

Document Preparation for Electronic Document Management System

Mohamad Azri Mohamad Radzi, Saiful Farik Mat Yatin, Nurul Azira Najihah Mohd Fadzil, Syed Hamzah Jamallulail Syed Abdul Aziz

Faculty of Information Management, Puncak Perdana Campus, UiTM Selangor, Malaysia

To Link this Article: <http://dx.doi.org/10.6007/IJARBSS/v8-i9/4583>

DOI:10.6007/IJARBSS/v8-i9/4583

Published Date: 13 October 2018

Abstract

This article discusses the topic of document preparation for archival institution or records repositories before the materials been digitize. Document preparation is a process that came across before the scanning process. Topics include types of document that are collected for digitizing, criteria that makes the documents collected for digitization, steps of document preparation for before the digitizing process, during the process and post document preparation. The methods used benefits and the challenges of preparing document for digitization are also included in the article. Recommendations also been proposed to propose new ideas and advice on the topic discussed. This article devotes readers into understanding more about the importance of document preparation process in digitization project implementation in organization.

Keyword: Document Preparation, Digitization, Scanning Document, Content Value, Document Digitization

Introduction to Document Preparation

Digitization is a process that converts information from documents into a digital form document. Paper or non-digital files and documents have a very important part in the growth of the world since the previous eras. However, these files and documents will become harder to be hold or even look at it clearly because of the fragility of the files and documents physical. Before digitizing those documents, it is a must to make a document preparation process. Document preparation process is a process of collecting documents or records to be digitize. However not all documents or records are necessary to be digitize. All documents or records need to be observed their appropriateness before digitizing them. When the physical state of the documents is not tolerable, it is supposed to be fixed first before undergoing a digitization process. Usually, the physical and characteristics of the documents or records are the reason why some of them need to be digitized. Document preparation is the process of determining all of these problems because it helps in easing the digitizing process. It is very important for the documents or records to be undergoing the digitize process. Without digitizing them, it

will be a problem to obtain information from the past in the future. This is because there might be important information on those documents that is needed in the future. Especially information agencies, they should be making digitization as a practice from now on as they hold so many important information.

Types of Document Collect for Digitization

The National Archives (n.d.) stated that documents that will be digitized are in various forms, not only focuses on paper documents. This is because there are many types of documents that are stored in the information agencies. These documents even take up a lot of spaces in the organization. Digitizing them helps to reduce the usage of spaces in the organization. Besides that, it helps those organizations to have an extra copy of the vital documents in case they would be lost or damage. There are several types of documents that will be digitized:

Standard files

One of the document collect for digitization is standard files. Standard files are those documents that are paper form and put in files. These kinds of documents are normally found in record center and archives. They are the most types of documents that are placed in record center and archives. It should be digitized because it is easy to deteriorate and damage. Digitization of the standard files documents is important as it allows the documents to be able to be use although the original copies of document have already been damage.

Maps and Plans

Maps and plans are a huge type of document. It is so big that it requires many spaces for the storage of the document. Some of them are also easy to deteriorate because it is also in paper form as those documents of standard files. By digitizing maps and plans, it helps the record center or archive to safe some space for other documents. In addition, it is also a way to make those documents have a backup copy if the original copy deteriorates or damage.

Photographs and Negatives

These types of document are not paper form and some of them require a tool to be able to use them. Photographs are also in paper form and it is also easy to deteriorate or damage. Negatives are used to produce photographs and it was used back in the old days as there are no digital cameras. It also allows photographs to be produce so many times by using the negatives. It helps to avoid a photograph from being loss. However, it is also a must to digitize these types of documents as there are probabilities of damage or deterioration.

Criteria in Preparation of Document

When it comes to digitizing the file and document, Archivist or Record Manager need to consider every requirements and assessment criteria over the file and document in the repositories. The numerous of file and document with different format in the archives repository require an assessment over the document need to be digitize. This is because the goal and objective of the preparation documents before digitization is to ensure that all records determine with value and rare are to be scanned and that scanning process is performed according schedule and period agreed. It is discussed by Jervis and Masoodian (2014), one of the most prevalent roles served by paper documents was the keeping of records, such as legal documents and correspondence. Although all the offices had records of some form, there was some variety as to what was kept, and how much information was kept.

So, based on that there is need for documents to be assessed with the criteria in preparing the document for scanning process. The digitization process could also develop a comprehensive policy to aid the archivist in preparation and collection of documents. The policy should outline a proper guide and criteria that related to preparation documents before digitization process. The following criteria were necessary to be considered by Archivist during the process of preparing and collecting the documents.

The Condition of Material

Documents that are chosen for digitization process need to have significant factor that convince the documents chosen is required to be place into digitization process. The document condition should be examined before taken it to the next process. For documents with the worse condition of physical, it first needs to be taken for preservation and conservation process. It is addressed in International Federation of Library Associations and Institutions, (2002) stated that selection of material for digitization will be affected both by its physical condition and by the existing quality of the bibliographical descriptions available for it. Material which is fragile damaged and in poor condition may present too many risks of further damage being caused by handling to allow it to be scanned without special care, or some basic conservation treatment. This will involve additional costs, and the institution will need to consider whether other collections in better condition should have priority, or whether the costs of preparation and conservation should be built in to the costs of the overall digitization project.

Risk of the Material Been Damage

The archivist should assess the risk of the material to be damage by documents going through digitization process. These are importance to know the damaged rate of documents if the document were schedule for digitization. These will inform archivist about risk and what action that is need to be taken before the document were schedule for digitization. This could be done by assessing the level of damage that is acceptable for document to be digitized. The level of damage helps document to be identified on the damage occurs and the level that of damage that are accepted into digitization. The documents need to have least preservation and conservation care to allow document to be digitize.

Demand and Priorities of the Document

The users play a significant role for digitizing the documents. A specific demand of users that exist on the document may be considered as the demands over the document. If the document has a higher demand over the physical document, it is indeed needed to be considered for digitization. As indicated by International Federation of Library Associations and Institutions (2002) did mention that the level of demand is of course of great interest when selecting source material for digitization. If the purpose is mainly to enhance access, the likelihood of significant use of a digitized material will probably govern the selection process. Involving scholars and other researchers in the original decision is therefore a traditional selection methodology.

Content Value of Material

The document that require to be chosen by the archivist need to hold an intrinsic value to and unique information. This information would bring value and worth to the institution to digitize the documents. The documents must bring a representative value to the larger collection that

the institution stored. Documents must be identified and analyze the value of information and the specific purposes for documents to be digitize. The core content of the documents must be unique and rare which could generate users interests over documents digitize. So, that document will be access and used by the users effectively. Therefore, every digitization project must have its own definitions of value based on the goals it trying to achieve. If the material being considered as a candidate for digitization lacks detailed cataloguing or descriptive data, it is essential for future access to such material to create such data, and it will therefore need to be considered whether the necessary costs of doing this can be included in the overall budget of the digitization project. It was mentioned in International Federation of Library Associations and Institutions, (2002). That the document selected should have a complete content and description of documents. These to avoid institution budget spend on the completing the content of the documents in the middle of digitization process.

Intellectual Properties

The institution should have the authority to digitize the documents and mount the content of the documents online. The intellectual property rights over the documents should be addressed earlier in the process of preparing document. This is because the institution may not have the right over the documents legally which will not allow the institution to produce the document into electronic format and disseminate the documents online. If the institution does require the copyright of the documents, the digital versions must be accessed only on the institution's premises. Obtaining permission from rights holders takes time, can be expensive, and is not always possible. Observation made by Liu, (2004) mention that the development of digital technologies and the popular use of digital documents also pose an unprecedented challenge to protect copyright. Copyright protection will become more and more difficult with the ease in the reproduction and distribution of documents. So, requiring the copyright is a concern that must be address in the preparation of documents for digitization process.

Technical Aspects

The documents that follow the criteria above must be assessing either the document form and formats are able to be capture and access online. In this aspect, institution technical capability to capture, describe, store, and make digital versions accessible. The technical aspect that need to be considered such as preparation of materials, the high-quality capture of the content, Creating the user interface, with all the necessary searching and navigational tools and other necessities work to perform. The technical aspects of digitization for text, images, audio, and other genres influence the process because information can be captured in many ways at many quality levels. The institution must determine whether it can provide digital versions of the quality users need. Archivist should aware on the capabilities of the institution to capture the documents and make accessible for a long term. A mass digitization of the documents could affect the quality of the digitization of documents which lower the images, or any form produces. There should be a limit to the reproduction of documents into online forms to create a quality forms and formats of documents. A strategy should be plan by institution before planning for digitization projects. Medicaid Information Technology System, (2010) suggested the requirement for preparing inbound documents, mailboxes sorting and batching to minimize the technical aspects in preparing documents.

Stages of Document Preparation

1st stages: Preliminary Document Preparation

Conducting a survey

The aim for conducting a survey is to assess the documents before deciding the appropriate equipment that will be used to digitize the documents. There are numerous of physical aspect that needs to be considered before performing scanning operation over the documents. The variation of documents exists cause a concern for technical staff to perform the digitization. An initial assessment provides information which allows technical staff to choose the right equipment need for digitization. This survey must be answer by the archivist which able to address the documents and the content value of those documents. The survey methods used by technical staff are by providing a complete listing of all documents holding in repositories to the archivists. Then, the archivist will enter the repositories and conduct the assessment to identify the documents that need to be pulling off for digitization project.

Collecting Data

After the initial assessment was conducted by the archivists, the technical staff need to collect all data from the survey and performed a review over the data. The data should consist the information about documents, forms and format, description and arrangement and other necessities data. An approval must be received by technical staff from top management or senior archivist to conduct the digitization over the documents selected.

Identify the document to be Scan

After receiving and approval from top management or senior staff, documents that are selected need to be collect and must be place together in the suitable environment or spaces before the document enter the scanning process. The documents need to be in safe and secure condition which protect documents from deterioration or worsen the documents condition.

Reviewing Document Format

The documents need to be sort according to the forms and format of their original document. The entire documents need to be label according to the forms and format of original documents. Every document need to have its unique identifier to arrange the documents according to their possible arrangement. Technical staffs need to identify the suitable equipment and method that will be used for every forms and format of original documents in scanning process.

Preparing Original Documents

All documents must be assessed according for their digitization appropriate prior to the digitization process of selected documents. This process will determine the extent of preparation required and ensure successful images are obtained from the digitization process. Considerations of the physical properties and attributes of the original documents will influence the image quality. Imaging Services Records and Information Management Service Alberta, (2013) explained that documents should be examined and if the poor quality of documents was found it should be notice by marking the document to alert the staff in scanning process. Technical staffs need to ensure when removing staples or other types of bindings, so that the original document is not damaged while scanning process. This institution also suggests that documents should be grouped together in batches to ensure

easier management and to be able to perform quality control and other procedures on a sampling basis. Document batches helps technical staff managing the process of preparing documents for scanning and ease scanning process.

2ND. STAGES: CURRENT

Sorting Document

Chmielewska (2013) stress that a rapid growth and success of presenting of electronic documents in the 90's and scientific environment to undertake building their own digital resources. Similar activities towards selecting documents for the collection, assumptions were made when planning the digital archives was that the core of the internet resources would be made from the documents from the public domain. But in addition, that the Polish copyright law does not protect normative acts or their official projects, official documents, materials, signs and symbols, published patent descriptions or simple press information.

Other than that, all of the documents and material for the digitization coming from the collections of the Warsaw University Library are been scanned within the library, some technical regulations according the digitization of the documents. Source files before the conversion into the viewing format could be modified and adjusted. For example, sharpening or color correction, if the condition of the object is fair enough, all of the documents published in e-bUW are also automatically processed with OCR.

Inspection and Separating Documents

Preparation of Records, if you just need a simple and quick way of just separating the documents in a batch, you can use a general separation sheet included with the software. Moreover, to be able to identify each document after the scanning, you specify a common filename for all the documents and a count number will be automatically appended during the automatic separation process. Other than that, another quick way of separating and naming documents in the batch is to use a yellow marker pen to highlight some text on the first page of each document. The highlighted text will be used as the filename and can be used to index it in the database afterwards.

Next, if the documents in the batch are always the same number of pages, you can just set a fixed interval to separate them. A common filename and auto-numbering is used to identify each document. Sometimes it is desirable or even necessary to let people do some Quality Assurance of the batch-scanned documents before they go into the database

If the standard for documents in your archive is strict or the material that you have to scan is of lesser quality, setting up a fully automatic process can be challenging. Some points that may be necessary to check by an operator is:

- Missing or misplaced pages
- Incorrect separation due to barcode errors
- Page order
- Proofing of readability

Rendering Document

Moreover, according to the Mike Jean (2010), there are a few things you can do to help us prepare your documents for scanning. These steps will also help to streamline the scanning process and allow us to input your documents faster. First, do not tape receipts on both sides of a single page. Do not fold and tape receipts so that they will fit on a blank piece of paper.

If you need to cut off unnecessary trailer information to adjust the receipt to fit, feel free to do so. Just make sure you are not trimming off important information. If you are uncertain, please contact the Records Management staff for guidance. When sending your documents over to services to be processed please use a large manila envelope and refrain from folding your document to fit in a standard mail envelope. Please take care of your receipts. Try to keep from folding or crunching receipts in your wallet or purse, and be sure not to leave them in direct sunlight or near heat as it will damage most receipt paper.

Integrating with software, it is possible to include a step of manual inspection and editing in the workflow. The scanned documents will automatically appear in PixEdit, where a QA person can browse through and correct any errors before they are saved and indexed. Or, if the errors are severe, he could decide the batch must be rescanned.

3RD. STAGES: POST DOCUMENT PREPARATION

After Scanning

Document that complete scanning process will be placed at suitable environment or places. These documents will be inspected for the condition of physical and reorganize into proper order. Documents are arranged and identify the documents reference number. Records Management Section, Government of Northwest Territories (2010) indicates that maintain indices of storage location to allow users to retrieve records when required. Use audit trails generated from the document logs or file access logs to track when original records are removed and returned from the storage location.

Rearrangement of Document into Repositories

The document transport into repositories and need to be arranged into the repositories according to their classification number. After that, the documents need to be arranged according to their arrangement order on shelves. The archivist need to be approved by signing a form as evidence document are being sent after the scanning process finished.

Importance of Document Preparation

Document preparation is an important process before doing the digitization process. It is a process to identify the type of documents that need to be digitized and the reason of digitizing them. Doing document preparation have some benefits for the record holders.

Only Valuable Documents Digitize

One of the benefits of document preparation is to make sure valuable documents digitize. Valuable documents need to be digitizing because it has vital information that cannot be lost. If these types of documents are lost, a certain organization or country might be in a loss because they have lost the most vital documents that can help their organization or country. This does not mean that the original record would be destroyed; it is just a way of making extra copy.

Ease the Digitization Process

When scanning documents for digitization, documents' pages need to be scanned one by one. Therefore, there are some documents that have bounds on them such as staples. Those things need to be removed first in order that the document is easier to be scanned. By doing document preparation, digitizer will be easier to do their job and will reduce time consumed to digitize.

Preserve Damage Documents from Loss

Documents that are damaged may have values too, so it is important to digitize them to make sure it is not damage even more when it is use by people. The more people are handling the document, the more easily it gets to be damage, so it is best to be digitized.

Identify Documents that Have High Demand

It is important that the documents that have a very high demand of access to be digitize. This is because it is used by so many people and it will easily get damage. So, it is better if the documents were digitized and it also allows more than one people to access the documents by digitizing. These benefits are very good for any organization to do a digitization. It is advisable that every organization that holds not only vital records but also any kinds of records to digitize. In addition, it will reduce the need of space and workloads to care for the documents.

Problem and Challenges

Non-Standard Paper Sizes

The document in batches needs to be in the same size and orientation because they can be put in place in the scanning feeder. If there is a different paper sized, document will be detached and scan individually using flatbed scanner.

Non-Standard Text Orientation

Standard text is oriented so that the text runs parallel to the short sides of the piece of paper. If text is oriented so that it runs parallel to the long sides of the paper it may be necessary to rotate the images after they have been scanned. It may be more difficult to run OCR (optical character recognition) software on these documents.

Thickness of Paper

Extremely heavy or light paper may not feed properly through an automatic document feeder. These types of papers should either be photocopied, and a scan made of the photocopy or they should be scanned by placing each page individually on a flatbed scanner. The different thickness of paper such as 70 GSM and 80 GSM influence the scanning method.

Smoothness, Texture and Finish

Three problems can be caused by rough paper. The first is that the texture of the paper causes the image to appear broken. The second is that the texture of the paper may be picked up by the scanner and reproduced in the image. The third is that rough papers may cause the document scanning feeder mechanism to jam.

Colored Paper

When scanning in black and white, colored papers may appear as black to the scanner. Scanner settings may need to be adjusted so that the scanner can distinguish the text from the background colour. Run a test scan of colored documents before scanning a batch.

Voluminous Documents

The enormous of documents in the repositories require a long period for archivist to assess every documents condition and content value. The long period taken for examining the documents has become a huge challenge for technical staff to prepare the documents.

Need to Meet Document Security Mandates on Security and Access

Staff technical needs to overcome security feature that secure the physical document. The documents security needs to be release in order for technical staff able to transport the documents outside the repositories. The access to the documents needs to be given to the technical staff for them to access the documents content.

Involve a Large Number of Manpower

A lot of staff needed to perform the assessment over the documents in the repositories. They also required preparing the document that needed for scanning process. The manpower could cause a huge budget for digitization projects as they are important to keep the process goes as schedule.

Material Condition and Format

The different type of documents format needed different equipment for scanning process. The scanning process need to have an enough equipment to support the digitization process. Otherwise, there are certain documents that can only be supported for scanning process. The material condition also can affect the quality of scan picture. The material need to be in the best condition for the document to be scanned.

Recommendation

As a recommendation, on the successful completion of this training the staff will be able to integrate the document of the document control of one or more management skill. For example, tackle the quality of the staff that provides staff knowledge's and safety aspect the training should be related on how to provide a document control and recording system and the management can measure the strength and the weakness of the staff, also quality of the facility of digitization is also important to be maintain efficiently so that the organization have good quality of digitizing the document of the agencies.

Moreover, instead of having to find and manage local storage for all your scanned documents and then deal with online archives, you can send scans to your online storage system of choice straight from the scanner. For example, the Fujitsu ScanSnap can send documents to SugarSync, Evernote, Google Drive, and other online locations, Dropbox, Evernote, and Google Drive. You'll need to spend some time configuring things, but once set up you'll find you can cut out the middleman effort of manually copying files to the cloud for backup.

Other similar services are focused exclusively on cloud storage and include document scanning as a feature. Check out [SimpleScan](#) and [Finovera](#), which are designed to wean you from reliance on local storage altogether by putting everything online, though some may have some initial concern over trusting their entire financial lives to a cloud service.

Conclusion

Document preparation is an importance process that requires a through selection and evaluation process of material that have unique and rare information and value to be digitize. Document preparation process is a preliminary stage that needs to be conduct by organization in digitization project. Preparing document require a meticulous planning and research before it was conducted by organization. The criteria need to be elaborate and assess on the material that has information and bring value to organization. The criteria must be assessed on the material which is listed to be digitized. This is to avoid material that is not important and invaluable material are digitize in the project. Once this have been clear, a firm

decision need to be perform whether or not to digitize the material and whether the digitization department have a sufficient resource and facilities in order to conduct the digitization on the material decided. Digitization project requires a huge sum of budget to execute the project; hence there is a need for proper planning over the project. As noted by Jervis, M. & Masoodian, M. (2014) it is clear that there is a need for systems that provide facilities for a more integrated management of paper and electronic documents. Various tools and techniques have been developed to assist people in dealing with problems associated with information fragmentation over different devices and systems.

People want to preserve the past that have a value to the individual, an organization and to the country. Today people or Internet users, can access the content directly from the library's home page or archives' home page, also have other possibilities of reaching the e-collection and historical document via various access point, domestic or foreign. For the record, there is a lot of challenges regarding digitizing document due to lack of staff and facility, but still Malaysia always try to make a softcopy of the document, so that we can preserve in a various technologies and standards, gives a leading model for the digitization projects not only in Malaysia but for the developing countries worldwide. Throughout this procedure, it is proved to be a configurable and flexible management open source tool for any digitization workflow. Hence, the preparation of material is a stage that need to be focus on by archivist and record manager to ensure the process of digitization execute effortlessly and efficient. The organization need to produce a guidelines or recommended practices for document preparation process and the implementation of such practices helps staff in preparing and evaluating the material properly.

References

- Chmielewska, B. & Agnieszka, W. (2013) "Providing access to historical documents through digitization", *Library Management*, Vol. 34 (4/5), pp.324-334, <https://doi.org/10.1108/01435121311328663>
- International Federation of Library Associations and Institutions (2002). Guidelines for digitization projects: for collections and holdings in the public domain, particularly those held by libraries and archives. *Unpublished manuscript*. Retrieved from <https://www.ifla.org/files/assets/preservation-and-conservation/publications/digitization-projects-guidelines.pdf>
- Jervis, M. & Masoodian, M. (2014). How do people attempt to integrate the management of their paper and electronic documents? *Aslib Journal of Information Management*. Vol. 66 (2), 134-155. Retrieved from <http://dx.doi.org/10.1108/AJIM-01-2013-0007>
- National Archives. (n.d.). *Preparation of records*. Retrieved Dec 21, 2107, from <http://www.nationalarchives.gov.uk/documents/information-management/preparation-of-records.pdf>
- Liu, Z. (2004). The evolution of documents and its impacts. *Journal of Documentation*. Vol.60 (3). pp. 279 – 288. doi: <http://dx.doi.org/10.1108/00220410410534185>
- Medicaid Information Technology System (2010). *Electronic Document Management System (EDMS): Document Preparation, Scanning and Manual IQC*. Retrieved December 17, 2017, from http://medicaid.ohio.gov/Portals/0/Providers/MITS/MITS%20Non%20Provider%20Communications/EDMS_DocPrep_Scan_IQC_Updated.pdf

- Records Management Section, Government of Northwest Territories (2010). *Guidelines for Scanning Projects*. Retrieved from https://www.inf.gov.nt.ca/sites/inf/files/procedures-guidelines_for_scanning_projects.pdf
- Imaging Services Records and Information Management Service Alberta. (2013). *Digitization Document Preparation: recommended Practice*. Retrieved from http://imtdocs.alberta.ca/RIM0002_Digitization_Document_Preparation_February2013.pdf
- Document Sorting Software. (n.d.). Retrieved December 17, 2017, from https://www.scanstore.com/Document_Sorting_Software/
- Null, C. (2013, December 30). How to set up a document scanning system. Retrieved December 17, 2017, from <https://www.pcworld.com/article/2082569/how-to-set-up-a-document-scanning-system.html>
- (n.d.). Retrieved December 17, 2017, from <https://www.pixedit.com/pixedit-software-suite/pixedit-desktop/batch-scanning/separation-and-indexing/>