



# THE POTENTIAL USE OF ARTIFICIAL INTELLIGENCE TECHNOLOGY IN THE PROCESS OF COLLECTING METADATA IN PHOTO ARCHIVE DESCRIPTION ACTIVITIES

Wiwiet Mardiaty<sup>1\*</sup>, Siti Aisyah<sup>1</sup>, Niko Grataridarga<sup>1</sup>, Retno Wulandari<sup>1</sup>

<sup>1</sup>Applied Social Science, Vocational Education Program, Universitas Indonesia

wiwiet.m@ui.ac.id

**Abstract.** In the current era of digital technology, the creation of photos produced in an activity is experiencing a significant amount of "explosion". These photos are often stored without adequate metadata, allowing a great deal of information and events to fade over time. In this case, information managers and archivists have a big task dealing with still image archives that must be identified and analysed immediately. Currently there has been research on how Artificial Intelligence (AI) can be used to help collect metadata and descriptions of still image archives. This paper will describe how still image archive identification activities are currently carried out at the Indonesian National Archives manually and provide an overview of how this activity can be assisted with the use of AI technology. It is hoped that the results obtained from this paper can serve as a basis for further research regarding the use of AI for archival activities in general, especially for Indonesian audio visual and still image archives such as photographs and other works of art, and can be a solution for creating more detailed archival descriptions and can later help improve findability in still image archives searches.

Keywords: artificial intelligence, metadata, photo archive

## 1 Introduction

In the current era of digital technology, the creation of photos produced in an activity is experiencing a significant "explosion". These photos are often stored without adequate metadata, so much information and events fade over time. In this case, information managers and archivists have the big task of handling photo archives that must be identified and analysed immediately. In 2021, internship activities were carried out at the Indonesian national archives by doing a description of photo archives which are part of the stages of processing their archives collections. Meanwhile, the archives described are photo archives of the Ministry of Information for the Jakarta area in 1955. The activity was carried out by apprentices with 12 archivists who produced 538 photo archives in a period of approximately 3 months.

There is currently research into how Artificial Intelligence (AI) can be used to help collect metadata and archive photo descriptions. Extracting metadata from non-digital photos using AI can be more challenging than with modern digital images because non-digital photos don't have any embedded metadata. However, AI techniques can still be applied to analyse visual content and conclude some metadata, so it is hoped that the

© The Author(s) 2023

D. V. Ferezagia et al. (eds.), *Proceedings of the International Conference on Vocational Education Applied Science and Technology (ICVEAST 2023)*, Advances in Social Science, Education and Humanities Research 783,

[https://doi.org/10.2991/978-2-38476-132-6\\_78](https://doi.org/10.2991/978-2-38476-132-6_78)

implementation of photo archive descriptions in the future can be done more quickly with a smaller team of archivists.

This paper will explain how the identification of photo archives is currently being carried out at the Indonesian National Archives manually and provide an overview of how this activity can be assisted using AI technology. It is hoped that the results obtained from this paper will become the basis for further research regarding the use of AI for archival activities in general, especially for audio-visual archives and Indonesian still images such as photographs and other works of art, as well as solutions for creating more detailed archive descriptions which will be available later. helps improve searchability in photo archive searches.

## 2 Literature Review

Hadiwardoyo (2002) [9] defines that a state archive is an archive must be stored and managed by an archival institution because it has value for national accountability. archives are produced by archive creators because they have historical value, have expired retention, and have permanent statements that have been verified either directly or indirectly by the National Archives of the Republic of Indonesia and/or archival institutions. Archive of secondary use value or have permanent use value managed by archival institutions as a result of systematic and selective acquisition of archive treasures created in the implementation of the activities of the agency that created them [4].

According to Anthology of Archives (Musliichah, 2016) [8] archives created as a by-product of administrative activities are indispensable as a resource in organizational management, archives have various roles/functions, including: for evidence of activity, for corporate Memory, for planning process, for decision-making process, and for supervision activity.

The processing of archives is the second stage in managing archives according to Constitution regulation number 43 of 2009. PERKA ANRI Number 27 of 2011 concerning Guidelines for the Preparation of Archive Retrieval Assistance Facility explained Processing of archives is the process of making a means of assisting the retrieval of archives based on applicable archival principles. Article 62 paragraph 2 of Law Number 43 of 2009 concerning Archives explains that processing of archives can be carried out based on standard archive descriptions. Types of archive retrieval aids produced in the processing of archives at archival institutions consist of archive lists, archival inventories, and archive guides. This is done for the accessibility of archives to users to make it easier to find the archive.

Archive processing activities can be carried out based on provenance and provenanceal order. These two principles must be applied in processing activities which can later be used as a reference in archive processing activities. Based on Government Regulation Number 28 of 2012 Article 96 explains that the processing of archives is carried out based on provenance and original order, as well as standard descriptions of archives. In the Regulation of the Head of ANRI Number 27 of 2011 concerning Guidelines for the

Preparation of Facilities to Assist the Retrieval of Archives, it is explained that the principle of processing archives cannot be separated from the context of the creator of the archive and its structuring system, so in its preparation it is necessary to pay attention to provenance, namely the principle/principle that is carried out to maintain permanent archives. managed in a single archive creator unit (Provenance), not mixed with archives originating from other archive creators, so that archives can be attached to the context of their creation. The principle of the original order is the principle that is carried out to keep the archives organized in accordance with the original arrangement (original order) or in accordance with the arrangements when the archives were still used for carrying out the activities of the archive creator. This means that archival institution archivists in recording and storing archives must pay attention to the work unit from which the archive creator originates and the subject matter and details of the archive. In this way it can guarantee systematics, control, and easy access to archives.

Based on the Head of PERKA ANRI Number 27 of 2011 concerning Guidelines for the Preparation of Archive Retrieval Assistance Facilities, it is explained that the procedure for processing archives in the framework of preparing a archive recovery aid in archival inventory is carried out through the following work stages: (1) identification of archives; (2) Preparation of technical plans; (3) Carry out tracing of data sources; (3) Compilation of a temporary scheme for managing archives; (4) Reconstruction of archives; (5) Archive description; (6) Compilation of a definitive scheme for archival arrangements; (7) Maneuver/unification of information and physical archives; (8) Definitive numbering; (9) Labeling of archives and arrangement in archive boxes; (10) Box labeling and box arrangement; (11) Writing a draft archive inventory; (12) Assessment and sampling; (13) Improvement of the results of the assessment and sampling; (14) Approval of the archive list.

In taking pictures, there are various camera viewpoints (Angle), including Normal angle, Wide angle, High angle, Low angle, Bird angle, Frog angle, Extreme long shot, Long shot, Medium long shot, Medium Shot, Medium close up, Close up, Big close up, and Extreme close up. This shooting was written to further clarify the description of the point of view in taking a photo.

**Photo Archive.** In PERKA ANRI Number 23 of 2011 concerning Guidelines for the Preservation of Archives it is explained that photo archives are archives whose information content is in the form of static images (still images), the creation of which uses special equipment. Photo archives are one of the audiovisual archives, which means that the contents of the information can be seen and/or heard. Based on the book *Archival Anthology* (Musliichah, 2016) Photo archives whose information content is in the form of still images which are created using special equipment. Audiovisual Archives of still images in the form of photographs, drawings, and others. still images archives are archives whose information content is in the form of still visual/images or not moving, included in this category are archives of photographs, transparent microfilm, pictures and posters.

**Archive Description.** According to PERKA ANRI Number 14 of 2018 concerning the Standard Archive Description, it is explained that an archive description is the creation of an accurate representation of an archival unit (an archive or a number of archives) in

full and all its components by capturing, analysing, organizing and recording information that is important for identifying, manage, find, describe archives along with the context and archive system that creates them. The purpose of the description is to identify and explain the context and content of the archive in order to increase the accessibility of the archive. Description is the preparation of means of retrieving archives to facilitate the control and use of archives according to the agency/organization so that they can be used for public services (Sauki Hadiwardoyo, 2002). Archive description is the preparation of an accurate description of an archive unit which is described in full and all its components [11]. The word description means exposure or description in clear and detailed words (Indonesian General Dictionary, 2008). According to Saur, K.G. Dictionary of Archival Terminology Description is a preparation for creating tools to facilitate the control and consultation of archives.

Physical description is the process of creating an archive retrieval tool that can help archive users access their archives. In the management of archives, the physical description of the archive includes the activity of providing information related to the physical characteristics of the archive, content and information on the contents of the archive. Archive description standards are the rules used in describing information or details of information contained in archives.

Archive descriptions can be done in stages starting from the macro, medium, and micro levels. Based on PERKA ANRI No. 14 of 2018 Article 2 paragraph 4 describes the description of archives to be carried out in stages with the provisions: Starting from general to specific information; Contains information that is relevant to the level of description being worked on; Do not repeat information; and provides an indication of the linkages between levels of description. In addition to the provisions in the archive description, there is a discussion regarding the depth of the description which must also be considered. The PERKA ANRI No. 14 of 2018 explains that the depth of description is a representation of the level of detail of an archive that will be described at the description level. Matters discussed further in the depth of the description are archival materials, archive processing, archive retrieval tools and rules or conventions used. Two basic things are needed in a description, namely each description must be accurate / precise / thorough (representing archives) and descriptions of structured data so that it facilitates data processing [17].

**Archive Description Elements.** Audio Visual description data elements consist of two, namely intellectual data and technical data. For the photo archive description data elements, namely intellectual data consisting of photo numbers (positive and negative), activity titles, information descriptions, time, place, character names, photographer names, publications, and descriptions. While the technical data for the description of photo archives consists of photo size (3R, 4R, 5R), photo color, photo format (portrait/landscape), type of photo (original/duplicate/repro), date of reproduction, type of photo paper (doff/glossy) , other formats: repro/copy (repro/copy number) and photo negatives (negative number). Five categories minimum in the description: Title, date, amount, physical description, and summary of archival information. (Bradsher. *Managing Archives and Archival Institution*, 1988).

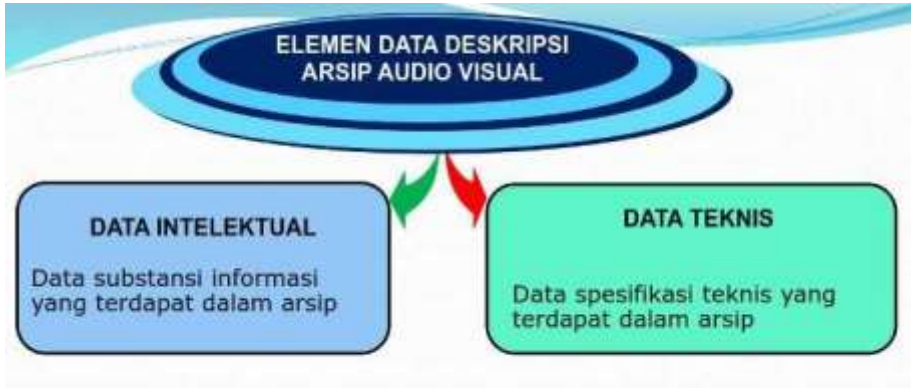


Figure 2. Audio Visual Description Data Element

Based on PERKA ANRI No 14 of 2018 Article 8 paragraphs 1 and 2 it is explained that general elements are elements that regulate matters that are not directly related to Archive material. Common elements consist of the following; access point or index; punctuation in descriptions; writing format; and language.

Whereas mandatory elements contain description elements that regulate matters directly related to photo archive material and are elements that must at a minimum be present for the benefit of exchanging archival information, namely: Unique Code; Title; Archive Period; Physical description; Level Description; Contributor; and Scope and Content.

This additional element completes the mandatory elements and if the source of information for the additional elements cannot be found, this element does not need to be filled in, such as: Processing Period; Archivist Records; Systematic Writing of Archive Retrieval Assistance; Notes; Archive Creator History; Archive History; Acquisition History; Archive Retention Schedule Information; Existence and Location of Original Archives; Existence and Location of Copy Archives; Publication Notes; Instructions for Using Archives

### **ARTIFICIAL INTELLIGENCE (AI)**

Artificial Intelligence (AI) refers to the simulation of human intelligence in computer systems, enabling them to perform tasks that would normally require human intelligence. These tasks may include learning, reasoning, problem solving, perception, understanding natural language, and decision making. AI aims to create machines that can mimic cognitive functions and exhibit "intelligent" behavior. (OpenAI, 2023, July 19).

AI systems are designed to analyse and process large amounts of data, identify patterns and make predictions or decisions based on that data. There are various approaches to AI, including: Narrow or Weak AI: This type of AI is designed to perform a specific

task or solve a specific problem. Examples include virtual personal assistants like Siri or Alexa, recommendation systems, and AI-based chatbots. **General or Strong AI:** This is a more advanced form of AI that has human-like intelligence and the ability to understand, learn, and reason across a variety of tasks similar to human abilities. This level of AI is still theoretical and has yet to be reached. **Machine Learning:** A subset of AI, machine learning involves training algorithms on data to learn and improve performance without explicit programming. This allows the AI system to adapt and improve based on new information. **Deep Learning:** A specific form of machine learning, deep learning uses neural networks to process and learn from large data sets. This has enabled significant advances in tasks such as image and speech recognition.

AI can be used in the world of archives to improve the efficiency, accuracy and accessibility of information in managing records and documents. Here are some of the ways in which AI is applied in the world of archives:

**Automatic Indexing:** AI can be used to perform automatic indexing of documents that enter the archive system. Through natural language processing, AI can identify and extract important information from documents, such as name, date, reference number, and topic, to make document search and classification easier.

**Fast and Accurate Search:** Using natural language processing and text analysis technologies, AI-powered filing systems can provide more accurate and relevant document searches. AI understands the context and meaning behind search queries, producing more relevant results than traditional keyword-based searches.

**Document Classification:** AI can help classify documents into proper categories based on their content. The use of machine learning algorithms allows AI to identify different patterns and attributes in documents, so that it can accurately classify documents according to predefined categories.

**Removal of Sensitive Information:** In the context of archives containing personal or sensitive information, AI can be used to detect and remove sensitive information from documents automatically. This helps maintain data privacy and security.

**Duplicate Document Detection:** Using text analysis and pattern recognition techniques, AI can help identify identical or duplicate documents in filing systems. This helps reduce redundancy and optimizes storage space usage.

**Sentiment Analysis and Classification of Emotions:** AI can be applied to analyze sentiments or emotions in documents, for example in letters or emails. This can help understand the responses or feelings associated with the content of a particular document.

**Handwriting and Optical Character Recognition:** AI-powered handwriting and optical character recognition (OCR) technology can help convert physical documents into digital formats, enabling them to be more accessible and searchable.

By implementing AI technology in archives, companies and organizations can increase efficiency, reduce information search time, avoid human errors, and improve the security and accuracy of document management.

## **AI FOR PHOTO ARCHIVE DESCRIPTION**

Artificial intelligence (AI) technology can be very useful for the process of collecting metadata in photo archive description activities. Here are some potential uses of AI in this context:

First, Automatic Metadata Extraction. AI can be used to automatically extract metadata from photos, such as date, location, object, scene and other relevant information. Deep learning models can be trained on large data sets of annotated photos to recognize and extract specific metadata elements. This automated extraction can speed up the process significantly and reduce manual effort.

Second, Image Recognition and Object Marking. AI models can analyze the visual content of photos and identify specific objects, people, landmarks or attributes in the image. By automatically tagging photos with object labels, it's easier to find and categorize images based on their content.

Third, Optical Character Recognition (OCR): AI-powered OCR algorithms can extract text from photos, including handwritten or printed text in old documents or captions. The extracted text can provide valuable metadata, such as the name, date, description or any textual information associated with the image.

Fourth, Image Similarity and Duplicate Detection. AI techniques can be used to identify similar or duplicate photos in large photo archives. By detecting duplicates, it becomes easier to organize and delete duplicate archives, ensures efficient storage usage and facilitates search and retrieval.

Fifth, Image Analysis and Classification: AI models can analyze the visual characteristics of photos and classify them based on various attributes such as style, genre, theme or historical period. This classification can help in creating structured metadata, enabling better organization and retrieval of images in archives.

Sixth, Enrichment and Data Linking. AI can help enrich photo metadata by connecting with external databases, such as geographic or historical repositories, to provide additional context. By leveraging AI-powered techniques, you can automatically collect relevant information about the location, event or person associated with a photo, enhancing its descriptive metadata.

Seventh, Natural Language Processing (NLP): NLP techniques can be used to analyze text related to photo archives, such as captions, annotations, or historical documents. AI models can extract information, entities or sentiments from textual data, contributing to the metadata collection process. AI technology offers the potential to speed up and improve the process of collecting metadata in photo archive description activities. They

can automate repetitive tasks, improve accuracy, and provide valuable insights that help efficiently organize, search, and store photo collections. (OpenAI. (2023, July 19)

## **AI FOR NON-DIGITAL PHOTO ARCHIVE DESCRIPTION**

Extracting metadata from old photos using artificial intelligence can be more challenging than with modern digital images because old photos may lack embedded metadata or EXIF data. However, AI techniques can still be applied to analyse visual content and infer some metadata. Here are some approaches:

First, Image Classification. Train a custom deep learning model using a dataset of old, labeled photos. This model can classify images based on various attributes such as time period, location, object or scene. Create a labeled dataset by manually annotating a set of old photos with relevant metadata. This can involve identifying objects, clothing styles, architectural features, or other visual cues that can provide insight into the context of a photo. Use trained models to predict metadata for old, unseen photos.

Second, Generative Adversarial Networks (GAN): Take advantage of GANs, which can generate new images based on certain categories or styles, to simulate old photo styles. Train the GAN model on old photo datasets to learn the characteristics of vintage or historic images. Apply a trained GAN model to generate new images that resemble old photos and then analyze the resulting images to extract the metadata.

Third, Caption: Use the caption model, which combines computer vision and natural language processing, to generate descriptions or captions for old photos. Train models on old photo datasets paired with appropriate captions or descriptions. Apply a trained model to generate textual metadata for invisible old photos. Crowdsourcing and Human Input: Consider leveraging crowdsourcing platforms or engaging human experts to manually analyze and annotate old photos. Crowd workers or domain experts can review and provide metadata based on their knowledge of the historical context, visual cues, and cultural references depicted in images.

It's important to note that extracting metadata from old photos using AI methods may not be as accurate or reliable as modern digital images. The success and quality of the results will depend on the availability and quality of the training data, the complexity of the tasks and the capabilities of the AI models used. Additionally, combining AI approaches with human expertise can help increase the accuracy and depth of extracted metadata. OpenAI. (2023, July 10)

## **3 DISCUSSION**

### **3.1. Photo archive description activities in the national archives**

In 2021, internship activities will be carried out at the Indonesian national archives by doing a description of photo archives which are part of the stages of processing archives. Meanwhile, the archives described are photo archives of the Ministry of Information for



the Jakarta area in 1955. The activity was carried out by apprentices with 12 archivists who produced 538 photo archives in a period of approximately 3 months.

The parties involved in carrying out the activity of describing the photo archives of the Ministry of Information for the Jakarta area in 1955, namely the first archivist, youth, and supervisors consisted of 12 people. The archive description standard refers to PERKA ANRINumber 14 of 2018 concerning Archive Description Standards. Describing archives, especially photo archives, there are elements that must be considered in carrying out description activities, namely general elements related to access points or indexes, punctuation, language, and writing formats. The mandatory elements consist of a unique code, title, archive period, physical description, level of description, contributors, scope and content while additional elements are used to complete the information in compiling a tool for retrieving archives, in this case the inventory of photo archives of the Ministry of Information for the Jakarta area in 2010. 1955.

The activity of describing photo archives of the Ministry of Information for the Jakarta area in 1955 was carried out for 86 working days according to SOP AP Number 63 of 2015 concerning Description Archive descriptions that have been made. Photo archive description activities were carried out by a total of 12 archivists including those led by the coordinator and secretary. The division of the number of description quotas is divided into 12 according to the number of members, each person gets approximately 538 archives to be described. Each day has a target of describing at least 15 archives according to the level of difficulty or convenience of the archives.

Table 1. Number of Ministry of Information photo archives 1955

<b>Total: 6.854 Archive</b>	6452 Master
	402 Duplicate
<b>Month</b>	<b>Amount</b>
April	51 sheets
May	413 sheets
June	74 sheets
<b>Total</b>	<b>538 sheets</b>

The procedure for the activity of describing photo archives of the Ministry of Information for the Jakarta area in 1955 had several stages so as to create a good quality photo archive description. The procedure for this description activity is regulated in the Standard Operating Procedures for Government Administration (SOP AP) Number 63 of 2015 concerning Archive Descriptions. However, along with the implementation in the field, there are additions or updates that must be submitted to the AP SOP to clarify the stages. So that this AP SOP can be used as fully as possible.

Of the several stages of implementing archive descriptions, there are two activities that require further investigation, namely Archive Identification and Data Source Searching.

In carrying out the identification of archives, the identification method is carried out using the archive identification form in accordance with the AP SOP for Compiling a List of Archives that are available by looking directly at the physical archives at the Archive Storage Depot. Identification is done by knowing the creator of the archive, type of archive, period, amount/volume, area, negative numbering model, duplication, condition of the archive, and storage location of the caption on the back of the photo archive. Overall, the physical photo archives of the Ministry of Information of the Republic of Indonesia for the 1955 Jakarta Region are in good condition and most (90%) are identified, so they are very helpful to the team when they are going to do a description.

In carrying out data source searching activities, it is carried out to obtain complete information relating to the context of the creation of photo archives of the Indonesian Ministry of Information such as the history of the institution, the regulations that form the basis of the institution's duties and functions, or other information relating to the archives being processed. In PERKA ANRI Number 14 of 2018 concerning the Standard for Archival Descriptions, it is explained that context is the organizational, functional and operational circumstances that surround the creation of material, signs of receipt, storage or use, and its relationship to other materials. Searching for data sources was carried out at several agencies related to the Ministry of Information, namely the Depok City Library, the National Library of the Republic of Indonesia, the National Press Monument, and the Information Museum.

In tracing sources at the Depok City Library, the first day the team received data and information was as follows: The book entitled "President of the Republic of Indonesia" was published by the Ministry of Education and Culture of the Republic of Indonesia; On the second day of tracing sources, the team obtained data and information as follows: The book entitled "Evaluation of the 1955-2019 Indonesian Elections A Perspective of Legal Pluralism" was published by PT. Rayyana Komikndo (2021) by M. Nizar Kherid, SH, MH.

Meanwhile, the results of searching data sources and references at the National Library of the Republic of Indonesia are: (1) Searching at the National Library of Indonesia's Photo, Map and Painting Collection Services found several photographs of figures that may be related to the photo archives being processed. (2) In the search on the Audiovisual Service, the National Library of Indonesia borrowed several old newspapers in microfilm form, including: National Public Daily 1955, Indonesia Raya, from the People by the People, and People's Daily 1954-1955. From some of these newspapers found things related to interesting figures or events in 1955. (3) In a search on the Open Monograph Service, found several books containing and relating to figures from the Indonesian Government in 1955, among others entitled History of Indonesia from the Proclamation to the Reformation Order by Ketut Sedana Arta, S.Pd., M.Pd. and Dr. I Ketut Margi, M.Sc. Publisher Graha Ilmu, Cabinet Development and Government of Indonesia author Bibit Suprpto, Jakarta 1950-1970 author Firman Lubis, Caring for Indonesia Looking for Figures and Events author Lukman Hakiem and Records of My Nation's Historical Events 1950-1975 Writer Mashuri published by

Intan Pariwara.

Source tracing at the National Press Monument produced newspapers that could be accessed, namely the Dwi Warna daily in July 1955, the Indonesian Suluh daily from April to December 1955. Some data regarding events were captured directly from the computer monitor in the access room, including regarding friendly meetings said General A.H. Nasution, Colonel Gatot Subroto, and PWI Chairman Sutan Sjahrir in the Army General Staff Hall; Departure of the Indonesian delegation chaired by the Minister of Foreign Affairs, Anak Agung Gde Agung, to the Netherlands; Jungschlager Trial; and Governors' Conferences throughout Indonesia and so on.

The results of tracing data sources from the Information Museum are in the form of books as follows: (1) Books on the History of the Asian-African Conference from the Ministry of Foreign Affairs of the Republic of Indonesia; (2) Books on Election Matters, Author Sigit Pamungkas. (3) Books on Visualization of New Order Development Results Pelita I-Pelita II-Pelita III.

Searching for data sources is the stage of processing archives that must be carried out to complete information that is not available in the archives. This supports the quality of the description during the description activity to find out the context of what happened in 1955. The information that has been collected will make it easier for archivists to describe when the captions on the archive are incomplete.

From the two activities above, it can be concluded that in the archive description activity, of course, there are several things that make the description of photo archives relatively easy and difficult, with an explanation as shown in Figure 2 below:

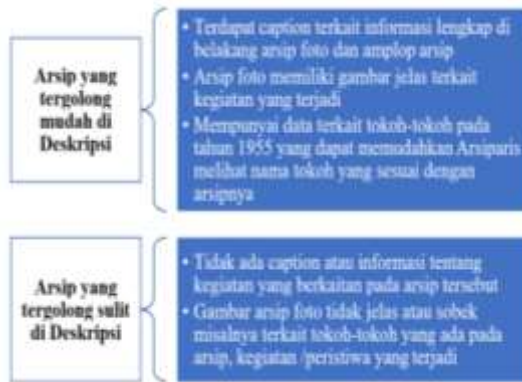


Figure 2. Archive descriptions that are classified as easy and difficult

### 3.2 Potential use of AI for Photo Archives Description

The activity above shows that old non-digital photos require significant time just to get in-depth information about the activities captured in an image. With the use of AI Technology, each investigation result can be used to train the AI application and later

on can assist to describe other photos, so it is hoped that processing time can be faster with more accurate results. Of course, old photos can pose unique challenges, such as image degradation, inconsistent metadata or cultural context so it is vital for archivists to collaborate with experts in the field, such as historians, to ensure accurate analysis and interpretation of old photos. using AI techniques. However, this use of AI needs to begin as soon as now.

In personal communication with ChatGPT (OpenAI, July 10, 2023), the use of AI for the non-digital photo collection above, can be done as follows:

(1) Digitizing photos: the initial step is to physically convert non-digital photos to digital format by scanning or photographing them, ensuring the digital copies are of good quality and resolution.

(2) Preprocess data: Clean and organize digital photos by removing duplicates, correcting orientation, and cropping or resizing images if necessary. Make sure images are properly labeled with relevant metadata, such as date, location, and any known information about the subject. This is of course easy enough for the 1955 photo collection, which is mostly (90%) complete with individual captions.

(3) Curation of representative data sets: Those 1955 photos can be aggregated and grouped for those representing different time periods, locations, and subjects to begin using to train an AI model to recognize and understand the characteristics of non-digital photos, so that they can continued in the next collection of photos.

(4) Define objectives: Specific tasks to be completed by AI in extracting metadata or information contained in images can be adapted to the current ANRI Code, such as classifying photos by era, detecting objects, detecting figures/faces, detecting famous buildings, extract text from text, or generate text automatically.

(5) Choose the appropriate AI technique: with the existence of a variety of AI techniques, of course, it is necessary to study further which AI techniques are suitable for analysing photo archive collections. The analysis performed can include image recognition, object detection, OCR, natural language processing, or even generative models to recreate missing parts of damaged photos. Thus, choosing a technique that aligns with your goals and available resources becomes very important.

(6) Using a trained AI model: in an ideal process, of course it is good if the archivist has got a suitable AI model that has been selected and trained according to the purpose. The problem is, If pre-trained models are not available or insufficient, Archivists may need to train their own models using transfer learning or from scratch. This requires labeled data and the use of deep learning frameworks.

(7) Train and refine models: Use curated old photo datasets to train AI models. Adjust the model architecture, hyperparameters and training techniques to achieve optimal results. Refine the model on old photo datasets to improve performance and adapt them to the unique characteristics of existing photo archive collections.

(8) Evaluate the model: Assess the performance of a model trained using an evaluation metric, such as accuracy, precision, recall, or F1 score. Model validation using discrete test sets or by manually verifying its output against known metadata.

(9) Implement AI systems: Integrate trained AI models into the photo archive collection infrastructure. This may involve developing custom software, leveraging APIs, or leveraging an AI platform with deployment capabilities. Ensure that the system can efficiently process a large number of old photos and provide the desired output.

(10) Continuously improve and refine: Regularly evaluate AI system performance and collect feedback from users. Combining user feedback and new data to refine and improve the model over time. Stay updated with advances in research and AI technology to enhance the capabilities of photo archive collection systems.

## 4 Conclusion

Currently, photo archive description activities are still carried out manually and require a large number of implementing teams and take quite a long time to do. AI is here to facilitate the process of processing photo archives, it's time to start using it for photo archive description activities.

## Reference

1. Arsip Nasional Republik Indonesia. (2018). Peraturan Arsip Nasional Republik Indonesia Nomor 14 Tahun 2018 Standar Deskripsi Arsip Statis. Diakses Maret 15, 2022, dari [www.Jdih.anri.go.id](http://www.Jdih.anri.go.id)
2. Republik Indonesia. (2009). Undang-Undang Nomor 43 Tahun 2009 tentang Kearsipan. Diakses Maret 20, 2022, dari [www.Jdih.anri.go.id](http://www.Jdih.anri.go.id)
3. Republik Indonesia. (2012). Peraturan Pemerintah Republik Indonesia Nomor 28 Tahun 2012 tentang Pelaksanaan Undang-Undang Nomor 43 Tahun 2009 tentang Kearsipan. Dipetik April 4, 2022, dari <https://jdih.kemenkeu.go.id>
4. Azmi. (2016). Deskripsi dan Penataan Arsip Statis. Tangerang Selatan: Universitas Terbuka.
5. Apriyani, E., Safira, D., & Rodin, R. (2020). Pengelolaan arsip statis di Dinas Kearsipan Daerah Provinsi Sumatera Selatan. *Al-Kuttab: Jurnal Kajian Perpustakaan, Informasi dan Kearsipan*, 2(1), 1-11.
6. Bell, G. (2004). Automatic image content extraction: A framework for machine learning and image retrieval in the humanities and social sciences [PDF file]. Retrieved from [https://www.ideals.illinois.edu/bitstream/handle/2142/1507/LibraryTrendsv52i2f\\_opt.pdf?sequence=1&isAllowed=y](https://www.ideals.illinois.edu/bitstream/handle/2142/1507/LibraryTrendsv52i2f_opt.pdf?sequence=1&isAllowed=y)
7. Colavizza, G., Franzini, G., & Kaplan, F. (2021). Archives and artificial intelligence: A survey of the literature. *ACM Journal on Computing and Cultural Heritage*, 15(1), Article 4. <https://doi.org/10.1145/3457811>
8. Musliichah. (2016). Bunga Rampai Kearsipan. Yogyakarta: Gadjah Mada University.
9. Amsyah, Z. (2003). Manajemen kearsipan. Gramedia Pustaka Utama.
9. Hadiwardoyo, S. (2002). Terminologi Kearsipan Nasional. Jakarta: Arsip Nasional Republik Indonesia.

10. Hegedus, I. (n.d.). How artificial intelligence and machine learning can help rethink archives? [PDF file]. Retrieved from <https://www.ica.org/sites/default/files/ICA-2019-Programme-Book.pdf>
11. International Council on Archives (ICA). (2000). Retrieved 5 April, 2022, from <https://www.ica.org/en>
12. The Korean Society of Conservation Science for Cultural Heritage. (2021). Digital Image Comparisons for Investigating Aging Effects and Artificial Modifications Using Image Analysis Software. *Journal of Conservation Science*, 37(1), 1-8.
13. OpenAI. (2023, July 10). "how to start using Ai for image archives?" [Message post]. AI Language Model. <https://www.openai.com/>
14. OpenAI. (2023, July 10). "how to start using Ai for old photos archives?" [Message post]. AI Language Model. <https://www.openai.com/>
15. OpenAI. (2023, July 19). "Apa Itu Kecerdasan Buatan?" [Message post]. AI Language Model. <https://www.openai.com/>
16. Saur, K.G. (1984). *Dictionary of Archival Terminology*. Retrieved 5 April, 2022, from <https://scalar.usc.edu/works/slaviccollection/media/Dictionary%20of%20Archival%20Terminology.pdf>
17. Nasional, I. D. P. (2008). *Kamus Besar Bahasa Indonesia: Pusat Bahasa*. Michael Cook. (1999). *The Management of Information from Archives*.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

