# Digitization Specifications of Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences on Physical Archives

## 1. Scope

These Specifications specify the technical and management requirements for physical archive digitization.

These Specifications are applicable to the management of the digitization process of physical archives using methods such as photography and scanning.

## 2. Normative References

The following documents are indispensable for the application of this document. All referenced documents with dates are only applicable to this document with the dated version. All referenced documents without dates apply to the latest version (including all amendments).

GB/T20530-2006 *Guidefor the Digitalize Processing of Document and Archives* DA/T1 *Basic Terminology of Archive Work*

DA/T18 *Archive Recording Rules*

DA/T31 *Digitization Specifications on Paper Archives*

DA/T62 *Digitization Specifications on Audio-visual Archives*

DA/T63 *Metadata Schemefor Audio-visual Electronic Archives*

DA/T68 *Specificationsfor Archive Service Outsourcing*

ISO/TR13028 Information and Documentation - Implementation Guidelines for Digitization of Records

（InformationandDocumentationImplementationguidelinesfordigitizationofrecords）

## 3. Terms and Definitions

The terms and definitions defined by GB/T20530-2006, DA/T1, DA/T18, and the following terms and definitions are applicable to this document.

Archive Object

Physical objects with specific shapes produced or acquired by national institutions, social organizations, or individuals in social activities, which have preservation value, including letters of commendation, trophies, banners, certificates, seals, products, specimens, tools, and equipment.

Digitization

The process of converting analog signals into digital signals using information technology.

Digitization of Physical Archives

The process of digitizing physical archives using equipment such as photography and scanning, converting them into digital files stored on media such as tapes, disks, and CDs, and establishing the relationship between catalog data and digital files according to the intrinsic connection of physical archives.

3D Scan

Digitally scanning the spatial shape, structure, and color of objects to reconstruct high-resolution virtual objects.

3D Scanner

A hardware and software system that can generate digital data, which can represent the overall or partial form of physical archive objects in three dimensions.

Feature Point of Physical Archives

The critical point where significant changes occur in the spatial shape, structure, and color of physical archive objects.

Point Cloud

A collection of points distributed discretely in 3D space.

Texture

A visual characteristic that reflects homogeneous phenomena in an image, reflecting the surface structure and organizational arrangement attributes of an object's surface with slow or periodic changes.

## 4. General

We shall plan and carry out the digitization of physical archives in a coordinated and scientific manner based on factors such as the rarity, openness, utilization rate, urgency of rescue, and digitization funds of the archives. The implementation of the digitization of physical archives shall follow the requirements and recommendations put forward by ISO/TR13028and GB/T20530-2006.

The basic stages of the digitization of physical archives mainly include: pre-digitization processing, establishment of catalog database, digitization acquisition, image processing, data linking, acceptance and handover of digitization results, etc.

Effective management and technical means shall be adopted to ensure the quality of the digitization results of physical archives. The digitization of physical archives shall follow the objective laws of archive management, truthfully reflect the relevant information of the physical archives, and maximize the presentation of the original appearance of the archives.

During the digitization process of physical archives, metadata such as digital project information, technical environment, and various technical parameters of digitization shall be preserved. The determination of metadata elements shall comply with the requirements proposed by ISO/TR13028 (Information and Documents - Implementation Guidelines for Digital Records).

Safety management of each link of the digitization of physical archives shall be strengthened to ensure the safety of archive entities and archive information.

When processing classified archives, work shall be carried out in accordance with the relevant confidentiality requirements for classified archives.

## 5. Organization and Management

5.1 Institutions and personnel

5.1.1 We shall establish the organization for the digitization of physical archives, carry out overall planning, organizing implementation, coordinating management, ensure security, providing technical support, conduct supervision and inspection, and accept results to ensure the smooth progress of digitization work.

5.1.2 Personnel with corresponding capabilities shall be equipped, including management personnel familiar with archive business and possessing a high level of investigative research and good organizational leadership abilities, technical personnel familiar with relevant standards and able to provide technical support for various links of the digitization of physical archives, and operators who have a certain knowledge of digitization and are familiar with their job duties. Standardized management of personnel shall be carried out through scientific and normative management systems.

5.1.3 To enhance the security of digitization work, external contracted personnel shall undergo rigorous review.

5.2 Infrastructure

5.2.1 A dedicated processing area shall be provided, with a reasonable layout forming work areas for archive storage, pre-digitization processing, archival recording, digitization acquisition, image processing, quality inspection, etc.

5.2.2 The selection of processing areas and the control of environmental factors such as temperature and humidity shall be conducive to the protection of archive entities. Facilities and equipment for fire prevention, waterproofing, prevention of harmful organisms, anti-theft alarm, video surveillance, etc. shall be provided in the area.

5.2.3 Facilities and equipment for physical archives digitization shall be reasonably planned, equipped, and managed to ensure they can meet the needs of physical archive digitization work.

5.3 Work plan

5.3.1 Based on thorough research, we shall develop scientifically reasonable work plans to ensure that the digitization of physical archives achieves the intended goals.

5.3.2 The work plan for digitization of physical archives shall include digitization targets, work objectives, work content, cost accounting, digitization technology methods and major technical indicators, acceptance criteria, personnel arrangement, division of responsibilities, schedule arrangement, safety management measures, etc.

5.3.3 We shall have the physical archive digitization work plan reviewed by experts to ensure its scientific, standardized, and reasonable nature.

5.3.4 The physical archive digitization work plan shall be strictly enforced after approval. The results of work plan approval shall be kept together with other documents generated during the digitization process.

5.4 Management system

5.4.1 We shall establish scientific and standardized management systems and strictly implement them during the work process to effectively guarantee archive security and the quality of physical archive digitization results.

5.4.2 The management system for digitization of physical archives shall include systems for position management, personnel management, site management, equipment management, data management, archive entity management, etc. The position management system mainly specifies the work objectives and responsibilities of various positions involved in digitization work, forming clear position business process norms, assessment standards, reward and punishment methods, etc.; the personnel management system mainly standardizes personnel's safety responsibilities, daily behaviors, verification and management of externally hired personnel information, registration of non-employees' visits, etc.; the site management system mainly standardizes the management of personnel access and working environment, infrastructure, network, monitoring facilities, on-site items, documents, etc.; the equipment management system mainly standardizes the management of all equipment

involved in digitization work at various stages; the data management system mainly standardizes the management of data generated in various stages of digitization; the archive entity management system mainly standardizes the handling, management, and storage of archive entities during the digitization process.

5.5 Workflow control

5.5.1 We shall establish relevant workflows, operation specifications, etc., to effectively control the entire process of digitization of physical archives and ensure the quality of digitization results. See Appendix A for the physical archive digitization process.

5.5.2 Strict safety management shall be carried out for the entire process of digitizing physical archives, and data generated at various stages of digitization shall be backed up in a timely manner.

5.5.3 A sound feedback mechanism shall be established to promptly feed back and correct problems discovered during the digitization process of physical archives.

5.6 Work document management

5.6.1 Work documents for physical archive digitization shall be tailored to practical requirements to strengthen the management of digitization work. These documents mainly include the work plan, approval documents, and workflow forms for physical archive digitization, data acceptance forms, project acceptance reports, handover lists of physical archive digitization results, etc. When outsourcing, they shall also include tender documents, bidding documents, notification of winning bids, project contracts, confidentiality agreements, etc. Refer to Appendix B for examples of some work forms.

5.6.2 Enhanced management of work documents for physical archive digitization shall clarify the requirements for organizing, archiving, and transferring work documents generated during the digitization process.

5.7 Outsourcing of archive digitization

5.7.1 If outsourcing of the digitization of physical archives is required, the relevant work shall comply with the relevant requirements of DA/T68 *Specifications for Archive Service Outsourcing*. Archives departments shall rigorously review the relevant qualifications of digitization processing companies in terms of company nature, shareholder composition, security and confidentiality, enterprise scale, registered capital, etc.; evaluate the technical capabilities of digital processing companies according to the requirements of Chapter 5 of GB/T20530-2006; examine the management capabilities of processing companies from aspects such as the establishment and completeness of regulations and systems. If it is necessary to review the confidentiality qualifications of digitization companies, it shall be carried out in accordance with the requirements of documents such as *Management Measures for Printing Qualifications of National Secret Carriers* (GBF [2012] No. 7).

5.7.2 During project implementation, strict safety management requirements shall be implemented from various aspects such as the archives departments, digital service agencies, digital processing sites, digital processing equipment, archive entities, acceptance and reception of digitization results, and equipment handling, according to the Regulations on Safety Management of Archive Digitization Outsourcing (DBF [2014] No. 7).

5.7.3 Archives department shall assign dedicated personnel to supervise and guide the outsourcing of physical archive digitization, completing tasks such as quality control, progress monitoring, investment monitoring, safety monitoring, and coordination and communication.

## 6. Archive Delivery

6.1 In accordance with the physical archive digitization work plan, the Archives storage departments shall conduct preliminary preparations such as archive retrieval, inventory, and registration for the digitization objects. After obtaining approval from the relevant responsible person, relevant procedures for retrieving archives are strictly carried out in accordance with the regulations of archive storeroom management. After jointly verifying the accuracy with the digital department, the archives are handed over for delivery.

6.2 During the physical archive digitization process, a storage warehouse located close to the digital processing site shall be set up based on work needs to temporarily store physical archives. Strict management shall be carried out for the receiving and return of physical archives, including careful inspection, inventory, registration, etc., to ensure the safety of physical archives.

## 7. Pre-digitization Processing

7.1 Determining the digitization objects of physical archives

In principle, complete digitization shall be carried out for physical archives identified as digitization objects, without missing feature points. If there are feature points that do not need to be collected, they shall be annotated.

7.2 Archive inspection

7.2.1 Conduct an external examination of physical archives. For those severely damaged or with other conditions unfavorable for digitization processing, appropriate technical treatments such as moderate cleaning shall be performed. The technical treatment shall be reasonably selected according to the specific situation of the physical archives.

7.2.2 Inspect the quality of physical archives and record any issues found.

7.2.3 Record information such as physical archive No., material, and shape.

7.2.4 Any items that require marking in the catalog database shall be marked accordingly.

7.3 Preparation of catalog data

7.3.1 According to the data rules established when creating the catalog database, the content of the catalog in the archives shall be standardized based on the relevant information of physical archives.

7.3.2 Any items that require marking in the catalog database shall be marked accordingly.

## 8. Establishment of Catalog Database

8.1 Data rules for the catalog database shall be formulated, including data field length, field type, field content requirements, etc. The establishment of catalog database data rules shall comply with the requirements of DA/T18 for archive recording. Strict adherence to these rules shall be observed during the preparation of physical archive catalogs and the establishment of catalog databases.

8.2 Consider universal data format conversion issues when selecting a catalog database to facilitate data exchange.

8.3 The design of database structure shall pay special attention to maintaining the internal connections of archives, facilitating the management and utilization of physical archive digitization results.

8.4 The results of modifications and supplements to physical archive catalogs during pre-digitization processing shall be entered into the database to form accurate and complete catalog data. Catalog data shall include at least file No., retention period, year, responsible party, title, material, type, cross-reference No., as well as information on digitization projects, technical environment, and various technical parameters.

8.5 A combination of computerized automatic proofreading and manual proofreading shall be used to check the quality of catalog data, including the completeness, standardization, and accuracy of recorded items. Any data found to be unsatisfactory shall be promptly corrected.

## 9. Digitization Collection

9.1 Basic requirements

9.1.1 Select appropriate information collection equipment based on the actual conditions of archive originals, digitization purposes, digitization scale, computer network, and storage conditions, and set and adjust relevant parameters accordingly. The settings and adjustments of parameters shall ensure that the scanned digital images are clear, complete, undistorted, and closest to the original appearance of the archives.

9.1.2 Pay special attention to protecting the physical archives during the collection process, and no damage shall be inflicted.

9.1.3 Before collection, clean, inspect, and adjust the corresponding collection equipment and venues to ensure the quality of the digitization results of physical archives.

9.1.4 Regular maintenance and upkeep of equipment shall be conducted following the relevant usage regulations.

9.2 Flat scanning

9.2.1 Applicable to physical archives that can be displayed in 2D static image form, such as letters of commendation, certificates, and seals.

9.2.2 Archives larger than the scanning size of the scanner used can be scanned using larger-format scanners or by dividing the scanning into smaller sections and then stitching the images together. When dividing scans, there shall be sufficient overlap between adjacent images, and the method of division shall be clearly indicated using templates or other means. If automatic stitching software is used later, the recommended overlap size shall be no less than 1/3 of the size of a single image corresponding to the original document.

9.2.3 For extremely precious and irregularly sized archives, for the convenience of visually displaying the original size, methods such as using templates and rulers to mark the size information of the original piece can be adopted.

9.2.4 To preserve physical archive original information to the fullest extent and facilitate various uses, it is advisable to scan using the color mode for all pages. The selection of scanning color mode and scanning resolution shall comply with the requirements and recommendations proposed in DA/T31.

9.3 3D scan

9.3.1 Applicable to displaying physical archives in 3D stereoscopic image form, such as trophies, products, specimens, tools and equipment.

9.3.2 The main technical parameters of 3D scanning for digitizing physical archives are shown in Table 1.

Table 1 Main Technical Parameters of 3D Scanning For Digitizing Physical Archives

| **Item** | **Requirements** |
|---|---|
| **Scanning Accuracy** | ≤0.05mm, 500mm (distance between lens and object to be scanned) |
| **Point Spacing** | ≤0.254mm |
| **Texture Color** | ≥32 bits color |

| | | |
|---|---|---|
| **Scanning Distance** | 400mm-600mm | |
| **Storage format** | Geometric data | .3DS, .3DMAX, .3DM, .FLT, .0BJ, .WRL, .DAE, etc. |
| | Texture data | .JPG, .TIFF, .PNG, .DDS, .TGA, etc. |

9.3.3 Texture resolution

The overall texture resolution of scanning physical archives shall meet the following requirements:

a) It should truly reflect the color, texture, shape, and pattern of the surface of the physical archives, and the texture of the same surface shall be consistent;

b) The number of pixels in length and width shall be the N power of 2 (N is a natural number), and the difference in length-to-width ratio of the texture shall not be too large;

c) Correction shall be made to reduce deformation caused by viewing angle or lens distortion, and glare and shadows shall be eliminated;

d) The image resolution shall not be less than 100dpi, with harmonious and natural color tones;

e) Seamless stitching shall be performed, with natural transitions;

f) After texture scanning, the digitization results shall be compared and corrected with a color card to ensure consistency with the color of the physical archives.

9.4 Digital photography

9.4.1 Applicable to displaying physical archives in 2D static image form, such as medals, trophies, and banners. Digital cameras shall be used for color photography, with a photography accuracy of no less than 16 million pixels.

9.4.2 During photography, firstly, select a lens with a focal length above the mid-focus and try to keep it horizontal to avoid image distortion; secondly, set the camera to a screen format without time display; thirdly, pay attention to uniform lighting; fourthly, fill the image frame with the subject; fifthly, for physical archives with regular shapes, photography is performed from multiple angles such as frontal, lateral, and top views.

9.4.3 The main technical parameters of digital photography for digitizing physical archives are shown in Table 2.

Table 2 Main Technical Parameters of Digital Photography for Digitizing Physical Archives

| | |
|---|---|
| **Resolution** | 4928×3280 or above; for photos used for system browsing and web browsing, the resolution can be set to 1920×1080 or 720×480. |
| **Exposure Mode** | Manual exposure / automatic exposure |
| **Color Mode** | ≥32 bits, RGB mode |

9.4.4 Long-term storage formats for physical archive digital images include TIFF, JPEG, or JPEG2000, with the choice of compression ratio based on actual application needs.

9.4.5 When using physical archive digital images, consider factors such as network browsing speed, ease of operation, and storage space occupation to convert images into other formats such as OFD and PDF.

9.4.6 Use the same storage format for the same photography method.

9.5 Panoramic photography

9.5.1 Applicable to displaying physical archives in the form of 3D panoramic images, such as production tools and facilities and equipment.

9.5.2 When conducting panoramic photography for physical archives:

a) Ensure that the lens is orthogonal to the surface of the photographed physical archives, and the photographer uses professional equipment such as digital SLR cameras, camcorders, fisheye lenses, panoramic heads, tripods, and aerial drones to shoot the physical archives from a 360° panoramic view, with no feature points of the physical archives being omitted;

b) Choose the appropriate exposure mode to ensure that the exposure parameters of each photo are the same and the tone is uniform;

c) Once the focal length is set, it shall remain unchanged until all photos are taken, and each photo shall be taken at intervals of 30°.

Table 3: Main Technical Parameters of Digitizing Physical Archives (360° Panoramic Photography)

| **Item** | **Requirements** |
|---|---|
| Camera shooting mode | Manual mode (M) / automatic mode |
| Pixels | ≥16 million |
| Color Mode | ≥32 bits, RGB mode |
| Image format | Common formats such as TIFF, JPEG, or JPEG2000 |
| Video format | Common formats such as AVI or MXF |

9.6 File naming

9.6.1 Digital files shall be named based on the file No., ensuring uniqueness.

9.6.2 When one directory corresponds to multiple digital files, the digital files can be named based on a combination of file No. and sequence number.

9.6.3 Scientifically establish the storage path for digital files to ensure accurate data mounting.

**10. Image Processing**

### 10.1 Rotation and correction

Images that do not conform to the reading orientation shall be rotated to restore them. Skewed images shall be corrected to visually eliminate skewness.

### 10.2 Cleaning

Digital images shall undergo cleaning processing to remove impurities such as spots, lines, and black edges that affect image quality during digitization. The principle of presenting the original appearance of the archives shall be followed, and traces formed during the storage process of physical archives must not be removed during processing.

### 10.3 Image stitching

Multiple images formed by sectional scanning shall be stitched together to form a complete image to ensure the integrity of the digitized archive images. The stitching process shall ensure smooth fusion at the stitching seams, with no obvious stitching traces on the entire image after stitching.

### 10.4 Cropping

Digital images shall be cropped to remove excess parts, effectively reducing the file size and saving storage space.

### 10.5 Noise reduction

Remove invalid points in the 3D scan point cloud caused by external factors (such as light, vibration) and factors inherent to the 3D scanner.

### 10.6 Data smoothing

To obtain smooth models and improve contour recognition, the results of 3D scanning shall undergo smoothing processing to improve data smoothness, making it closer to the geometric characteristics of the physical archives.

10.7 Image quality inspection

Digitization personnel shall inspect the quality of digital files, including image skewness, clarity, distortion, etc. Those that do not meet quality requirements or those whose digitization results are incomplete or unclear shall be re-digitized; if the digital file organization does not match the physical archive, adjustments shall be made promptly.

**11. Data Mounting**

11.1 Use relevant software to mount the catalog data in the database to its corresponding physical archive digital image to establish the association between catalog data and digital images.

11.2 Check the mounting results item by item, including the accuracy of the association between catalog data and digital images of physical archives, consistency between mounted digital images and the actual scanning quantity, whether digital images can be opened normally, etc. Correct errors promptly when found.

**12. Digitization Results Acceptance and Handover**

12.1 Acceptance method

12.1.1 It is recommended for the archives department to establish a specialized acceptance team to accept the physical archive digitization results.

12.1.2 A combination of computer automatic inspection and manual inspection shall be used to accept the physical archive digitization results.

12.2 Acceptance content

12.2.1 The physical archive digitization results include digital images, catalog data, metadata, work files generated during digitization, storage media, etc.

12.2.2 Acceptance of catalog data shall mainly include the accuracy and completeness of the content and format of each item in the database, whether mandatory items are filled in, etc.

12.2.3 Acceptance of metadata shall mainly include the completeness and normativeness of metadata elements, etc.

12.2.4 Acceptance of digital images shall mainly include the accuracy of digitization parameters, storage paths, naming accuracy, integrity of images, accuracy of arrangement sequence, image quality, etc.

12.2.5 Acceptance of data mounting shall mainly include the accuracy of the association between catalog data and digital images, etc.

12.2.6 Acceptance of work files shall mainly include the completeness and normativeness of work files.

12.2.7 Acceptance of storage media shall mainly include the usability and absence of viruses.

12.3 Acceptance criteria

A 100% inspection shall be conducted, with a qualification rate of 100%.

12.4 Acceptance conclusion

12.4.1 The quality inspection of the results of each physical archive digitization results meets the requirements of these Specifications 12.2 and

12.3 , and is accepted as "pass". If acceptance is not passed, rework or modifications shall be made as necessary, and then re-accepted.

12.4.2 After acceptance is completed, it must be signed by members of the acceptance team. The conclusion of "pass" acceptance must be reviewed and signed by relevant responsible person to be valid.

12.5 Handover

Data that has passed acceptance shall be handed over in a timely manner according to the physical archive digitization plan, and handover procedures shall be completed. See Appendix B.8 for the handover form example.

## 13. Return to Archive Storage

The original archives that have undergone digitization processing shall be handled and counted according to the relevant requirements for archiving of physical archive, and archival storage procedures shall be followed. Examples of special digitization warehouse in/out forms of physical archives are provided in Appendix B.9, and examples of data quality inspection return forms are provided in Appendix B.10.