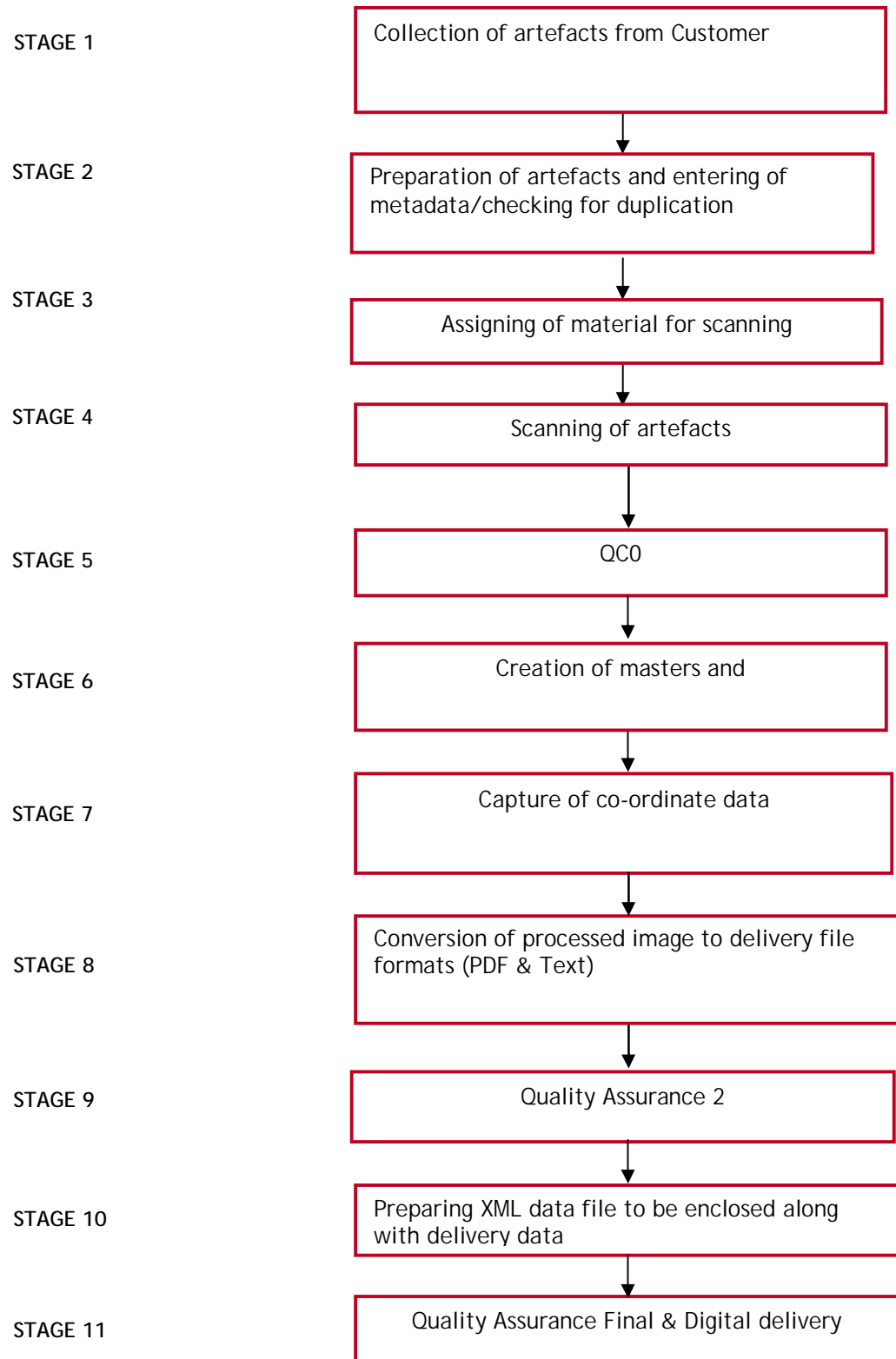


2.1 Workflow Block Level:



Stage 1: Collection of Artefacts from Customer for scanning

Artefacts collected from Customer will be individually be counted and accounted for. The condition and the categories of the documents will be noted and will be segregated accordingly. A receipt would be created for the documents which are initialled by representatives of Customer.

Format for receipt

Date:

Receipt No:

S.No	Artefact Index	Description	Category of Artefact	Type of media	No of Pages	Condition of media	Need for filming Y/N	Metadata Y/N	Date of receipt	Receiving person signature	Dated of return	Remarks and signature on return
01	A001-P001	Diary of Mr. Sharma	A	Book -	125	Good		Y				
02	B001-F006	Microfilms of Andhra Patrika 1942	B	Micro films	650	Fair		N				
3	A002-P006	Photo Albums of Independence Day	A	Photos	45	Needs repair		N				

Stage 2: Preparation of Artefacts, checking for duplication and entering of metadata

On receipt of the Artefacts from Customer the artefacts the following is done

- a. Accounting is done according to the receipts wrt the received quantity.
- b. All artefacts are cleaned removing lint and dust
- c. Minor repairs to documents are done
- d. Film rolls/microfiches are ensured to be free of lint, and dust this is of a severe consequence as these artefacts are scanned with magnification ratios of 10 -29.6 and any spec on the artefact can become a very large noise band on the scanned image.
- e. Each artefact is bar-coded and numbered so that the subsequent accession is through this barcode
- f. Metadata is associated with the accession numbers and entered into the Workflow
- g. During entry of the metadata Duplication is checked and duplicates kept aside.

The scanners that are used are going to be overhead scanners and as far as possible we would like to avoid debinding of any artefacts. Only in cases where the text enters into the bind would we need to de-bind.

The prepared artefacts are now ready for the assignment to scanning

Stage 2A: Preparation and calibration of scanners/equipment and scanning rooms.

All equipment that is used for the archiving process need to be calibrated to the local environment. Factors such as direct light reflections from walls reflections from the artefact to be scanned etc will need to be considered and the equipment accordingly calibrated to the local/room conditions. Calibratable monitors are used wherever the scanning is to be done in color.

While preparing the site it may be required to repaint the walls and the ceilings of the scan area with inert colours which are non reflective. Further the ambient lights will need to be consistent avoiding shadows and glares.

API/ Kodak test targets will be used for the calibration. While this is not a one time effort the frequency may vary from many times in a day depending on the mother media to once in a week with consistent media.

Following are the key image acquisition parameters that will be considered while calibrating

- a. Linearity- To ensure that the size of the artefact remains the same as the original.
- b. Resolution- To ensure that there is true reproduction of resolution
- c. Scan depth/ field of vision - to ensure that for thicker artefacts image acquisition is consistent across the artefact and no blurring.
- d. Color Reproduction- To ensure that the final color/ grey scale tones are in consonance with the original artefact.
- e. Density of light across the scan bed.

The tables and the chairs that would be used for scanning would be ergonomic to ensure comfort for the people working with the project and the artefacts.

Stage 3: Assigning artefacts for scanning

This activity is crucial and the artefacts received are assigned to the various processes depending on the category of the artefact and the kind of artefact.

- f. Microforms- Require Microfiche and Microfilm scanners with 300 dpi grey scale

- g. Paper – Using overhead scanners with A, B, C categories requiring 600dpi/300dpi, color/grey scale and of various sizes.

Treatment of each of these artefacts will therefore differ and image acquisition will need to happen through different equipment operators of each with skills in handling the respective media and the equipment.

The supervisor will through the workflow assign the respective artefacts by their accession nos to the respective persons which will appear in their scan queue in their workflow logins. This ensures that only assigned artefacts are scanned and with the details of their categories and types the settings of the scanners can be pushed through the workflow.

Once the artefacts have been assigned the scanning operators would also physically take custody of the artefacts from the local store/book bank.

Stage 4: Scanning of artefacts

Every artefact is on different mother media, has categories has a different set of digital parameters for scanning. From the metadata entered which will contain the type and category of the artefact the workflow will push to the scanner the settings for the resolution file type etc. This will to ensure that these are adhered to automatically and will reduce the error in scanning and more importantly avoid handling of the physical artefacts the second time.

Each operator will only be able to scan the artefacts that have been assigned to him which will appear in his scan queue once he logs into the work flow.

When he clicks on the scan button, based on the agreed schema the folder names and the file names are created. Normally the names of the files are the artefact names preceded by the accession codes all spaces are replaced by underscores, and the files are named with 8 digit numbers with preceding zeros. This is configured in the work flow and the naming details are taken from the metadata of the artefacts. This mechanism ensures that there are no errors brought in by misreading or typos.

While scanning the operator through the workflow screen will also have a matrix where he enters any problems with the artefact or part of it in while scanning. These are normally done if there is a paper torn or if there are any special comments he needs to make on the papers or the frames within the artefact, essentially to highlight any exceptions with the artefacts. Once he completes the scanning process the workflow automatically takes the scanned artefact into the queue for the next process, QC0.

Stage 5: Assigning to QC0

QC0 is the quality check at the 0th level of the workflow process. This is the primary check but the most important check as if the quality of any further processing can be only as good as the original raw scan. Once the scanning of an artefact is over it then appears as a queue element of the QC0 stage. The work flow also maintains a track of the location of the digital data that has been scanned.

The supervisor then allocates/assigns the digitized artefacts to the QC0 operator.

Stage 6: QC0 -ensuring the quality of the raw scanned artefacts

As mentioned the Raw scanned images form the basis for the master and rendering copies and also the archiving copies it is necessary to ensure that they are of utmost quality standards.

Once the QC0 operator logs in he finds in his work queue the artefacts that he has to check for the quality control. He has then an option of either copying the files from the source location or checking them on the source location itself. This consideration is made based on the sizes of the files on the hard disks.

At the QC0 level there are two kinds of checks made on the artefacts.

- a. Parametric
- b. Qualitative

The parametric check is an automated check through the Audit tool that is incorporated in the work flow. This ensures that the measurable parameters like the

- a. image type(color/grey scale/Bw)
- b. Resolutions
- c. Compression
- d. Blank pages
- e. File size limitations if any
- f. Linearity of the documents
- g. No of pages against the not of pages in metadata

If the Parametric check fails the artefact is rejected and sent back to the production floor for rescanning, with a list of errors in the artefact. This list is a text file that is generated through the Audit tool and has page wise conformance/non conformance it will reside in the respective folder of the digital artefact.

If the parametric check is cleared the artefact is subjected to an eye ball check of every page. Any loss of information within pages or pages missed out (if numbered pages) or if there is either a skew, flying text, noise, visibility, or any other image acquisition/reproduction problems it is rejected. There is a facility to make comments on the respective pages that have the problems or if the entire artefact has problems. Even if one page is not conforming to the Quality standards the entire artefact is rejected.

On passing the QC0 the digital artefact has qualified to become the Archive Copy and can be sent to the next stage to be made to a Master copy. The Archive copy can at this stage be copied onto a backup for storage.

Stage 6: Assigning to Image Processing

Once an Archive copy has been created it would need for a master copy to be generated from the Archive copy. A QC0ed artefact would now come automatically in the queue for the next stage in the workflow i.e. assigning to Image processing. A supervisor would then through his console assign the digital artefacts to the respective image processing operators.

Stage 7: Image processing and image enhancement of digital artefacts

Once the Processing operator logs into the workflow he would in his work queue find the digital documents that have been allocated to him for image enhancement.

The source location of the digital copies of the digital artefacts would also be prompted by the workflow and the operator has the option of either copying the images to a different location or working of the source location.

The very first operation that is done in image processing is cropping of the images to remove peripheral noise and crop the image. This is essential as DE skewing of the images etc will not be possible until the images are free from peripheral noise. Any other large noise bands or noise artefacts that are present on the image are also deleted. Once the

cropping is over the image enhancement tools are applied. Various parameters are used depending on the mother media of the artefact.

For paper documents- If the documents are old, paper acidic or stored in areas of too high or too low humidity the paper tends to have acid blotches, or noise or tear marks. Old typewriting and print tends to leave spaces within characters.

For Microfiches/films- The images that have been captured need to be processed to eliminate noise and enhance to bring out the relevant information clearly. Often images, especially those that have been captured from old microfilm / microfiche or scanned images of typewritten documents, have variations in image quality from one end of the image to the other based on lighting conditions while capturing the image or on the quality of the medium that the image has been stored upon. Microfilms and Microfiche are especially susceptible to problems such as fogging of microfilm, blurred images due to camera shaking or inappropriate lighting at the time of image capture, under or over exposure at the time of microfilm making, fading of images due to both poor quality original images as well as poor microfilm storing techniques, and severe contrast issues from one end of the microfilm image to the other.

While the image processing is the same irrespective of the mother media additionally tonal consistency and DE blurring are done besides the following processes.

- a. DE skewing
- b. Bringing tonal consistency across the page
- c. De-blurring if within the readable scale
- d. Correcting
 - i. Brightness
 - ii. Contrast
 - iii. Gamma Correction
 - iv. Smoothing/Sharpening the image
 - v. DE speckling
- e. DE speckling in case of black and white images
- f. Padding of the images for a uniform border
- g. Resizing all pages in case the original artefact is a book or news paper with pages of same size.
- h. Applying compressions to the digital artefacts

The software for the above processes would currently be Book Expert/Image Process Plus or Scanfix for b/w images. Any kind of image processing tools could be integrated with the workflow. The image processed artefacts would now go the next stage and appear in the QC1 assignment queue.

Stage 8: Assigning to QC1 - Quality check for the master images

Once the images of the artefacts are cleaned and processed the quality needs to be checked for ensuring that the master images are of acceptable quality. The supervisor now assigns the artefacts to the QC1 operator. Once assigned the artefacts appear in the Work queue of the QC1 operator.

Stage 9: QC1 Quality check for creation of master images.

Once the operator logs in, he will find that the digital artefacts that he has to check in his work queue of his workflow.

QC1 is like QC0, a two staged quality check with an automated parametric check and an eyeball check. The parameters would again depend on the specs of the masters for the

various mother media and the category of the media. Here again the automated check checks for

- a. Image type and color depth (color/greyscale/bw)
- c. Resolutions
- d. Compression
- e. Blank pages
- f. File size limitations if any
- g. Linearity of the documents
- h. Additionally consistency of page sizes -against original is consistent.
- i. No of pages scanned and processed are consistent with the metadata.

If the artefact fails the parametric test it is sent back to processing of it passes the second eyeball test is done and the following are checked for

- a. Skew in pages
- b. Noise/Noise artefacts
- c. Background color
- d. Padding
- e. Blurring
- f. Consistency across the page

If the digital artefact clears the QC1 then it qualifies a Master copy and can be archived a digital copy. An automated process of creation of thumbnails is done. The digital document is ready for the next stage of internal indexing and would appear in the queue for assigning to Indexing, structural metadata and/or OCR.

Stage 10: Assigning for structural metadata.

The digital artefacts will have to be internally indexed and the structural metadata is created depending on the category of the artefact following are the categories

- a. Newspapers
- b. Photo albums
- c. Books
- d. Collections
- e. Handwritten manuscripts

Strategy for the internal indexing is predefined depending on the contents of the artefacts and is assigned to the operators who are skilled and tooled with the respective tasks.

Stage 11: Creation of the Structural metadata

As each of the Structural metadata operators logs in he will find he artefacts in his queue. He selects the artefact and does the indexing depending on the media.

1. Newspapers
 - a. Zoning of the articles
 - b. OCRing the heading of the articles (only English)
 - c. Automated saving of the metadata in the prescribed schema
2. Books- As per the contents page following are not exhaustive and dependent on the book.
 - a. Contents page
 - b. Publisher
 - c. Chapters
 - d. Illustrations
 - e. Indexes
 - f. Blank pages

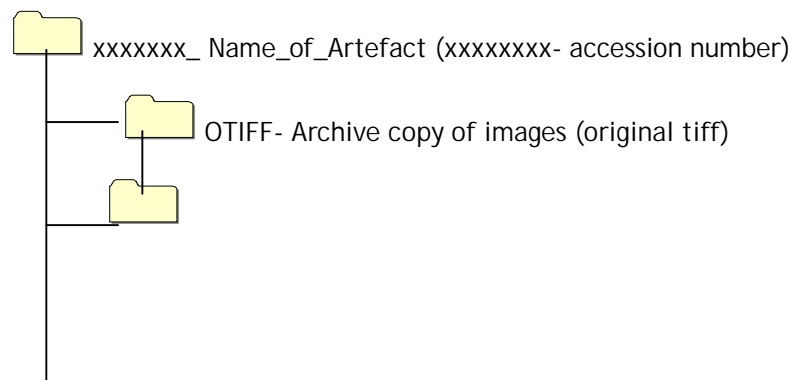
- g. Other classifications as per the contents page
- 3. Photograph Albums
 - a. Photographer
 - b. Photograph Date
 - c. Description
 - d. People in photo
 - e. Donor
 - f. Context
 - g. Negative number
 - h. Place taken
 - i. Event
- 4. Collections
 - a. Event
 - b. Description
 - c.
 - d.

Note: the schema that has been devised is beyond the fields that the operator would fill in. We have from the RFP and discussions in pre-bid meets gathered that these are the only fields. Lot of fields may not be available on the artefacts but facility has been provided to enter these fields. Many may need subject experts for population of the fields.

Stage 12: Packaging of the Digital deliverables

Once the artefacts have been structurally indexed they would be packaged for digital delivery following would be a suggested delivery mechanism. The details can be frozen before the commencement of the project.

Folder structure



Stage 13: Final quality checks

Once the digital deliverables have been packaged a Document format auditor is used to ensure that all the files are consistent with the original and the following

- a. All file and folder naming conventions are adhered to
- b. No of files are consistent across the 4 digital copies
- c. All the files in the respective folders areas per the digital standards specs.
- d. All the metadata files are present
 - i. Artefact Meta data
 - ii. Structural metadata
 - iii. Administrative metadata
 - iv. QC reports at all stages

Once the quality check is made the digital delivery is made to the IDMS which will segregate the respective folders and metadata to the various backup or rendering locations or databases and

Stage 10: Uploading of the images to IDMS

Stage 11: Return of media to Customer