

Planning Digitization Projects

Instructions for Using This Job Aid

Use this job aid when you are assigned to planning local digitization. Check off each step when you have completed the step.

Procedure	Actions/Decisions	
1. Define the goal of the proposed scanning project.	1. Identify stakeholders.	<input type="checkbox"/>
	2. Plan and schedule a meeting or similar mechanism to distribute project information.	<input type="checkbox"/>
	3. Explain to stakeholders why the project is necessary and what will be achieved by scanning.	<input type="checkbox"/>
	4. Collect feedback about the project from stakeholders to create a goal that is specific, measurable, achievable, relevant, and time-bound.	<input type="checkbox"/>
2. Define the scope of the proposed scanning project.	1. Identify what records will be scanned.	<input type="checkbox"/>
	2. If scanning active records, then determine whether to scan related records as a single file.	<input type="checkbox"/>
	3. Organize a meeting with end-users to determine the following: <ul style="list-style-type: none"> • Are they accessed frequently? • Do they require access by multiple users? • Do they have a long retention period? • Are the records essential to the organization for disaster recovery? Based on collected information, decide whether to scan related inactive records.	<input type="checkbox"/>

	<p>4. Determine the volume of records to be scanned.</p>	<input type="checkbox"/>
	<p>5. Determine whether the project will be one-time or is part of a recurring business process.</p> <p>A. If records to be scanned are active and created as part of an ongoing business process, then the project will be recurring.</p> <p>B. If records are inactive or no longer being created, then the project will be a one-time occurrence.</p>	<input type="checkbox"/>
	<p>6. Determine whether the original documents will be retained or disposed of after scanning.</p> <p>A. If the original documents have an approved temporary retention period, then consult with the office of origin regarding whether they can be destroyed after quality verification of the digital images.</p> <p>B. If the original documents are to be retained, then estimate those resources necessary to keep the textual records for their authorized retention period, whether onsite or offsite.</p> <p>C. If the original documents are to be replaced by scanned images, then plan for their destruction as non-record material after scanning is completed. Document destruction costs as part of the project's cost-benefit analysis.</p> <p>D. If the original documents are scheduled for permanent retention, then consult with the office of origin regarding whether the digital images will serve as reference copies only or are of sufficient quality (according to NARA standards) in terms of pixel bit depth (black/white, gray, or color) and resolution (dots per inch), to replace the scanned documents. However, if the scanned original documents have intrinsic archival value, as identified by the office of origin, then retain them permanently, regardless of the quality of the scanned duplicates.</p>	<input type="checkbox"/>
	<p>7. Choose the appropriate file format for the scanned images (TIFF, JPEG, PDF, etc.).</p> <p>A. If available, use open-source, non-proprietary file formats.</p> <p>B. If not available, then use accepted proprietary formats that are industry standards and well supported by developers.</p> <p>C. If the scanned records are Federal and to be retained permanently, then select a file format that is approved by NARA for transferring permanent electronic records.</p>	<input type="checkbox"/>

	<p>8. Identify the required quality of the digital images (dots per inch [dpi] resolution, black and white vs color, etc.).</p> <ul style="list-style-type: none"> A. Choose a resolution at a minimum of 300 dpi. B. If only black and white images are to be displayed, use 1-bit depth, if only shades of gray are to be displayed, use 8-bit depth, and if color is to be displayed, use 24-bit depth. 	<input type="checkbox"/>
	<p>9. Determine whether the digital images will be compressed to save server or offline storage space. If data compression is acceptable, then decide whether to use lossy or lossless format.</p> <ul style="list-style-type: none"> A. Decide whether compressing a file is an acceptable option to save digital storage space or whether the file must be stored as scanned. The amount of available storage space (either online or offline), the content and purpose of the file, and its eventual disposition all influence this decision. B. If compression is acceptable, then determine whether to use a lossless or lossy format. C. If the file is to appear identical to the uncompressed text version and/or be retained long-term (more than 10 years) or permanently, then use a lossless compression format. D. If the file is to be retained temporarily and/or an approximation of the original document is acceptable, then use a lossy compression format. 	<input type="checkbox"/>
	<p>10. Identify those metadata fields (tags) that are necessary to identify, index, and view the images efficiently, and also provide for their preservation and eventual disposition. The type and number of metadata fields selected will be based on user requirements and transfer.</p> <ul style="list-style-type: none"> A. Determine what fields will be automatically filled in by the scanning system or software, such as date, format of the records, and a unique identifier for each image. B. Determine what other metadata fields will be manually added for retrieval purposes, such as document title, type, records series, creator, or disposition date. C. Estimate the resource (i.e. time, money, staff, etc.) costs involved in any required manual data entry. This may influence the number of metadata fields ultimately added. D. Establish either keystroke or sight verification processes for keying error detection and estimate resource costs for the selected quality audit process. 	<input type="checkbox"/>

	<p>11. Standardize metadata element and term assignment for identified metadata fields (Step 2.8).</p> <p>A. If standard metadata elements and terms are desired, then choose an accepted convention, such as the Dublin Core Metadata Initiative documented in NARA Bulletin 2015-04.</p> <p>B. If not, use metadata elements and terms provided by the records' office of origin.</p>	<input type="checkbox"/>
	<p>12. Determine image quality control/audit protocols.</p> <p>A. If the scanned images will be the record copies, then visually check every image for positioning, missing pages, legibility, and printability.</p> <p>B. If the scanned images are for reference only, review a sample of the scans. Determine the percentage of images to be reviewed based on business need.</p> <p>C. Confirm that the number of original documents scanned match the number of images produced.</p> <p>D. Check indexing metadata at the same time as the images, using the same record or reference use criteria documented in A. and B.</p>	<input type="checkbox"/>
	<p>13. Identify how the images will be maintained and stored during their life- cycle.</p> <p>A. Identify available storage choices, both online and offsite.</p> <p>B. If the scanned data is to be stored onsite, then consider the organization's present and future digital storage capacity by consulting with IT.</p> <p>C. If immediate access is required, then store images online in a shared or restricted drive or near-line.</p> <p>D. If delayed access is acceptable, then store scanned images offline.</p>	<input type="checkbox"/>
	<p>14. Coordinate with IT to ensure there are provisions for periodically backing up the digital data to prevent unintended loss.</p>	<input type="checkbox"/>
3. Assess the documents to be scanned and indexed to determine project workload and costs.	<p>1. Determine the physical condition of the original records. (Are they damaged, faded, oversize, onion skin, etc.?)</p>	<input type="checkbox"/>
	<p>2. Determine how much document preparation (removing staples and paper clips, repairing documents, sequencing, etc.) will be required to make the files ready for scanning.</p>	<input type="checkbox"/>

	3. Determine whether the records have access restrictions due to content. If the records to be scanned have access restrictions, then so will the scanned images.	<input type="checkbox"/>
	4. If access is to be restricted, then consider linking access to log-in type, flagging confidential records so they are filtered out of search results, or segregating confidential records so access can be controlled through a separate process.	<input type="checkbox"/>
	5. Identify the retention period(s) of the records to be scanned by referring to the organization's records schedule(s). If the records to be scanned have a short retention period, then consider in Subtask 5 whether scanning them is cost effective and serves a business need.	<input type="checkbox"/>
4. Based on the type and quality of the documents, decide whether to use Optical Character Recognition (OCR) for enhanced image searching and retrieval.	1. Ask potential users if full-text searching or editing of the scanned documents is a desired capability or required for Section 508 compliance.	<input type="checkbox"/>
	2. Determine the additional costs, if any, of implementing OCR, for inclusion with the cost-benefit analysis in Subtask 5.	<input type="checkbox"/>
5. Conduct a feasibility study regarding the proposed scanning project.	1. Identify potential options for the proposed project: perform the scanning in-house, outsource to a vendor, or continue using textual documents as is.	<input type="checkbox"/>
	2. Perform a cost/benefit analysis (CBA) for each option. <ul style="list-style-type: none"> A. Document the current process. B. Estimate future requirements. C. Consider options. D. Collect cost data for each option. <ul style="list-style-type: none"> a) If hardware and software for in-house scanning must be procured, then collect cost information from quality sources. 	<input type="checkbox"/>

	<ul style="list-style-type: none"> b) If considering outsourcing to vendor, account for all service fees. c) If continuing the “as is” process, then account for current costs. E. Estimate the cost of each option, normally in dollars. F. Estimate the benefit of each option, normally in dollars. G. Determine those option(s) with a net benefit. If benefits exceed costs, then a net benefit results. H. Identify the preferred option from the organization’s perspective. 	
	3. Confer with IT to determine its technical feasibility.	<input type="checkbox"/>
	4. Gauge potential customer satisfaction with expected project deliverable. If resistance to the proposed project from either internal or external users is expected, then decide whether the project is still worth initiating.	<input type="checkbox"/>
6. Based on the feasibility study and customer feedback, request management decide whether the scanning project will be completed in-house, outsourced to a vendor, or not done at all.	1. Recommend the preferred option from the feasibility study and request management’s decision on proceeding with the project.	<input type="checkbox"/>
	2. If the feasibility study and/or customer response does not support scanning, then advise management not to proceed.	<input type="checkbox"/>
7. If scanning is the chosen option, then develop a preservation strategy for any digital images that are records and require long-term or archival retention.	1. Identify those digital series that require long-term or permanent retention.	<input type="checkbox"/>

- | | | |
|--|---|--------------------------|
| | <p>2. Choose a preservation strategy based on available resources.</p> <ul style="list-style-type: none">A. If financial resources are available, then choose migration or emulation to preserve and keep the digital images accessible.B. If financial resources are not available for migration or emulation, then plan to copy the digital records to new media periodically. | <input type="checkbox"/> |
|--|---|--------------------------|