Digitization Specifications of Shenzhen Institute ofAdvanced Technology, Chinese Academy of Sciences on Paper Archives (Trial)

1 General

To establish a sound system for digitizing archives and standardize the digitization of paper archives in the Institute, fully leveraging the role of electronic archives in our scientific research and management work, in accordance with the *Guide for the Digitalize Processing of Document and Archives* (GB/T20530-2006) issued by the National Archives Administration, the *Basic Terminology of Archive Work* (DA/T1), the *Archive Recording Rules* (DA/T18), and the *Information and Documentation - Implementation Guidelines for Digitization of Records* (ISO/TR13028), these Specifications are specially formulated. These Specifications specify the technical and management requirements for the digitization of paper archives and are applicable to the management of the digitization process of paper archives using scanning equipment.

2 Terminology

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

3 Definitions

3.1 Digitization

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

3.2 Digital Image

An integer array representing a physical image. A two-dimensional or higher-dimensional sampled and quantified function generated from continuous images ofthe same dimension.

3.3 Digitization ofPaper Archives

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

3.4 Resolution

The number of points or pixels contained in a unit length of an image, generally expressed in dots per inch (dpi).

4 Basic requirements

4.1 The Comprehensive Archives Room of SIAT shall plan and carry out the digitization of paper 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 paper archives shall follow the requirements and suggestions proposed by ISO/TR13028 and GB/T20530.

4.2 The basic stages of the digitization of paper archives mainly include: pre-digitization processing, establishment of catalog databases, archive scanning, image processing, data linking, acceptance and handover of digitization results, etc.

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

4.4 During the digitization process of paper 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.

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

4.6 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 The Comprehensive Archives Room is responsible for establishing the organization for the digitization of paper archives, carrying out overall planning, organizing implementation, coordinating management, ensuring security, providing technical support, conducting supervision and inspection, and accepting results to ensure the smooth progress of digitization work.

5.1.2 The Comprehensive Archives Room shall be equipped with personnel with corresponding capabilities, 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 paper 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. To enhance the security of digitization work, external contracted personnel shall undergo rigorous review.

5.2 Infrastructure

5.2.1 The Comprehensive Archives Room shall be equipped with dedicated processing areas and arrange them reasonably to form work areas for archive storage, pre-digitization processing, archival recording, archive scanning, 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 not be detrimental to the protection of archive entities. Facilities and equipment for fire prevention, waterproofing, prevention of harmful organisms, anti-theft alarm, video surveillance, and other safety management shall be provided to cover the entire area.

5.2.3 Facilities and equipment for paper archives digitization shall be reasonably planned, equipped, and managed to ensure safety, advancement, and the ability to meet the needs of digitization work.

5.3 Work plan

5.3.1 Based on thorough research, the Comprehensive Archives Room shall develop scientifically reasonable work plans to ensure that the digitization of paper archives achieves the intended goals.

5.3.2 The work plan for digitization of paper 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. The determination of digitization targets shall consider factors such as the rarity, openness, utilization rate, urgency for rescue, and digitization funds of the archives.

5.3.3 If necessary, the Comprehensive Archives Room may need to have the digitization work plan of paper archive reviewed by experts to ensure its scientific, standardized, and reasonable nature.

5.3.4 The paper 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 The Comprehensive Archives Room is responsible for establishing scientific and standardized management systems and strictly implementing them during the work process to effectively guarantee archive security and the quality of paper archive digitization results.

5.4.2 The management system for digitization of paper 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 The Comprehensive Archives Room shall establish relevant workflows, operation specifications, etc., based on relevant laws, regulations, and various technical standards, to effectively control the entire process of digitization of paper archives and ensure the quality of digitization results. See Appendix A for the paper archive digitization process at SIAT.

5.5.2 Enhanced safety management shall be applied throughout the entire process of digitizing paper archives.

5.5.3 A comprehensive feedback mechanism shall be established to provide timely feedback and corrections for issues identified in the backend stages of the paper archive digitization process.

5.6 Work document management

5.6.1 Work documents for paper 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 paper archive digitization, data acceptance forms, project acceptance reports, handover lists of paper 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 paper 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 paper archive digitization needs to be outsourced, the Comprehensive Archives Room will 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.

5.7.2 During project implementation, strict safety management requirements shall be implemented from various aspects such as the Comprehensive Archives Room, 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 ofArchive Digitization Outsourcing* (DBF [2014] No. 7).

5.7.3 The Comprehensive Archives Room assigns dedicated personnel to supervise and guide the outsourcing ofpaper 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 paper archive digitization work plan, the Comprehensive Archives Room conducts preliminary preparations such as archive retrieval, inventory, and registration for the digitization objects. After obtaining approval from the Public Finance Department leadership, 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 paper archive digitization process, a storage warehouse located close to the digital processing site shall be set up to temporarily store paper archives. Strict management shall be carried out for the receiving and return ofpaper archives, including careful inspection, inventory, registration, etc., to ensure the safety ofpaper archives.

7 Pre-digitization Processing

7.1 Determining scan pages

In principle, all paper archives identified as digitization objects shall be scanned, and selective scanning is not advisable. Any pages that do not require scanning shall be clearly marked.

7.2 Numbering pages

7.2.1 Page numbers shall be renumbered for archives without page numbers or with incorrect page numbers.

7.2.2 When renumbering pages, the numbers shall be written in a uniform position without covering the archive content.

7.2.3 The pens, ink, etc., used for writing page numbers shall not damage the original documents or have a long-term impact on the preservation of archives.

7.2.4 Special cases such as damaged pages or missing pages shall be recorded.

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 content of the original documents.

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

7.4 Dismantling bindings

Whether to dismantle bindings shall be determined based on the principle of protecting paper archives. If bindings need to be removed, care shall be taken to ensure that the archives are not damaged, and archives with incorrect sequence shall be rearranged. For special bindings that need to be restored after dismantling, the

original appearance of the archives shall be recorded through methods such as photography to facilitate restoration.

7.5 Technical repairs

7.5.1 For paper archives with severe damage or unable to be scanned directly, technical repairs shall be carried out by professional technicians.

7.5.2 Paper archives with wrinkles or unevenness that affect scanning quality shall undergo corresponding technical treatments such as flattening first.

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 paper archive catalogs and the establishment of catalog databases.

8.2 Database selection shall consider compatibility with universal data formats for ease of 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 paper archive digitization results.

8.4 The results of modifications and supplements to paper archive catalogs during pre-digitization processing shall be entered into the database to form accurate and complete catalog data.

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 Archive Scanning

9.1 Basic requirements

Archive scanning shall involve selecting appropriate scanning equipment based on the actual conditions of paper archive originals, digitization purposes, digitization scale, computer network, and storage conditions, and setting and adjusting 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.

9.2 Scanning equipment

9.2.1 The selection of scanning equipment shall pay special attention to the protection of archive entities, preferably using scanning equipment with minimal damage to archive entities for digitization.

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

9.3 Scanning color modes

9.3.1 To preserve archive original information to the fullest extent and facilitate various uses, it is advisable to scan using the color mode for all pages.

9.3.2 Pages containing letterheads, seals, photos, color illustrations, or multicolor text shall be scanned in color mode.

9.3.3 Page images in black and white, with clear text and no illustrations, can also be scanned using black and white binary mode.

9.3.4 Page images black and white pages, but with poor text clarity or containing illustrations, can also be scanned using grayscale mode.

9.4 Scan resolution

9.4.1 The selection of scan resolution shall ensure clear and complete images after scanning, taking into account factors such as the later use of digital images.

9.4.2 The scan resolution shall not be less than 200dpi. For documents with small, dense, or unclear text, a scan resolution of not less than 300dpi is recommended.

9.4.3 For other purposes such as COM output, simulation replication, or printing and publishing, the scan resolution can be adjusted as needed. Documents requiring COM output shall have a scan resolution recommended not less than 300dpi; for documents requiring high-precision simulation replication, a scan resolution of not less than 600dpi is recommended; for documents requiring printing and publishing, appropriate resolutions shall be selected based on factors such as document size, printing format, and printing accuracy.

9.5 Storage format

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

9.5.2 When using paper 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 PDF.

9.5.3 Use the same storage format for the same batch of documents.

9.6 Image naming

9.6.1 Digital images shall be named based on the file No. The selection of image naming methods shall ensure the uniqueness of image names.

9.6.2 It is recommended to store digital images as single-page files and name them based on a combination of file No. and image serial number.

9.6.3 When digital images need to be stored as multi-page files, the file No. of the document can be used for naming.

9.6.4 Scientifically establish the storage path for paper archive digital images to ensure accurate data mounting.

10 Image Processing

10.1 Image stitching

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

10.2 Rotation and correction

Digital images that do not match the reading direction shall be rotated to restore them. Skewed images shall be corrected to visually eliminate skewness.

10.3 Trimming

If digital images need to be trimmed, the images shall be cropped at least 2 to 3mm from the outer edge of the page.

10.4 Cleaning

If digital images need to be cleaned, impurities such as spots, lines, or black edges that affect image quality generated during the scanning process shall be removed. The principle of presenting the original appearance of the document shall be followed, and the process shall not remove original paper fading spots, water stains, spots, or binding holes.

10.5 Image quality inspection

10.5.1 When digital images are incomplete, indistinguishable, or heavily distorted, they shall be re-scanned.

10.5.2 For cases of missed scanning, repeated scanning, or excessive scanning, corrections shall be made promptly.

10.5.3 When the arrangement sequence of digital images does not match that of the original documents, adjustments shall be made promptly.

10.5.4 Inspect the stitching, rotation and skewness correction, trimming, and cleaning of digital images. If they do not meet the image quality requirements, image processing shall be redone.

11 Data Mounting

11.1 Use relevant software to mount the catalog data in the database to its corresponding paper 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 paper 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 paper archive digitization results.

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

12.2 Acceptance content

12.2.1 The paper 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 ofmetadata elements, etc.

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

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

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

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

12.3 Acceptance criteria

Projects that can be automatically inspected by computers shall undergo 100% inspection using computer automatic inspection methods, and the acceptance rate shall be 100%. For projects that cannot be automatically inspected by computers, manual inspection can be conducted on a sample basis by case or volume. The sampling rate shall not be less than 5%. For the accuracy ofthe association between database entries and digital image content, the acceptance rate of sample inspection shall be 100%, and for other content, the acceptance rate of sample inspection shall not be less than 95%.

12.4 Acceptance conclusion

12.4.1 Each batch of paper archive digitization results shall meet the requirements of Sections 12.2 and 12.3 of these Specifications for quality inspection, and be 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 leaders to be valid.

12.5 Handover

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

13 Return to Archive Storage

13.1 Archive Binding

After the paper archive digitization is completed, if documents that have been bound need to be rebound, care shall be taken to maintain the original appearance of the documents, ensuring safety, accuracy, and completeness.

13.2 Return to Archive Storage

Handle and count the paper archives according to the requirements for archive storage and complete the procedures for archive storage.