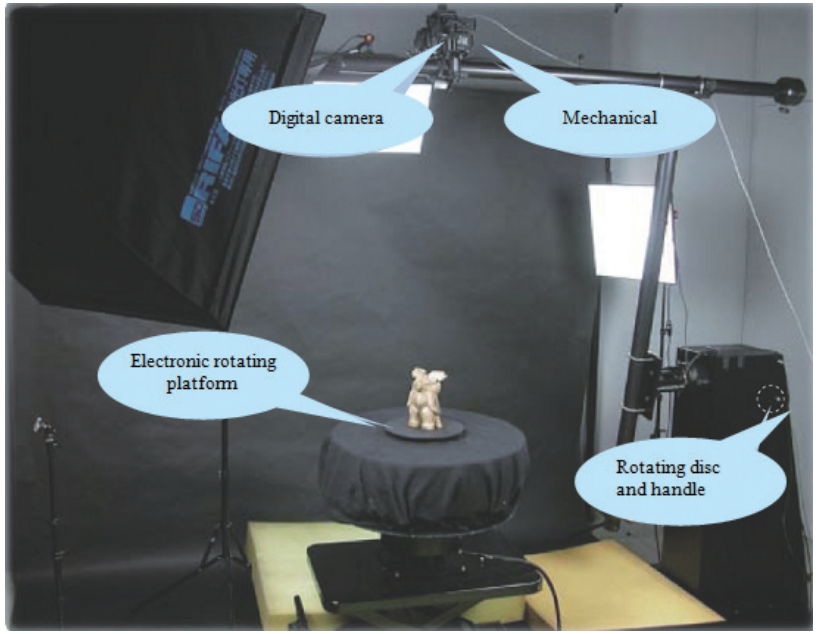


Fig.4-1: Equipment parts include: digital camera, mechanical arm, rotating platform, and manual handle, electronic rotating platform

Source: The Project of Historical and Cultural Heritages Developed in the Institute of History and Philology, Institute Academia Sinica

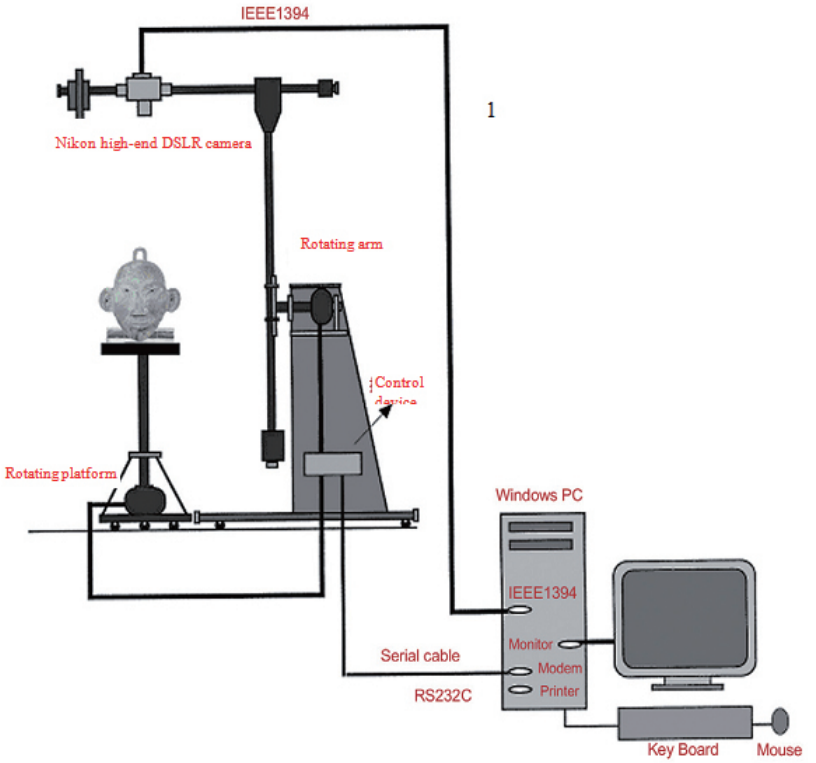


Fig.4-2 Object photography device
Source: The Project of Historical and Cultural Heritages Developed in the Institute of History and Philology, Institute Academia Sinica

2. 3D Photography procedures:

i. Object photography

(1) Position measurement, platform arrangement

A.Use a tape measure to find the center of the rotating platform, and adjust the camera so that it is level with the center of the object being photographed.

B.The safety of the object must be considered when placing and photographing it, especially when rotating the platform. Use auxiliary tools to stabilize the object, e.g. Styrofoam, black flannelette, transparent acrylic stands, and foam (for the convenience of background removal, white foam should be covered with black cloth).

(2) Place the object

A.After protective measures are completed, personnel should take the object and place it on the rotating platform.

B.Start measuring the horizontal, vertical position and center point of the object when placing it. Institution personnel should be responsible for moving the object, while the photographer completes measurements. (One photographer completes measurements, while the other completes computer calibrations and uses the software.)

(3) Trial photograph

The photographer positions the object based on estimates by the naked eye, and then takes a picture of the object. Object Master is then used to measure whether or not the object is centered vertically and horizontally.

(4) Position adjustment

A.The rotating arm still needs to be used to measure the object from different angles to see if it is located at the center of the platform. At the same time, test if the camera collides with the object or blocks the light when it ascends.

B.Object size varies, so the photography position also varies. Sometimes a tape measure needs to be used to measure if the height is correct.

(5) Light adjustment

After all position measurements are complete, the photographer should adjust lighting to display the object's volume and texture. Be careful to prevent the mechanical arm from colliding with or blocking the object.

(6) Place the color calibration card

Place the color calibration card in from of the object and photograph them together. Use this image as a basis for future color calibrations.

(7) Second trial photograph

The other photographer views the image on the computer monitor to determine if lighting is appropriate. If the lighting is not appropriate, the other photographer makes adjustments for another trial photograph

until the image color is close to the original object.

(8) Photograph the object

A.After all preparations are complete, the photographer responsible to operating the computer starts setting the photography angle and number of pictures and presses “start photographing.” At this time the camera and rotating arm will automatically start taking pictures.

B.If during the photography process the software detects an angle wasn’t properly photographed, the rotating platform will return to the previous position and the object will be re-photographed.

C.Avoid unnecessary vibrations or walking around when object are being photographed, the slightest vibration might affect lights and the image.

(9) Changing objects

A.After photographing an object, personnel should return the object to the transport cart and take out another object for photography.

B.If 360 pictures are taken for each object, then one object will require roughly 2 hours, meaning that 3-4 objects can be photographed each day.

ii. Post processing – image synthesis

In order for the full appearance of objects to be accessible online, the numerous images of an object must be synthesized into a 3D image using post processing software. Main 3D imaging procedures are as follows:

(1)Use Photoshop for image background removal (images can only be effectively synthesized if the background is removed), and save images in TIFF format.

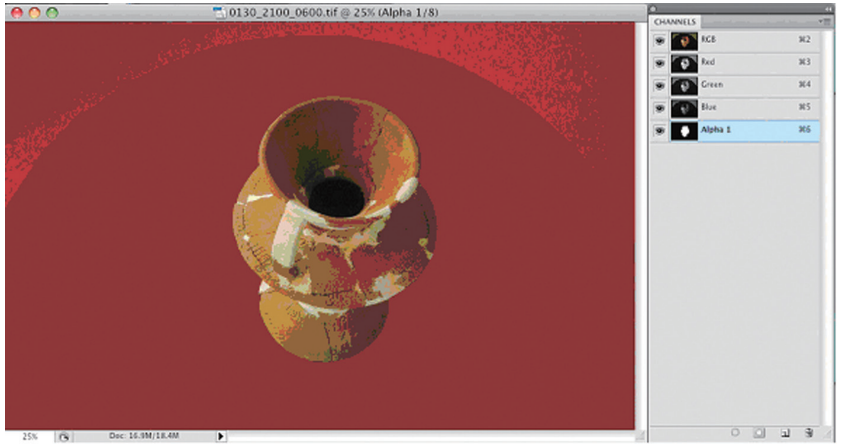


Fig.4-3 Using Photoshop to remove the background

Source: The Project of Historical and Cultural Heritages Developed in the Institute of History and Philology Institute, Academia Sinica

(2)Use VR Tool Box to synthesize the image and output it as a Quick Time VR file (image size should give consideration to network bandwidth).

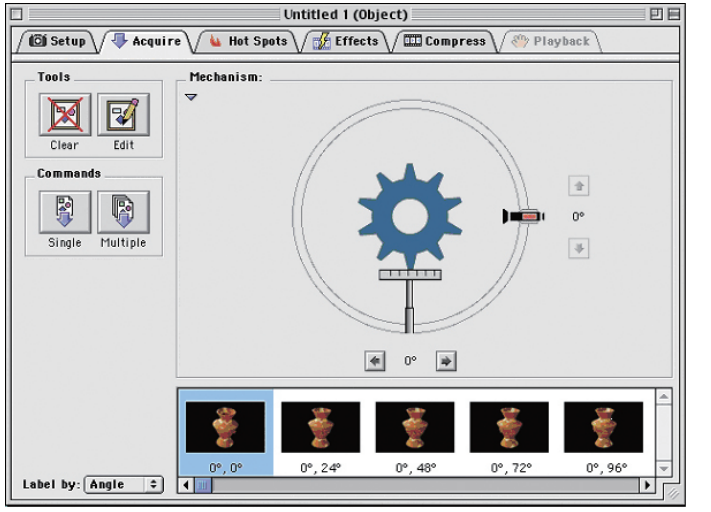


Fig.4-4 Image synthesis

Source: The Project of Historical and Cultural Heritages Developed in the Institute of History and Philology Institute, Academia Sinica

(3)Use your browser to examine if the image is correct.

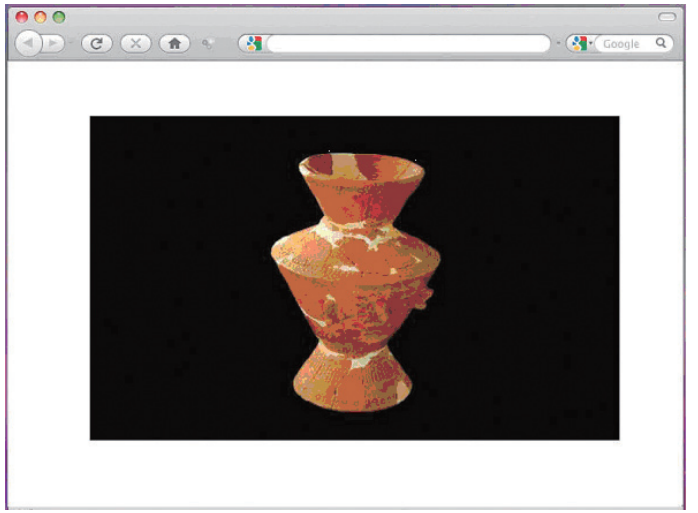


Fig.4-5 Examine image results
Source: The Project of Historical and Cultural Heritages Developed in the Institute of History and Philology Institute, Academia Sinica

For institutions with limited funds and collections of relatively small solid objects, a manual rotating platform and synthesis software can be used as an alternative. Together with the original digital camera and lighting equipment, 2D photos can still be taken every fixed number of degrees and then synthesized in a 3D image. Figures 4-6 and 4-7 show 3D photography equipment used by Chang Jung Christian University in the "Pingtung County Paiwan Tribes Traditional Arts Relics Digital Archives Project".[13]

[13] "Ceramics Digitization Procedures Guideline" by Chen Hsiu-Hua, Tsai Hsing-Chen, and Kao Yu-Ju, TELDAP Taiwan Digital Archives Expansion Project, March 2009, page 35.

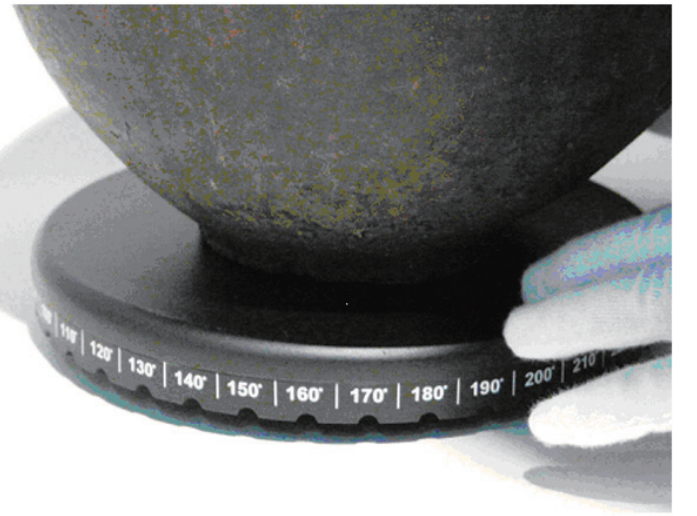


Fig.4-6 Manual rotating platform
Source "Pingtung County Paiwan Tribes Traditional Arts Relics Digital Archives Project"

Fig.4-7 3D photography of solid objects
Source "Pingtung County Paiwan Tribes Traditional Arts Relics Digital Archives Project"

3. Other – Producing a 3D model of the object restored

Relics excavated from archeological sites are mostly shattered pieces; using 3D technology to virtually repair these relics is not only another type of digital archive technology, but also benefits academic research by providing the full appearance of objects. Using the National Museum of Natural Science Archeology Department as an example, ceramic pieces with research value are selected based on digital archive requirements for 3D model repair. Use 3D MAYA professional graphics software to establish curve models of pieces. (Fig.4-8)

Fig.4-8 Curve model production
Source: National Museum of Natural Science
Anthropology Department Archeology Division

Next, compare the 3D image with pieces of the physical object from different angles and make adjustment.

Fig.4-9 Adjust different angles of the object
Source: National Museum of Natural Science
Anthropology Department Archeology Division

Furthermore, simulate special decorations, if there are any.

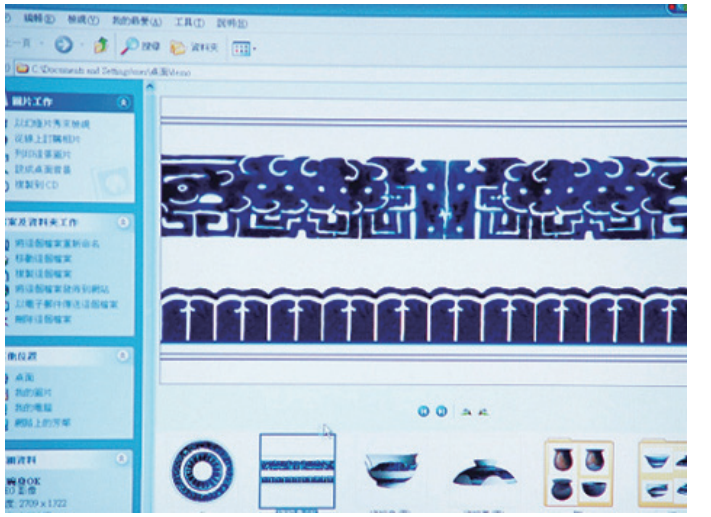


Fig.4-10 Simulate the decorations
Source: National Museum of Natural Science
Anthropology Department Archeology Division

Finally, view the complete picture of the 3D model, print it out on paper, and then save the file for import into the archives management system, allowing it to be provided for multimedia teaching and research.

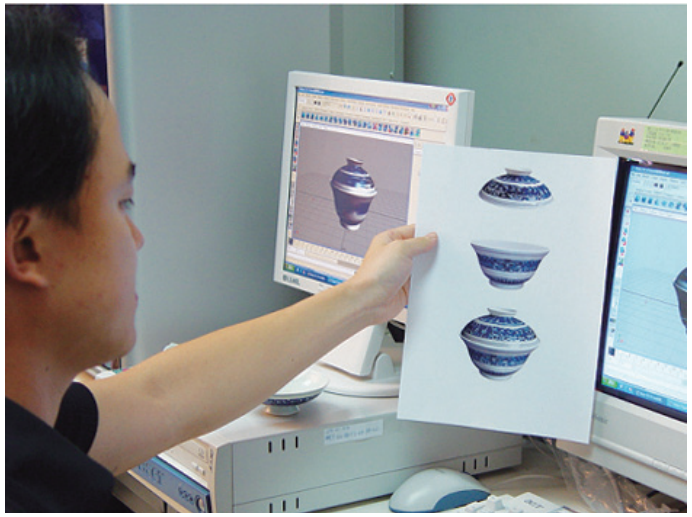


Fig.4-11 Output a sample
Source: National Museum of Natural Science
Anthropology Department Archeology Division

FIVE. Metadata and Database Establishment

Object digitization is a complicated process in the digitization workflow, in which another foundational task is "metadata and database establishment." Depending on the type of collection being digitized, different metadata standards may be adopted. Selection of metadata categories aims to help users fully understand the contents of collections.

I. What is Metadata?

Besides creating digital images of collections for preservation, digital archiving work also includes text descriptions of objects for the convenience of file management and database search. These text descriptions are what we call metadata, which is "data about data."[14]Metadata is used to describe contents and characteristics of digital archives, allowing digital archives to be effectively found and displayed in a digital environment, at the same time interoperable with data of other institutions so that resource sharing can be achieved. Numerous metadata standards have been developed by domestic and international academia for different fields. For example, there was a metadata standard adopted for biology called the Darwin Core, and was used to describe digital biological data; for archives there is the international metadata standard Encoded Archival Description (EAD).

From the paragraph above we can clearly know that there isn't a fixed standard for metadata establishment, and that metadata can widely vary. Due to different fields, collection quantity of institutions or individuals, professional understanding of collections, and collection conditions and purposes, there are different requirements on metadata establishment. For instance, in order to establish metadata for museum collections, one must first understand the types and characteristics of museum collections, as well as museum user requirements. However, when setting metadata fields for collections of private institutions or individuals, then simpler metadata fields can be created to best suit their needs.

Metadata standards suitable for archaeological sites, remains and artifacts are not common. Considering that overseas institutions have a relatively early start in digital archiving work and establishing digital knowledge bases, international metadata standards for archaeology, art history and art collections can be used as reference for establishing metadata for archaeological data. Such standards include MIDAS (The Monument Inventory Data Standard) developed by England's The National Monuments Record (NMR); The CIDOC Information Categories of the International Committee of Documentation, which is under the International Council of Museums (ICOM); the VRA Core Categories constructed by the Visual Resources Association (VRA); and CDWA (Categories for the Description of Works of Art), which was developed and is maintained by the Getty Research Institute.[15]After referring to the international standards listed above, domestic archaeology research related institutions can establish metadata fields that best suit their needs. The metadata fields of Academia Sinica Institute of History and Philology Archaeology Department (please see Appendix 3 for details) can also be used as reference for metadata establishment.

II. Archaeological Metadata Establishment

1. Architecture and categories of Archeodata

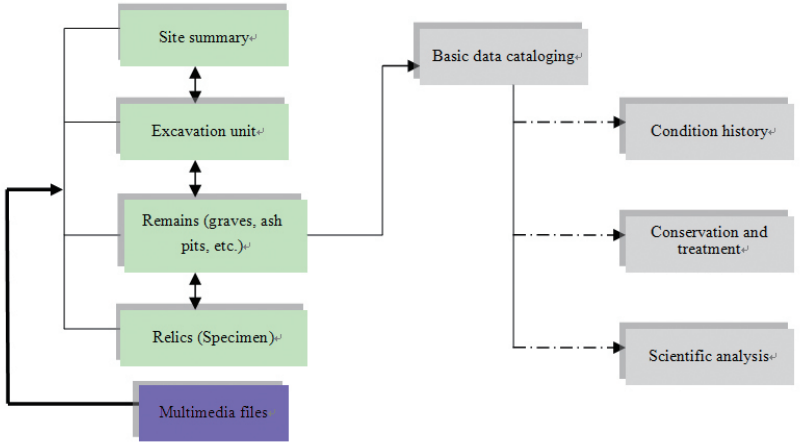
The most important task of database establishment is to create metadata fields that satisfy similarities between different collections, while making convenient for users to search for information. Refer to current domestic and international standards to establish metadata fields that best meet requirements. Using the Project of Historical and Cultural Heritages Developed in the Institute of History and Philology Institute, Academia Sinica as an example, its archaeological database was constructed based on overall considerations of archaeological excavations.

Archeodata is a system with four layers (site, excavation unit, remains, and artifacts) linked together, that further establishes essential information of archaeological work – context. Main work items include archive management, exhibition, repair, maintenance, regular preservation records, scientific analysis, and research. In terms of data content the archaeological data system can preserve a variety of data, including photos, illustrations, rubbings, field records, maps, and display them according to their properties, such as site, remain, or artifact. In terms of media type, text records, images, sounds, documentaries, 3D images can all be included in archaeological data systems.

[14]"A Preliminary Study on Metadata," by Chen Ya-Ning and Chen Shu-Chun, "Academia Sinica Computing Center Newsletter" volume 15 issue 5, March 1st, 1999.

[15]Academia Sinica Institute of History and Philology Archeology Division, "Digitization Procedures – Archeology Thematic Group," Taipei: Contents Development Division, 2004, page 75.

Table 5-1 Architecture of Archeodata[17]



Archaeological data can be divided into cultural artifacts, ecological remains, anthropological remains (human bones), and geological remains. Cultural artifacts are extremely similar to museum collections, such as archaeological data in the Institute of History and Philology, and jade, stone, ceramics, and bronze ware collected in the National Museum of History. Other categories, ecological remains, anthropological remains (human bones), and geological remains are unique to archaeological data. First establish core categories when developing archaeological databases; when describing other archaeological data, first describe their basic characteristics in core categories, and then select specialized categories accordingly. Core categories of Archeodata are as shown in the table below.

Table 5-2 Core categories of Archeodata[18]

Archaeodata Categories	Function group in the digital archives system
CATALOGING INFORMATION	Internal system information
SITE	Basic information
PIT	
FEATURE	
BOJECT TYPE	
TITLES OR NAMES	
CULTURE/ PERIOD	Culture/Period
DATING	Dating
PHYSICAL DESCRIPTION	Artifact description
DECRORATION	
FUNCTIONAL CLASSIFICATION	
MATERIALS	
TECHNIQUE	
SUPFACE MARKS	
INSCRIPTIONS/MARKS	Inscriptions/Marks
SET & PIECES	Set records
REPOSITORY	Repository
ACQUISITIONS	
MEASUREMENTS	Measurements
CONDITION/EXAMINATION HISTORY	Condition/Examination History
CONSERVATION/TREATMENT RECORD	Conservation and treatment record
SCIENTIFIC ANALYSIS & RESEARCH RECORD	Scientific analysis

2. Categories of multimedia files

Archaeology field work always produces great quantities of data, organizing and analyzing this data is an important link to producing research results. Specimen illustrations, important decorations and engravings, X-ray analysis of copper ware structure, and infrared photography all require conversion into digital files via digital scanning or photography. Various files then need to be converted into formats that can be circulated on the internet, e.g. the *.jpg format. Files for online browsing do not have as much information as original files, and therefore have different space for value-added applications. Therefore, following the development of digitization methods and preservation technology, it has become necessary to include information of the original object in metadata fields. In the light of this, categories for multimedia files can be divided into two areas “original object information” and “digital file information,” and include the digitization method (Table 5-3):

[17] Same as 16.

[18] Same as 16.

Table 5-3 Multimedia file categories of Archaeodata

Digital file information		Original object information	
Number		Number	Number
Registration number/Code			Type
Image description		Media type	
Data type		Format/Dimensions	
Business type		Archive information	Unit number
		Unit name	
Whole/Partial		Collection location	
Digitization Method	Type	Photography/ Illustration information	Name
	Brand		Form
	Specification		Date
File quality level		Condition	
File size			
File format			
Digitization Method	Gray scale		
	Color mode		
	Color depth		
Digital file creator	Name		
	Form		
	Date		
Archive information	Unit number		
	Unit name		
	Collection location		
Limited access			
Priority			

III. Metadata Cataloging and Database Establishment Procedures

1. Requirements evaluation and content analysis

Metadata analyzers interview institutions or individuals about to implement digitization projects to understand project properties, goals and metadata requirements. Next, they analyze related metadata standards and application cases for the project execution unit to understand metadata types used by similar projects, and then they use work forms (e.g. metadata elements requirements, metadata element codes, metadata cataloging examples) to more precisely analyze the institution's metadata requirements. After arriving at a conclusion, they recommend metadata that best meet the institution's requirements.

2. Metadata functional requirements and database specifications

Metadata functional requirements and database specifications serve three purposes:

i. To allow institutions to verify their metadata functions and database specifications.

ii. To serve as a basis for communication between institutions, the metadata task force, and system designers.

iii. To serve as a basis for institutions to modify metadata functions and database specifications.

After specifications are completed, evaluate the development potential of the metadata system and database, so that institutions can decide whether to adopt a system and database of a similar project, to develop new ones by themselves, or jointly develop new ones with other institutions. [19]

3. Metadata system development and database establishment

During system development the institution, metadata task force, and system developers should continue to meet for discussions and exchange opinions. After a prototype of the system and database is complete, the institution and metadata analyzers should conduct tests and provide results to system developers as reference for making adjustments. [20]

4. Evaluation, maintenance and update

Based the institution's requirements, metadata evaluation items should include: quality of metadata, search benefits of the standard that was adopted, and usability of the tool for generating metadata; the purpose of evaluation is to enhance the service quality of the metadata mechanism. In addition, the database should be updated on a regular basis; designated personnel should be responsible for maintaining the database's stable operation.

[19] The metadata requirements and evaluation procedures above referred to the website of the "Metadata Architecture and Application Team" (http://metadata.teldap.tw/index.html).

[20] "Introduction to Digital Archive Technology," by Tsai Yung-Cheng, Huang Kuo-Lun, and Chiu Chih-Yi et al., Taipei City, NTU Publishing Center, November 2007, pages 68-72.

SIX. Digital Content Protection

"Digital content" is the product of using information technology to digitize images, text, video, and audio. The purpose of digital archives is to show precious collections to the world without hurting the original object, achieving the purpose of research, promotion and education. Digital archiving results should be shared with the general public as much as possible. Following the rapid development of technology, large quantities of text, sounds and images can easily be converted into digital files, but these digital files can also easily be illegally downloaded, copied or tempered with. These actions infringe copyrights and creativity, and can only be avoided via digital content protection and rights management, which will help creators maintain their motivation for making creative works. Therefore, below we will introduce common digital content protection methods, including "digital watermark," "CC licensing," and "Digital Rights Management."

I. Digital Watermark

Digital watermark is a technology developed to protect intellectual property rights of digital content, and proves the legitimate owner of creative works by embedding the creator's name, owner's name, issuer, trademark, or address into digital files. Depending on the file format, digital watermarks may be text, static images, dynamic images, or audio signals. From the appearance, digital watermarks can be divided into visible watermarks and invisible watermarks. The former refers to watermarks directly visible on digital files; its disadvantage is that it damages the original image and makes the image lose its value, but the advantage is that this intimidates people and prevents them from illegally using the image. The later refers to watermarks not visible to the naked eye; images embedded with invisible watermarks look entirely the same, maintaining image quality, but require special techniques to identify the watermarks.[21]

In the experience of Academia Sinica Institute of History and Philology Archaeology Department, a watermark LOGO is first designed (visible), the position of the watermark is determined for each digital image, and then the Multimedia Center uses an automated process to embed the watermark. After digital files are completed, they are uploaded in batches to the Computing Center (Fig.6-1). To avoid affecting image quality, visible watermarks of lower density are added (lower right of Fig.6-1); the provider of the original file is also noted on the image (upper left of Fig.6-1).

Fig.6-1 Image of vase with a watermark on the lower right corner
Source: Academia Sinica Institute of History and Philology
"Archeodata" http://archeodata.sinica.edu.tw/

Digital watermarks are currently the most widely adopted digital content protection technology by digital archive projects, but like any other encryption technology, watermarks are still at risk of being undone by others, so special attention must be paid to encryption quality. Double encryption provides more protection than single protection; besides a visible watermark, an invisible watermark is also added to the digital image. However, higher protection levels also result in poorer image quality.

II. CC (Creative Commons) Licensing

In this age of booming digital technology, people can share text, images, video and music around the world. Most digital content creators who publish their works online have the intention to share works with the general public. Due to the openness of information, users can duplicate, reproduce, and even modify the original work

[21] Chia Hsin-Chieh, "What is a watermark?" "National Digital Archives Program Newsletter" http://www2.ndap.org.tw/newsletter06/news/read_news.php?nid=732. Search: January 2010.

at will. However, most users often overlook the importance of listing the source, or make modifications out of the reasonable range. The lack of a clear license label might result in the infringement of the creator's rights. Creative Commons licenses provide a simple and legally effective way for creators to specify the scope of which their creative works may be freely used, while retaining partial rights; on one hand allowing users around the world to use their works or create derivative works, on the other hand retaining rights the copyright owner wishes to retain.

The Creative Commons licensing terms announced in 2002 were developed based on laws of the United States. Therefore, Creative Commons is concerned that digital content must be usable by people in different countries and regions. Creative Commons has four licensing conditions: [22]

1. Attribution

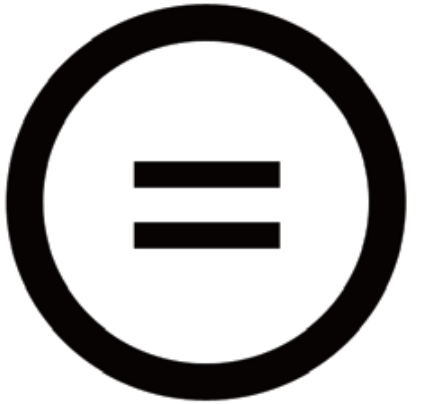
You let others copy, distribute, display, and perform you copyrighted work and derivative works based upon it, but only if they give credit the way you request. As to how to "attribute," there are the following notices:

If the creator provides the following information, then it should be verbatim copied on the media or tool used:

[22] From "CC-Creative Commons Taiwan" (http://creativecommons.org.tw/), Search: December 2009.

i. The copyright statement.
ii. The name or penname of the creator, licenser or a third person designated in the copyright statement or licensing terms.
iii. The name of the creative work.
iv. The website of the creative work provided by the creator or licenser.

2. No Derivative Works



You let others copy, distribute, display, and perform only verbatim copies of your work, not derivative works based upon it.

3. Non-Commercial

You let others copy, distribute, display, and perform your work, but for non-

commercial purposes only.

4. Share Alike

You allow other to distribute derivative works only under a license identical to the license that governs your work.

At present, CC provides several licensing terms for users to choose from. Using the Taiwan version CC licenses as an example, if users choose the license "Attribute – Non-commercial – No Derivative Work," anyone who wishes to use digital content protected by the CC license will have to credit the copyright owner.

Some project results shown in the TELDAP portal currently use CC licensing. Please visit the Creative Commons Taiwan website for different licenses. To complete licensing for your own work, follow the procedures on the website and add the system generated code to your own website.

*CC website: http://creativecommons.org.tw/
*CC video introduction: http://creativecommons.org.tw/cc_intro_anime

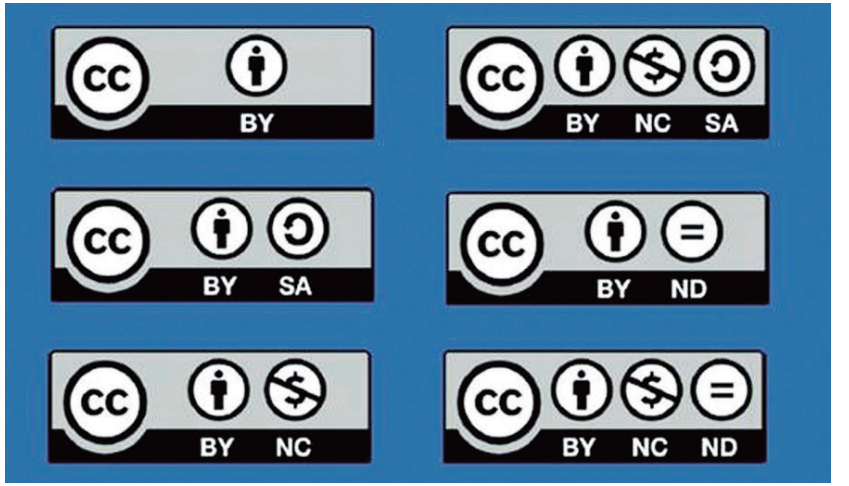


Fig.6-2 Six combination of license conditions

III. Digital Rights Management (DRM)

DRM technology is a digital content protection and management mechanism achieved via file encryption and integrates hardware with software. The Internet Data Center (IDC) defines DRM as: "The chain of hardware and software services and technologies confining the use of digital content to authorized use and users and managing any consequences of that use throughout the entire life cycle of the content." [23]

To protect the rights of digital content creators and providers, set access rights of digital files after digital content is produced, e.g. limiting the number of times a file can be read, saved, copied, transferred, and recorded, or the number of times a file can be played, or whether or not it can be copied, printed, expiration date, or requiring a password to open the file, or if the file is read only. Some restrictions even require users to open the file with specific software or hardware. Using the Multimedia Center system as an example, which was developed by the digital archives technology development division of Academia Sinica Institute of Information Science, users are required to acquire an account and password to login the system. When users click on a file they wish to download, the system will tell

[23] Dahl Joshua and Kevorkian Susan, "Understanding DRM Systems", An IDC Research White Paper, 2001.

the user the file's access rights and certification methods. After passing certification for digital content, the file can be played in the user's computer, during which the system will also notify the user of the file's usage limitations (e.g. number of times it can be played, can it be added, can it be printed, etc.).

Fig.6-3 Multimedia Center
(http://ndmmc2.iis.sinica.edu.tw/System/Index.jsp)

Therefore, this content protection and management technology (DRM) allows creators or copyright owners to set access rights of their digital content, e.g. download, output, copy, modify, etc. However, although usage limitations protect the digital content provider's rights, they require the use of specific software or hardware to open the file. Digital content possess great value, which results in high license prices. This is the foundation of economic operations, but it discourages circulation and utilization, and lowers users' willingness to use the content, raising quite some debate.

SEVEN. Equipment and Cost Analysis

In this section we will explain the selection of equipment for digital photograph. Equipment used by digital archive projects include: digital camera, lighting, computer system, color management, and storage equipment. Depending on the value of the collection or utilization purpose, different digitization methods may be adopted (e.g. 2D or 3D photography, but the later is not adopted for all collections), resulting in different equipment and cost.

I. Equipment

Equipment used by digital photography of archaeological data can be divided into digital camera, lights, computer system, color management, and storage equipment:

1. Digital camera

Considering future applications of digital images, DSLR cameras or traditional medium to large cameras used with digital camera backs are recommended. Not only do they allow the lens to be changed, but also more detailed adjustments to aperture, shutter and view depth, which produces better image quality.

i. Digital camera back:

This is the combination of a medium to large format (120 to 8x10) camera and a digital camera back used as replacement of film; prices can range from hundreds of thousands NTD to several million. The most fundamental consideration of selecting digitization equipment is to suit the usage purpose. Cameras with higher resolution should be selected to produce images of better quality or larger size. Digital camera backs currently in the market all have at least 20 to 30 megapixels.

ii. DSLR camera:

This refers to typical cameras with built-in CCD and allow lens changing; prices range from NT$30 thousand to NT$200 thousand. For institutions with limited funds, this is also an option. Basically, the larger the size of CCD or CMOS in a digital camera, the more color information it can process, producing more effective pixels and lower noise. Therefore, when selecting a digital camera, CCD or CMOS size is a major consideration. At present, DSLR cameras equivalent to the traditional 135 full frame camera (24mm×36mm) can now be found in the market; focal length is doubled after putting on the lens, more details can be captured in shadows, and larger digital image files can be produced, making it suitable for more applications.

2. Lights

Light sources used in the photography studio can be divided into two types of light sources, continuous and flash:

i. Flash light source

Flash light

This is the most commonly used artificial light source by photographers, especially when outdoors. Flash lights are close to natural light because their color temperature is roughly 6000K, which doesn't produce color shift. Flash lights produce little shadow because they generate instant light. Regular batteries can be used (110V) without special need for a transformer. Flash lights are suitable for photographing people.

ii. Continuous light source

(1) Cold light

Fig.7-1 KinoFlo cold light

Unlike tungsten lights, which are spot light sources, cold lights are plane light sources, which allow objects to be evenly illuminated. When used together with a high performance high frequency electronic stabilizer cold lights can generate stable lights. The low power consumption and low heat of cold lights save cost on power. The life span of cold lights is between 7000~10000Hrs, in which color temperature is maintained between 5000~6000K.

(2) Tungsten light

Tungsten lights are a common light source in the photography studio,

and come in two color temperatures, 3200K and 3400K. Tungsten lights are continuous, direct, heat light sources, color temperature changes along with power, and light bulbs have relatively short life spans, ranging between 50~1000Hrs. The main disadvantage of tungsten lights are the high heat they generate.

iii.Quartz light

Also known as quartz halogen lights, quartz lights can be spot light sources or linear light sources, and are common artificial light sources in the photography studio. Quartz lights are relatively small in size and are also heat light sources (but with relatively low temperature) with color temperature maintained at 3200K.

To reduce the effect of artifacts being under high temperature and ultraviolet rays for long time periods when photographing archaeological data, cold lights are recommended for lighting.

Table 7-1 Comparison of flash and continuous light sources

	Continuous light source (cold light)	Flash light source (flash light)
Advantages	1. Low power consumption 2. Long life span 3. Stable color temperature 4. Low temperature 5. No waiting time recharge required 6. High brightness 7. Shock resistant	1. Strong brightness and short flash time allows the capture of clear images of moving objects. 2. Low temperature suitable for photographing people and products. 3. Color temperature close to daylight, stable hue, and allows the use of daylight color film. 4. Suitable size, easy to carry, and easy to use.
Disadvantages	High price	1. Besides large flash lights that have simulation lights, brightness and results cannot be predetermined before objects are photographed. 2. Typical light meters cannot measure the instant brightness of flash lights. 3. Requires waiting time for it to recharge.

3. Color management system

The main purpose of color management systems is to make different output devices output the same colors. Digital image color management is mainly divided into three stages: image input, image processing, and image output. These three stages all go through different media, such as:

i.Image capture: Digital camera (Phase One P45+, Imacon IXPRESS 528C, NIKON D300, etc.) and scanner (Imacon Flextight 949, etc.)

ii.Image processing: Adobe Photoshop (version 7 or above required for processing images with 16 bit color depth).

iii.Image output: Monitor (EIZO CG211), ink-jet printer, photo output machine, and printing factory.

Different devices have different color characteristics; each stage needs to be included as a part of a complete system for color management to be truly implemented.

Color management is divided into three stages; each stage has three important procedures:

i. Calibration:

Different equipment have different calibration methods; for example, scanners and digital cameras are associated with gray scale balance (gray card: X-rite White Balance Card) or color gamut calibration (color card: GretagMacbeth Semi-Gloss)[24], while printers involve ink and concentration settings. For monitors, color temperature and gamma value need to be adjusted; the color management software ProfileMaker and spectrophotometer Eye-one Pro are used to calibrate the monitor so that colors displayed match standard colors.

ii. Characterization:

The difference between colors of different devices is stored by the color management software in an ICC Profile, which describes characteristics of device.

iii. Conversion:

When ICC profiles are created for all equipment in the digitization process, color values of the source image can be converted into values of the destination device, which is called color conversion.

4. Computer system

Computer systems can be divided into software and hardware. Hardware is typically divided into personal computers, workstations and servers, while software refers to different operating systems.

[24] GretagMacbeth: www.gretagmacbeth.com/index.htm. Search: January 2010.

Table7-2 Computer system ahalysis chart

	Personal Computer	Workstation	Server
Operating system	Apple's Macintosh OS9, MAC OS X; Microsoft's Windows XP, Vista, 7; Other, e.g. Linux; UNIX	Apple's Macintosh OS9, MAC OS X; Microsoft's Windows XP, Vista, 7; Other, e.g. Linux; UNIX	Apple's MAC OS X Server; Microsoft's Windows Server; Other, e.g. Linux; UNIX
Functionality	Daily operations, such as text processing	Image processing and 3D animations	Possesses high processing ability, which allows it to process interactive information of multiple users at the same time, e.g. database operations, file management and website functions
Processing resources	Typical	High	Massive
Price (NTD)	<50,000	50,000~100,000	Several tens of thousands for personal servers, and up to tens of millions for large enterprise servers

Computer equipment used for digital photography is as follows: [25]

i. Computer connected to the camera back: MAC Pro 2.66, 4GB RAM, HD 500GB, EIZOCG211, UPS, and high speed network connection.

ii. Backup: File server, RAID, UPS, and high speed network server.

iii. Calibration, post processing: MAC Pro 2.8, 4GB RAM, HD 500GB, EIZO CG241W, UPS, and high speed network connection.

iv.For processing large images, workstations with 2GB RAM or above and a monitor with built-in color management functions, e.g. EIZO CG series, are recommended.

5. Storage equipment:A brief list is as follows:

Table7-3 Storage Equipment

Equipment	Description	Image
Tape Cartridge	Used to backup large quantities of data; divided into tape driver and tape cartridge.	
Disk Array RAID	Used to store digital images. Multiple hard disc drives are combined into a disc array that performs better than a single hard disc with huge storage capacity. RAID has following the advantages over a single hard disc: enhanced data integration and error tolerance. RAID is divided into RAID-0, RAID-1, RAID-1E, RAID-5, RAID-6, RAID-7, RAID-10, and RAID-50.	
Recorder	Use a recorder and DVDs to backup photographed images.	

6. Other equipment

i. Camera stand: used together with the digital camera.

ii. Background stand, background paper, and hydraulic lift: used during photography.

iii. Copy stand: if a digital camera back is not used for 2D photography, then this is used along with the digital camera.

iv. Light meter: used to ensure color shift doesn't occur due to lighting or color temperature.

v. Level: used to prevent deviation of the image.

There are quite a few options for digitization equipment in the market, so equipment selection should give consideration to the project institution's budget, individual operation ability, and scope of future applications. Before purchasing equipment, besides personally seeing the product, ask vendors to test the products for you so that unsuitable products aren't purchased. [26]

Source of the equipment listed above: The Project of Historical and Cultural Heritages Developed in the Institute of History and Philology Institute, Academia Sinica

II. Cost Analysis

1. Composition of cost

[25] Source: The Project of Historical and Cultural Heritages Developed in the Institute of History and Philology Institute, Academia Sinica

[26] Also refer to "Ceramics Digitization Procedures Guideline" by Chen Hsiu-Hua, Tsai Hsing-Chen and Kao Yu-Ju for information on the equipment listed above, Taipei City: Taiwan Digital Archives Expansion Project, April 2009.

Costs required for archaeological data digitization work include: material cost, labor cost, and miscellaneous costs:

i. Material cost is the cost of consumables used for digitization work.

ii. Labor cost is mainly the salaries of project execution personnel.

iii. Miscellaneous costs can be divided into direct costs and indirect costs:

(1)Direct costs include costs and amortization expenses of information equipment, mechanical equipment, and cost of information software.

(2)Indirect costs include amortization expenses or rent, renovation cost, utilities and other.

The cost analysis above should only serve as reference for cost calculation. Equipment and human resources should be based the actual budget of institutions, or the purpose of digitization results. Cost estimation can also be used as a basis for calculating licensing fees in the future.

2. Cost estimation

Calculation method:Cost calculation can be divided into two methods based on equipment amortization:

i. Service life

[Labor cost (NTD) + Equipment amortization expenses (NTD)] / Digital output (Pages) = Cost per page (NTD/Page)

A.Labor cost is mainly the salaries of project execution personnel.

B.Equipment amortization expenses = (Equipment cost + Software cost – Remaining value) / Service life

ii. Total digital output

Labor cost (NTD) / Digital output (Pages) + (Equipment cost + Software cost) / Digital output (Pages) = Cost per page (NTD/Page)

Furthermore, renewing backup data and storage equipment is another cost associated with the long-term preservation of digital data; other factors that should be considered include the risk of data being damaged, human resource management and other intangible costs.

EIGHT. Outsourcing

The funds, equipment, human resources and time invested in digitization work often affect the quality of digitization results. Institutions can consider outsourcing all or part of digitization work based on their own conditions. Outsourcing is defined as "To authorize an external supplier or service provider to handle management function(s) along with associated assets within an agreed time period for a price agreed by both parties (with a proviso)".[27]From a management perspective, under the premise of not increasing existing human resources, outsourcing is another option for digitization work.

Some institutions cannot afford the expensive equipment required for digitization, or are worried about equipment maintenance and update. Under such circumstances, outsourcing is an option that reduces work time and effectively achieves cost requirements; it is indeed a way to reduce human resource requirements in an institution. Besides object digitization, database establishment, hardware and system maintenance, and even quality examination systems are technical problems that institutions often need to outsource. When institutions consider outsourcing, they should begin planning specification establishment, selection of a suitable contractor, communication with the contractor, and results acceptance procedures. In this chapter we will briefly explain outsourcing procedures and matters that need attention.

I. Before Tender Invitation

1. Outsourcing project evaluation and analysis

Before deciding to outsource a project, institutions should first determine whether or not the project is suitable for outsourcing based on considerations of human resources, budget, and operation conditions. Evaluate the benefits of outsourcing before deciding on which outsourcing model to adopt. Besides human and material resources in the institution, potential costs should also be considered during the preliminary evaluation, such as mistakes, personnel circulation, and communication issues.

2. Planning outsourcing projects

After thorough evaluation and analysis, once you decide to outsource a project, begin establishing specifications of requirements. Specifications should include budget, file format, quantity, quality standard, acceptance process, workplace, and even confidentiality clauses.

[27] Kao Chih-Tung, Lin Fang-Chih, Chen Hsiu-Hua, Chen Mei-Chih, "Digitization Procedures Guideline: Outsourcing Management", Taipei: Taiwan Digital Archives Expansion Project, April 2009.

II. Tender Selection

Institutions must first understand potential issues that must be dealt with when outsourcing projects. This is to avoid difficulties in the tender process, especially when government invitations for tender are required to follow specific procedures and regulations. When selecting a contractor, base your decision on cost, the contractor's technology, scale, operation method, financial condition, and past work records; you can also ask for reference from institutions that cooperated with the contractor in the past. The project can establish a selection team to be responsible for evaluating, discussing and deciding on a winning contractor. To prevent inappropriate outsourcing from damaging precious objects or archaeological data, institutions can list standard procedures in the announcement, or even ask contractors to perform demonstrations.

III. After Tender Selection and Execution

1. Outsourcing project execution

After an outsourcing project begins, institutions should contact and communicate with the contractor on a regular, making adjustments whenever problems arise. For example, matters that need attention when moving or photographing objects, and how the database should be established or how it should be presented all require communication with the contractor. Institutions should also inspect progress on a regular basis, the inspection process should be based on the contract, and contractors should be asked to make corrections whenever errors are found. Effective and conscientious communication benefits the work efficiency of both parties.

2. Project conclusion and feedback

Records of the outsourcing process are important data; whether the contract was terminated in advance or successfully completed, the entire process should be kept on record. Records include important events, goals, contents, execution performance and analysis, and even letters, meeting minutes and technical documents; these should all be provided in the project report. After a project is concluded, maintenance of technical aspects, such as the information system, should under warranty, and the contractor should provide necessary training to

institution personnel.

However, outsourcing isn't a perfect implementation method, and some institutions decided to implement digital archive projects themselves because of issues encountered during outsourcing. Miss Lin Yu-Yun of Academia Sinica Institute of History and Philology indicated that during early stages of its digital archive project a great amount of work was outsourced in hopes that digitization results could rapidly achieve cost effectiveness and be published, but after encountering numerous issues they decided to implement digitization themselves. Issues encountered by the Institute of History and Philology are common during early stages of many projects, including (1) Insufficient experience in equipment selection and poor after-sales service quality; (2) Poor digitization results as a result of not following standard operating procedures, e.g. unfamiliar with color management and printing procedures; (3) Communication problems during the photography process between the contractor, photographer and personnel; and (4) Time cost was hard to estimate because institution staff needed to perform conservation work on artifacts during the photography process.

Digitization is a long-term task, after considering budget planning and annual tender invitations, self-production is sometimes more economic than outsourcing. When the number of objects requiring digitization reaches a certain quantity and the safety of objects cannot be ensured in the photography environment, institution personnel will be required to make safety precautions, making it difficult to calculate time cost. Self-production eliminates management issues associated with the contractor, and quality is easier to control. Another advantage of self-production is the accumulation of knowledge and experience of institution personnel, which benefits further improvement of digital quality. For details on matters that need attention in the outsourcing process, please refer to "Digitization Procedures Guideline: Outsourcing Management".

NINE. Benefits and Prospects

The main purpose for writing this "Archaeological Data Digitization Procedures Guideline" is to provide public/private institutions and those interested in archaeology with digitization related knowledge and technologies, as well as specific procedures of digitization. The writing process of this book was still at the exploratory stage. We hope to gather the opinions of more scholars and specialists on this basis, to gradually expand the width and depth of contents, and make this guideline more mature. Benefits we hope to achieve or areas we hope to strengthen are as follows:

I. Present the workflow and provide detailed procedures

We hope that this guideline can present the complete digitization workflow, so that institutions or individuals with the intention to become engaged in digitization work can have an overall understanding of concepts associated with the entire process. We also provide a simple and clear reference standard, so that they can effectively carry out digitization work by referring to standard procedures.

II. List costs and equipment for selecting the most suitable combination

We have listed specifications, equipment performance, human resource allocation, cost estimation, and budget utilization, in hopes that digitization personnel at different levels can select the most suitable digitization method and equipment based on their condition. Due to the fact that main digitization equipment, such as scanners, computer equipment and storage media, are advancing at rapid speeds, the equipment listed in this guideline can only serve as reference, the latest technologies or products should be your main options when purchasing equipment.

III. Promote and introduce management systems

Effective management of the digitization workflow, maintaining a smooth workflow, and establishing a digitization workflow management system is truly necessary. Not only do they reduce manual operating procedures, ensure the integrity and accuracy of data, but also enhance workflow management and progress control via system analysis. We believe that paper copies or electronic lists will gradually be replaced. Therefore, we hope that in future revisions of this guideline a section of management systems can be included.

IV. Provide an exchange platform to facilitate experience sharing

Digitization procedures recorded in this guideline are based on investigations and interviews with institutions implementing digital archive projects, and then referring to experiences of other thematic groups when digitizing similar objects. This guideline provides a standard so that institutions will not be at a loss as to what to do. Institutions or personnel currently engaged in digitization work can use this guideline and their own workflow for evaluation, comparison, and finding techniques to make up for their weaknesses, which will enhance their digitization work efficiency. However, this guideline has its limitations, not all special requirements were covered in discussions, so not all projects are applicable.

Although this guideline is still limited, we still hope to continuously make improvements. In terms of technologies, revisions of this guideline can introduce the latest equipment and technologies to those who need it. From another aspect, the opinions and responses of reader will help refine procedures and techniques provided in this guideline, so that in the end this guideline will become all-inclusive.

The development of a technology or standard is not something that can instantly be completed, and must undergo continuous corrections and improvements to gain the best results. Writing this "Archaeological Data Digitization Procedures Guideline" is only a start, in the future we will continue to devote even more effort into perfecting digitization.

TEN. Conclusion

The 21st century is an age of digital technology. “Digitization” has not only become a policy of countries around the world, but also a development strategy of countless institutions and organizations. Institutions have thus invested great amounts of human and material resources in digitization.

Digital archive projects stretch over two massive fields of knowledge, technology and humanity. Becoming familiar with technologies or knowledge in both fields is difficult for specialists in any one of the fields. It takes time to cultivate a group of talents that are proficient in both technology and humanity, which shows the importance of digitization related knowledge management and experience accumulation. The establishment of digitization procedures guidelines not only requires a foundation of data, but also the knowledge and experiences of specialists in different fields. We hope that this guideline will help more institutions and staffs to join together in completing the cultural mission of establishing digital archives.

References

Books

夏鼐、王仲殊，《中國大百科全書 • 考古學》，台北：中國大百科全書出版社，1986 年。

劉益昌，《台灣的考古遺址》，台北縣：台北縣立文化中心，1992 年。

張光直，《中國考古學論文集》，台北市：聯經出版事業公司，1995 年。

劉克竑撰文、李瑾倫繪圖，《瑄瑄學考古》台北市：行政院文化建設委員會，1997 年。

臧振華，《台灣考古》，台北市：行政院文化建設委員會，1999 年。

徐明景，《數位攝影的技術》，台北市：田園城市文化事業有限公司，2001 年。

張希強編譯，《上班族自學方案—數位影像篇》，台北市：華彩軟體股份有限公司，2001 年。

臧振華、林秋芳等，《探索台灣古文明，國立台灣史前文化博物館導覽手冊》，台北：雅凱電腦語音有限公司台灣分公司，2002 年。

林明美，《十三行博物館》，台北縣：台北縣立十三行博物館，2003 年。

林明美，《我的八里左岸旅遊書 • 筆記》，台北縣：台北縣立十三行博物館，2003 年。

簡松村、陳耀東，《數位圖檔建置流程及其加值應用之現況—以國立故宮博物院為例》，台北：博物館數位化國際研討會—器物及書畫類，2003 年。

國立台灣大學人類學系—台灣大學典藏數位化計畫，< 數位化工作流程：考古學主題小組 >，台北市：數位典藏內容發展分項計畫， 2004 年。

Gonzalez Woods 著，《數位影像處理》，繆紹綱編譯，台北：普林斯頓，2004 年 10 月，初版二刷。

中研院歷史語言研究所，〈中研院歷史語言研究所考古資料數位化工作流程〉，《數位化工作流程：考古學主題小組》，台北市：數位典藏內容發展分項計畫，2004 年。

國立故宮博物院—故宮文物數位典藏系統之研製器物數位典藏子計畫，< 數位化工作流程：器物主題小組 >，台北市：數位典藏內容發展分項計畫，2005 年。

江仁傑，《制訂後設資料的三種方案》，未刊行，2005 年。

蔡永橙、黃國倫、邱志義等著，《數位典藏技術導論》，台北市：台大出版中心，2007 年 11 月。

高芷彤、林芳志、陳秀華、陳美智，《數位化工作流程指南：委外製作》，台北：數位典藏拓展臺灣數位典藏計畫，2009 年 4 月。

李佩瑛、王雅萍、高朗軒 ，《數位化工作流程指南：色彩管理》，台北市：數位典藏拓展臺灣數位典藏計畫，2009 年 4 月。

陳秀華 、蔡幸真、高鈺茹，《 瓷器數位化工作流程指南 》，台北市：數位典藏拓展臺灣數位典藏計畫，2009 年 4 月。

Journal Papers

陳亞寧、陳淑君，〈Metadata 初探〉，《中研院計算中心通訊》第 15 卷 5 期，1999 年 03 月 01 日。

林玉雲，〈VRA Core、CDWA 與 Archaeodata 比較研究初探〉，《圖書館學與資訊科學》，第 31 卷第 2 期，2005 年 10 月。

賈馨潔，〈什麼是浮水印？〉，《數位典藏國家型科技計畫電子報》http://www2.ndap.org.tw/newsletter06/news/read_news.php?nid=732 。檢索：2010 年 01 月。

Online Resources

臺灣研究網路化 臺灣考古遺址資料庫，http://twstudy.iis.sinica.edu.tw/archeotw/。

中央研究院歷史語言研究所「數位典藏計畫分項一：考古發掘遺物照片紀錄與檔案」http://archeodata.sinica.edu.tw/allindex.html。

MAAT 數位典藏與數位學習國家型科技計畫 • 後設資料工作組 http://metadata.teldap.tw/index.html。

創用 CC 網站 http://creativecommons.org.tw/。

X-Rite, http://www.xrite.com。

Appendix

Appendix 1 Image file formats

Due to different hardware and software, image files have numerous formats. Common file formats for digital images are as follows:

1. RAW:

RAW is the original image file generated by digital cameras. RAW files are not compressed, can render rich image colors and layers, and allow adjustments to white balance and brightness; it is a good storage format if image files will be further applied in the future. The downside is that image processing requires software provided by the camera manufacturer or professional software (Adobe Photoshop, Adobe Photoshop Lightroom or Aperture). Still, RAW is gradually being adopted by digitization units as a means for permanent preservation because it holds the original image.

2. TIFF (Tagged Image File Format):

TIFF is widely applied in different platforms and software. TIFF uses the LZW algorithm to achieve lossless compression, and is therefore suitable for preserving original images, which can be used for future processing or printing high resolution images. The downside is that TIFF has low compression rates, requiring longer time to open and save files; TIFF files are rarely used for download or upload.

3. JPEG/JPG (Joint Photographic Experts Group):

The compression rate is decided when JPEG images are saved; high compression rates will result in low image quality, while low compression rates will result in high image quality close to the original image. JPEG is not suitable for image preservation, which requires high image quality, because it is a lossy compression method. However, its highly effective compression method can create extremely small files that are suitable for browsing online.

4. GIF (Graphics Interchange Format):

GIF can only save at most 256 colors and cannot accurately save the original image. However, it is relatively smaller than other image files and is therefore suitable for internet applications.

Appendix 2 Image files name coding principles (2009 version)

The Project of Historical and Cultural Heritages Developed in the Institute of History and Philology Institute, Academia Sinica.

I. The Purpose of Image File Name Coding:

1. Image File Management:

Image files should be named when they are created to avoid further effort in organizing them when they accumulate, especially in digital archives projects. Original image files meant for permanent preservation are extremely large and still inconvenient when opened with a typical computer. Therefore, file naming is a significant part of file management.

Typically, digital files are numbered according to their creation date, which is used as their file names, and although such a naming method is convenient, it makes identification of image contents extremely difficult. Some people recommend using folders to organize files, but if file names are still numbers representing their creation sequence, then files can only be identified by opening one by one once they leave their folder. This is why file names or coding principles that allow easy recognition of file contents should be established.

2. Corresponding image files to catalog data

Whether or not the archaeological database is completed, image file names serve as a basis for connecting image files to catalog data.

i. Before the archaeological database is completed:

In this stage naming the large number of image files is an important preparation for batch operations on image files (uploading and processing numerous image files at the same time). Which artifact, site or remain an image file belongs to should be clearly specified. Once the system is complete and digital files are compressed, converted, and embedded with digital watermarks (or other digital rights management methods), digital files can be uploaded in batches and link to catalog data, saving time on manually checking them one by one.

ii. After the archaeological database is completed:

After the system is complete and digital files are compressed, converted, and embedded with digital watermarks (or other digital rights management methods), digital files can be linked to catalog data, and corresponded to data in the artifacts layer, site layer or remains layer. Therefore, the utmost priority is for file content to be identifiable by file name, and not by opening individual files. In this project, the components of file names should entirely correspond to codes in the digital

archives system. Please refer to the following sections for more details and examples.

II. Image File Name Coding Principles and Examples

Digital images of artifacts

The coding principle for file names is (S + Taiwan administrative area code + Site code) + Collection number (7-10 digits) + Affair code + Image type + Whole/Zoom + Direction/Part code + subcategories. The main purpose of coding principles is to achieve easy recognition of file contents, and for being linked to text files. After the system goes online, file names can be automatically generated for new images files by filling in the column below.

1. S: "S" indicates Site, taking the first letter of the English word. This is only used for images of archaeological data in Taiwan.

2. Taiwan administrative area code: This is only used for images of archaeological data in Taiwan. Taiwan administrative area codes consist of four digits; please refer to Archaeological Sites in the Taiwan Region issued by the Ministry of the Interior for details.

Administrative Area	Code
Taipei County Pali Township	0123
Nantou County Renai Township	0813
Nantou County Puli Township	0802
Penghu County Makung City	1601
Penghu County Husi Township	1602

3. Site code: Site codes are only used for images of archaeological data in Taiwan. Taiwan administrative area codes consist of four digits; please refer to Archaeological Sites in the Taiwan Region issued by the Ministry of the Interior for details. For example: S0802TML: S indicates site, 0802 is the code for Nantou County Puli Township, and TML is the code for Tamalin.

4. Collection number: Same as coding principles for collection number. For example, R1751=>R001751, 0s are added to make up for insufficient digits.

5. Affair code

Affair type	Code	Notes
Basic cataloging	B	
Condition Report	C	
Preservation	P	
Scientific analysis	S	
Exhibition Record	E	

6. Image type

Image type	Code	Notes
Glass	G	
Black/white Reversal Film	E	
Black/white Negative Film	B	
Color Reversal Film	C	
Color Negative Film	N	
Black/white Picture	BP	
Color Picture	CP	
Drawing (includes cross section profile)	D	
Rubbing	R	
X-Ray image	X	
Infrared Image	I	
micrograph	O	
Other image	Y	
Reconstruction	U	
3D image	3	
Video	V	
Audio	A	
motion Picture	P	
other	Z	
Map	M	
Text	T	

7. Whole or zoom

Whole/Zoom	Code	Notes
Whole	W	
Zoom	Z	
Context	C	

8. Direction/Position

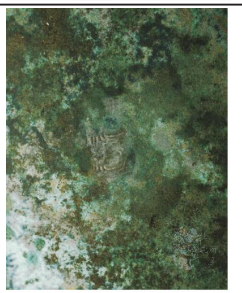
Object that cannot be determined should be proposed for discussion.

Direction / Position	Code	Meaning
Whole/Direction	A	FRONT
	B	BACK
	L	LEFT (west)
	R	RIGHT (east)
	T	TOP (north)
	D	DOWN (south)
	K	OBLIQUE

Zoom/Position	M	ANIMAL
	O	OPENING
	Q	BODY
	C	DECORATION
	E	EAR
	F	FOOT
	H	HANDLE
	I	INSCRIPTION
	L	LID
	N	NECK
	P	PLANT
	S	SHOULDER
	Z	Other

9. Example 1: Ten images of a single handle Chinese cooking vessel from Mainland China registered under R001110, each image and its file name is as shown below:

No.	Image	Image file name
1		R001110BCWA
2		R001110BCWB
3		R001110BCWD
4		R001110BCWL
5		R001110BCWR
6		R001110BCWT_1

7		R001110BCWT_2
8		R001110BCZI
9		R001110BRWH
10		R001110BRWS

Example 2: Six images of ceramic goblet from Taiwan registered under T9200119, each image and its file name is as shown below:

No.	Image	Image file name
1		S0813CPT9200119BCWA
2		S0813CPT9200119BCWB
3		S0813CPT9200119BCWD
4		S0813CPT9200119BCWL
5		S0813CPT9200119BCWR
6		S0813CPT9200119BCWT

10. Combined images:

If an image holds several artifacts at the same time, the registration number portion is replaced with a "combined image code." For the combined image code, the smallest registration number of the artifacts in the image is used and added with a Z in front. If the smallest registration number was already used by another combined image, then the second smallest registration number is used. For example: S0802SWKZT9102187BCWZ_1 (S indicates site, 0802 is the administrative area code, SWK is the site code, and ZT9102187 is the combined image code).

11. Other matters that need attention:

a.If two or more images have the same code, then add numbers (1,2~N) after the code with "_" separating the two, e.g. R001751BCWK_2

b.If the whole/front image has two focal points on the left and right as a result of the object's curve, then take two photos and separate the two images by adding L (focal point on the left) and R (focal point on the right) according to the focal point used.

c.Charts can be further divided into three categories "tracing paper," "graph paper," and "writing paper." Considering the sequence images are displayed, use A, B and C to indicate each type of paper; if two or more images of the same type of paper exist, then directly add the number of pages after A, B or C. For example, The coding method for two pages of tracing paper would be: Database code + Affair code + Image type code + Whole/Zoom code + Chart type code + Number of pages, in which the last two characters would be A2.

d.Photographs taken in the field of artifacts being unearthed should all be linked to under remains data and not artifact data.

e.If the image is the container of a artifact, then its coding method is to add an X after the artifact's registration number, e.g. PR00045PR00046_XBCWA (the container held two artifacts, so both registration numbers are used), but such images are not added to the database.

f.If the image is of an inscription on an object, due to the fact inscriptions may appear on different parts of the same object, the position code should be added after the image file name (please see 8. Direction/Position) separated with a "_", e.g. R001077BRWI_L.

Field data

1. Site data

The coding principle in this section is the same as for artifacts, detailed categories are as follows:

i. Database code + Site code

a. Database code

Add "S" at the front for corresponding multimedia image files of sites to catalog data and for the convenience of identification.

b. Site code

Add the site code after the database code, e.g. the site code of Hsiaotun is "Y", adding it to the database code would be "SY". If the site is in Taiwan, then add the four digit Taiwan administrative area code, e.g. the code of the Tapengkeng site is 0123TPK, in which 0123 is the administrative area code and TPK is the site code. The "site code" in the file name should be consistent with the "site code" in the multimedia file record, and the "site code" in the site catalog data.

The Institute of History and Philology's codes for sites are as follows:

	Site name	Code
Taiwan	Tapengkeng	0123TPK
	Tamalin	0802TML
	Shuiwaku	0802SWK
	Chuping	0813CP
	Kuoyueh	1602KY
	Sokang	1601Sok
Henan	Hsiaotun	Y
	Houkang	H
	Ssupanmo	SPM
	Warping dam and Huojia village	WH
	Xun County Hsin village	HS
	Glass Pavilion in Hui County	HL
	Ji County Shanpiao township	SP
	Houchiachuang Hsipeikang	HPK
	Houchiachuang south	S
	Houchiachuang Kaochingtaitzu	K
	Wukuan Nanpatai	NP
	Tasikong village	TSKT
	Tung Lok township	TLT
	Fanjaizhuang	FT
	Anshang village	A

Shandong	Wuwa village	WW
	Takutui	TKT
	Chengtzuai	CTA
Gansu	Foyemiao	FYM
	Lamawan	LMW

ii. Affair code

Affair type	Code	Notes
Basic cataloging	B	

At present, only basic cataloging data are included in the database.

iii. Image type

Image type	Code	Notes
Glass	G	
Black/white Reversal Film	E	
Black/white Negative Film	B	
Color Reversal Film	C	
Color Negative Film	N	
Drawing (includes cross section profile)	D	
Rubbing	R	
X-Ray	X	
Infrared Image	I	
micrograph	O	
Other image	Y	
Reconstruction	U	
3D image	3	
Video	V	
Audio	A	
motion Picture	P	
other	Z	
Map	M	
Text	T	

iv.Whole or zoom:

Whole/Zoom	Code	Notes
Whole	W	
Zoom	Z	
Context	C	

v. Direction/Position

Sites that cannot be determined should be brought up for discussion.

Direction/Position	Code	Meaning
Whole/Direction	A	Front
	B	Back
	L	West
	R	East
	T	North
	D	South
	K	OBLIQUE
	Z	Other

vi.If two or more images have the same code, add a number (1, 2~N) after the code with a "_" separating the two.

vii.Example: Using black/white photos of sites as an example, digital image files are coded as follows:

Affair type	Site Name	Code	Example of file name coding for digital image files of typical black/white photos
Taiwan	Tapengkeng	0123TPK	S0123TPKBBWA…
	Tamalin	0802TML	S0802TMLBBWA…
Henan	Hsiaotun	Y	SYBBWA…
	Houkang	H	SHBBWA…
	Ssupanmo	SPM	SSPMBBW…
	Warping dam and Huojia village	WH	SWHBBW…
	Xun County Hsin village	HS	SHSBBW…
	Glass Pavilion in Hui County	HL	SHLBBW…
	Shanpiao township	SP	SSPBBW…
	Houchiachuang Hsipeikang	HPK	SHPKBBW…
	Houchiachuang south	S	SSBBW…
	Houchiachuang Kaochingtaitzu	K	SKBBW…
	Wukuan Nanpatai	NP	SNPBBW…
	Tasikong village	TSKT	STSKTBBW…
	Tung Lok township	TLT	STLTBBW…
	Fanjaizhuang	FT	SFTBBW…
	Liuzhuang	LA	SLABBW…
	Anshang village	A	SABBW…
Shandong	Wuwa village	WW	SWWBBW
	Takutui	TLT	STLTBBW
	Chengtzuai	TKT	STKTBBW
Gansu	Foyemiao	FYM	SFYMBBW
	Lamawan	LMW	SLMWBBW

2. Archaeological excavation work records

The coding principle for this category is the same as for site data, but the site of which archaeological excavation work was recorded should be specified, on one hand providing a fast identification method, on the other hand providing a basis for being automatically linked to the database.

i. Site code

For accuracy and effectiveness of corresponding digital images to catalog data of "archaeological excavation work records" under the same site, the file name of images photographed in the field should start with "S" + Site Code.

ii. Database code:

The code of "Archaeological Excavation Work Record" is "E". For example: Photos taken during the second excavation at Hsiaotun are coded with "Y02"; this code should be consistent with "Archaeological Excavation Work Record Codes" in the database. Therefore, file names of the second excavation at Hsiaotun start with SYEY02 + codes after the third category. If the excavation work record is from a site in Taiwan, the four digit administrative area code should be added in front of the site code, e.g. S0813CPE01BCWA_01: S indicates site; 0 is the code for Chu Ping; S0813CP, 0813 is the administrative area code, CP is the original site code; E indicates the image is an archaeological excavation work record; 01 indicates it is from the first excavation.

iii. Affair type

Affair type	Code	Notes
Basic cataloging	B	

iv. Image type

Image type	Code	Notes
Glass	G	
Black/white Reversal Film	E	
Black/white Negative Film	B	
Color Reversal Film	C	
Color Negative Film	N	
Drawing (includes cross section profile)	D	
Rubbing	R	
X-Ray	X	
Infrared Image	I	
micrograph	O	
Other image	Y	
Reconstruction	U	
3D image	3	
Video	V	
Audio	A	
motion Picture	P	
other	Z	
Map	M	
Text	T	

v. Whole or Zoom:

Whole/Zoom	Code	Notes
Whole	W	
Zoom	Z	
Context	C	

vi. Direction/Position

Sites that cannot be determined should be brought up for discussion.

Direction/Position	Code	Meaning
Whole/Direction	A	Front
	B	Back
	L	West
	R	East
	T	North
	D	South
	K	OBLIQUE
	Z	Other

vii.If two or more images have the same code, add a number (1, 2~N) after the code with a "_" separating the two.

viii.If the image file contains a field diary, then a capitalized letter is added between the first and second coding category (this letter is added because sometimes field diaries of the same site on the same day might be written by different people, but this letter does not specify any recorder in particular; the letters A, B, C are added to indicate the number of people who wrote on the field diary, A indicates one person wrote the diary, whoever it may be.), followed by the date of the field diary entry; dates are coded in the form YYYYMMDD, and then following coding categories are added. For example:

SYEY01A19281013BTWA_1

ix.Examples:

a. Field work photos of each excavation at Hsiaotun:

Type	Coding method	Example
1st Excavation at Hsiaotun	S + Site Code + E + Archaeological Work Record Code + Affair Code + Image Type Code + Position Code + …	SYEY01bbw…
2nd Excavation at Hsiaotun	S + Site Code + E + Archaeological Work Record Code + Affair Code + Image Type Code + Position Code + …	SYEY02bbw…
3rd Excavation at Hsiaotun	S + Site Code + E + Archaeological Work Record Code + Affair Code + Image Type Code + Position Code + …	SYEY03bbw…
4th Excavation at Hsiaotun, and so on…	S + Site Code + E + Archaeological Work Record Code + Affair Code + Image Type Code + Position Code + …	SYEY04bbw…

b. Hsiaotun Field Diary:

Type	Coding method	Example
1st Excavation at Hsiaotun	S + Site Code + E + Archaeological Work Record Code + Recorder + Date of Record + Affair Code + Image Type Code + Position Code + …	SYEY01A19281013BTWA
2nd Excavation at Hsiaotun	S + Site Code + E + Archaeological Work Record Code + Recorder + Date of Record + Affair Code + Image Type Code + Position Code + …	SYEY02B19290307BTWA
3rd Excavation at Hsiaotun	S + Site Code + E + Archaeological Work Record Code + Recorder + Date of Record + Affair Code + Image Type Code + Position Code + …	SYEY03A19291017BTWA
4th Excavation at Hsiaotun, and so on…	S + Site Code + E + Archaeological Work Record Code + Recorder + Date of Record + Affair Code + Image Type Code + Position Code + …	SYEY04C19310321BTWA

c. Work photos and field diary of excavation work at the Chu Ping Site:

Type	Coding method	Example
1st Excavation at Chu Ping	S + Site Code + E + Archaeological Work Record Code + Affair Code + Image Type Code + Position Code + …	S0813CPE01BCWA_01
3rd Excavation at Chu Ping	S + Site Code + E + Archaeological Work Record Code + Affair Code + Image Type Code + Position Code + …	S0813CPE0319851128BTWA_01

3. Excavation unit data

The coding principle is the same as the above, but the following matters need attention:

i. The code for each pit should be unique and listed before digitization.

ii. The “excavation unit code” in the file name should be consistent with the “excavation unit code” of the multimedia file record and the “excavation unit code” of catalog data.

Example: Archaeological data of Mainland China

Site	Excavation Unit	Coding Method	Example
Hsiaotun 1~5	Pit 14	S + Site Code + P + Excavation Unit Code + Affair Code + Image Type Code + Position Code +…	SYPH14bbw…
Hsiaotun 1~9	A31	S + Site Code + P + Excavation Unit Code + Affair Code + Image Type Code + Position Code +…	SYPA31bbw…
Hsiaotun 13~15	A126	S + Site Code + P + Excavation Unit Code + Affair Code + Image Type Code + Position Code +…	SYPA126bbw…

WaWu Village WW	Ww Ww01 Ww02 Ww03 And so on Ww51	S + Site Code + P + Excavation Unit Code + Affair Code + Image Type Code + Position Code +…	SWWPWWBBW...
	wwBAS	S + Site Code + P + Excavation Unit Code + Affair Code + Image Type Code + Position Code +…	SWWPWW01BBW…
	WWSD	S + Site Code + P + Excavation Unit Code + Affair Code + Image Type Code + Position Code +…	SWWPWWBASBBW...
	Wwgathering	S + Site Code + P + Excavation Unit Code + Affair Code + Image Type Code + Position Code +…	SWWPWWSDBBW...

Taiwan Archaeological Data

Site	Excavation Unit	Coding Method	Example
Chichiawan	Trial Pit TP1	S + Site Code + P + Excavation Unit Code + Affair Code + Image Type Code + Position Code +…	S0621CCWPTP1BTWA_03
Chichiawan	Pit T0P0-5	S + Site Code + P + Excavation Unit Code + Affair Code + Image Type Code + Position Code +…	S0621CCWPT0P0-5BDWA_01

4. Layer data

No layer data currently exists, but it should be represented with an L if found. The file name for such data should be coded as: S + Site Code + Excavation Unit Code + L + Layer Code, remaining details are left for future discussions.

5. Archaeological remains

The coding principle is the same as the above, but special attention should be paid to the special principles below:

i. The “site code” in the file name should be consistent with the “site code” of the multimedia file record and the “site code” of catalog data.

ii. Computers generally have limitations on file names; please refer to related rules for details. Using the target site of this project – Hsiaotun Site, the site code includes a colon (1:M01), but many operating systems consider it to be an illegal file name. The colon does not affect identification of the file name, so delete the colon if the file name is not accepted by the operating system.

Examples are as follows

Archaeological data of Mainland China:

Type	Site	Coding Method	Example
Hsiaotun (1st Excavation)	1:M01	S + Site Code + P + Excavation Unit Code + Affair Code + Image Type Code + Position Code +…	S0621CCWPTP1BTWA_03
(Code changed to 01M01	S + Site Code + F + Archaeological Remains Code + Affairs Code + Image Type Code + Position Code…	SYF01M01bbw…	S0621CCWPT0P0-5BDWA_01
Hsiaotun (1st Excavation)	1:H01		
	S + Site Code + F + Archaeological Remains Code + Affairs Code + Image Type Code + Position Code…	SYF01H01bbw…	
Hsiaotun 13~15th grave	YM001		
	S + Site Code + F + Archaeological Remains Code + Affairs Code + Image Type Code + Position Code…	SYFYM001bb…	
Hsiaotun 13~15th ash hole	YH001		

	S + Site Code + F + Archaeological Remains Code + Affairs Code + Image Type Code + Position Code…	SYFYH001bb…	
Hsiaotun 13~15th rammed earth	YB001	S + Site Code + F + Archaeological Remains Code + Affairs Code + Image Type Code + Position Code…	SYFYB001bb

Taiwan Archaeological Data:

Type	Site	Coding Method	Example
Chu Ping	House foundation No.15	S + Site Code + F + Archaeological Remains Code + Affairs Code + Image Type Code + Position Code…	S0813CPFF15BCWA_01
Chu Ping	Grave No.115	S + Site Code + F + Archaeological Remains Code + Affairs Code + Image Type Code + Position Code…	S0813CPFM115BCWA_02

Appendix 3 Metadata fields and system functional requirements of Academia Sinica Institute of History and Philology for archaeological materials

Tag Name		Content
Basic Cataloging		
Data processing status		
Artifact identification code		
Site	Site	
	Archaeological excavation work records	
Excavation unit	Unit code	
	Layer code	
Remains	Remains	
	Position number	
Record type		
Artifact name		
Other names	Type	
	Name	
	Date/Period	
	Remarks	
Remarks		
Display sequence		
Collection Location		
Registration number		
Registration date		
Holding institution	Code	
	Name	
	Location	
Inventory number	Type	
	Number	
	Image board number	
Acquisition method		
Acquisition date		
Purchase amount		
Provider	Role	
	Name	
Artifact in exchange		
Artifact Description		
Physical appearance	Description	
	Other descriptions	
	Type	
Decoration	Name	
	Quantity	
	Position	
	Tool	
	Material	
Artifact position/sequence		
Function	Type	
	Item	
Material	Type	
	Category	
	Subcategory	
Art category	Material	
	Type	
	Method	
	Technique	
	Description	
Surface traces	Position/range	
	Category	
	Subcategory	
	Description	

Measurements			
Measurement	Type/Method		
	Complete/remains		
	Range/Portion		
	Value		
	Unit		
Measurement time			
Measurement date			
Remarks			
Source			
Dating Data			
Type			
Method			
Material			
Time			
Dating results			
Original data	Value		
	Value		
Corrected data	Value		
	Value		
Source			
Remarks			
Culture/Era			
Culture sequence	Name		
	Stage		
	Section		
Archaeology culture	Name		
	Area		
	Stage		
Historical period	Start	Dynasty	
		Age	
		Era name	
		Year	
		Month	
	End	Day	
		Dynasty	
		Age	
		Era name	
		Year	
		Month	
		Day	
Nation/Feud/Ethnic Group			
Inscriptions			
Inscription	Position		
	Lines		

	Direction		
	Production method		
	Script		
	Copy		
	Explanatory note	Script	
	note	Explanation note	
		Source	
	Remarks		
Preservation Records			
Description			
Changes			
Recorder			
Record Date			
Remarks			
Source			
Conservation and Treatment			
Description			
Conservation method			
Conservation process			
Tools used			
Drugs used			
Samples			
Conservator			
Conservation date			
File creator			
File creation date			
Afterthoughts and difficulties			
Source			
Scientific Analysis			
Type			
Method			
Description			
Instrument			
Result			
Analyzer			
Date			
Remarks			
Source			
Compilation			
Artifact Name			
Registration Number			
Relationship			
Researcher			
Date			
Explanation			

National Central Library Cataloging in Publication
Archaeological Data Digitization Procedures Guideline

Advisory Unit: National Science Council, Executive Yuan
Issuer: Lin Fu-Shih
Editor-in-Chief: Peng-Sheng Chiu
Executive Editors: Lin Yen-Hung, Lin Ting-Li, Lin Fang-Chih, Kao Lang-Hsuan
Authors: Wang Ya-Ping, Lu Chun-Yi
Reviewer: Associate Researcher Tzu-Ning Li, Nation Taiwan Museum
Translator: International Collaboration and Promotion of Taiwan e-Learning and Digital Archives Project
Publisher: Taiwan Digital Archives Expansion Project, Taiwan e-Learning and Digital Archives Program
Address: No.128, Sec.2, Academia Rd., Nankang District, Taipei City, 115
Institute of History & Philology, Academia Sinica
Tel: +886-2-2782-9555 extn.288
Fax: +886-2-2786-8834
Website: http://content.teldap.tw
Email: content@gate.sinica.edu.tw
Typesetting: Near Image Design Co., Ltd.
Printing: Near Image Design Co., Ltd.

Publish Date: 1st Edition December 2011
ISBN: 978-986-03-1292-8

本書譯自拓展臺灣數位典藏計畫出版數位化工作流程指南：考古資料