Miscellanea Archivistica Studia 225

International Archives Symposium
Internationales Archivsymposion in Namur (2025)

Open Data and AI. New chances for Archives?
Open Data und KI. Neue Möglichkeiten für Archive?

PROCEEDINGS/ANNALEN

Editing - Redaktion
Els Herrebout

International Archives Symposium / Internationales Archivsymposion
in Namur (2025)

Open Data and AI. New chances for Archives?
Open Data und KI. Neue Möglichkeiten für Archive?

—

Proceedings / Annalen

Editing – Redaktion
Els HERREBOUT

2026

ALGEMEEN RIJKSARCHIEF
EN
RIJKSARCHIEF IN DE PROVINCIËN

ARCHIVES GÉNÉRALES DU ROYAUME
ET
ARCHIVES DE L'ÉTAT DANS LES PROVINCES

GENERALSTAATSARCHIV
UND
STAATSARCHIVE IN DER PROVINZ

MISCELLANEA ARCHIVISTICA
STUDIA

225

Generalstaatsarchiv – Algemeen Rijksarchief – Archives générales du Royaume
Ruisbroekstraat 2 rue de Ruysbroeck
1000 Brüssel – Brussel – Bruxelles

ISBN : 978 94 6391 661 5
D/2026/531/032
Publ. 6668

Offprint / Sonderdruck

HEIMES, Daniel (2026). The legacy of a Nazi photographer in Gau Moselland - workshop report on AI-supported indexing in collaboration with the Fraunhofer IAO. In Els HERREBOUT (Ed.), International Archives Symposium in Namur (2025). Open Data and AI. New chances for Archives? Proceedings (Miscellanea Archivistica Studia, 225, pp. 71–79). State Archives of Belgium.

DOI: https://doi.org/10.5281/zenodo.20059363
Handle: https://orfeo.belnet.be/handle/internal/14702

The legacy of a Nazi photographer in Gau Moselland - workshop report on AI supported indexing in collaboration with the Fraunhofer IAO[1]

Daniel Heimes[2]

Herbert Ahrens was a photographer integrated into the structures of the Nazi regime. He was most likely employed by the *Nationalblatt* in Koblenz or had a freelance contractual relationship with it. The *Nationalblatt* was the propaganda newspaper of the NSDAP in the *Regierungsbezirk Koblenz* (administrative district of Koblenz). Ahrens produced around 66,000 photographs between 1933 and 1945. Most of these were taken in the Gau Moselland. Before the annexation of Luxembourg the designation was Gau Koblenz-Trier. The photographer's legacy from the Nazi era also includes images from other parts of the German Reich as well as the Western Front and the Eastern Front. The result is a photographic legacy of unique significance.

Years of negotiations by various organisations to acquire this legacy failed again and again. The Rhineland-Palatinate State Archives Administration seized the fortunate opportunity to purchase this legacy in 2022. It received financial support from the *Kulturstiftung der Länder* (Cultural Foundation of the Federal States). Not only was ownership acquired, but also the rights of use and exploitation arising from the copyright. This is far more important from an archive's point of view.

The largest part of the legacy consists of negatives. The indexing will have to show whether a relevant part of the prints on paper are not available as negatives. The negatives proved to be a particular challenge. They are nitrate films, also known as cellulose films. In Germany, these are subject to the *Gesetz über explosionsgefährliche Stoffe*, *das Sprengstoffgesetz* (law on explosive substances, the Explosives Act. The nitrofilms could not be stored in the *Landeshauptarchiv Koblenz* (State Archives Koblenz).

[1] The translation from the German original was done with DeepL. The post-editing was done by humans.

[2] Daniel Heimes is archivist at the Landesarchivverwaltung Rheinland-Pfalz / Landeshauptarchiv Koblenz (Rhineland-Palatinate State Archives Administration / State Archives Koblenz), Head of the Archival Services and official data protection officer.

The *Bundesarchiv* (Federal Archives) provided support with its special storage centre in Berlin. The nitrate films were digitised by a specialist company. When this was achieved in May 2024, the enormous challenge was to make these digitised files accessible. As part of a training course on artificial intelligence in September 2024, the *Landesarchivverwaltung Rheinland-Pfalz* (Rhineland-Palatinate State Archives Administration) was suddenly presented with unprecedented opportunities for rapid indexing. This would have taken many staff years under traditional circumstances. The networking made possible by this training led to contact with the Fraunhofer IAO[3]. It has particular expertise in supporting the introduction of AI into work processes. A contract was signed between the Fraunhofer IAO and the Rhineland-Palatinate State Archives Administration in 2024. The content is to support the AI-based indexing of the Herbert Ahrens photographic legacy and, as a sustainable product, the simultaneous development and creation of a tool that can make such photographic legacies and collections AI-supported accessible in the future.

The project is divided into various phases, which are described below. Because this project is developing something new and the two project partners come from different, in some cases completely unfamiliar areas, it was clear from the beginning that it would have to be a project with evaluation cycles. Mistakes and wrong turns must also be justified to find the better way.

First of all, a contract for order processing in accordance with Art. 28 General Data Protection Regulation had to be concluded so that the photographs, some of which are subject to legal restrictions, for example personal retention periods from the *Landesarchivgesetz Rheinland-Pfalz* (Rhineland-Palatinate State Archives Act), could be processed by the Fraunhofer IAO. The entire files were then transferred to a protected area at the Fraunhofer IAO.

At the same time, common requirements for the result of the indexing were developed. The Rhineland-Palatinate State Archives Administration works with indexing guidelines, which also contain a special section on the indexing of photographs. Their specifications became part of the joint requirements.

The Fraunhofer IAO also researched suitable multimodal language models. Initial tests were also carried out with various providers. The basic principle for all applications was that they were carried out in a protected area from which no data was or is fed back into internet-based AI models. The image

[3] The official English designation is Fraunhofer Institute for Industrial Engineering. The official German Designation is Fraunhofer-Institut für Arbeitswirtschaft und Organisation.

indexing models used for the analysis mirror the entered data back to the internet. The assurances sometimes given by the operators of AI models for paid services that the data would not be used for training purposes and would not be mirrored back do not offer the necessary reliability for such data, some of which is very sensitive. For this reason, the Fraunhofer IAO has proposed working with a local version of Gemma 3, among other things, with all the limitations that this entails - both positive and negative. At this stage, we assume that not all the requirements for indexing the legacy can be met by a single AI model, which is why other models will have to be used for different tasks, as will be shown later on. In the end, however, there should be an overall model that integrates these individual tasks.

The next step was to enable facial recognition by means of a special AI model, which first had to be selected from a range of different models, prepared with regard to the existing requirements and is still being prepared so that it becomes a sub-tool for the entire workflow of AI-supported indexing of image data. In simple terms, this works without any problems using the basic configuration of the AI models for well-known personalities of the Nazi state - Adolf Hitler, Rudolf Hess, Robert Ley, Hjalmar Schacht, etc. come to mind here. Persons in the second and third ranks, as well as unknown persons, had and still have to undergo so-called labelling. In principle, this is a manual process, which requires the intervention of archival colleagues. In the case of the Herbert Ahrens legacy, however, there are also documents that are part of the legacy. One of them is a list of photos compiled mainly with a typewriter. This list was only compiled by Ahrens (or complete or in parts perhaps his daughter – it is not sure) after 1945. There is also a list that is to be regarded as a primary source, which was created at the same time as the photographs. Therefore it reflects the unfiltered view of the Nazi photographer. This is Herbert Ahrens' photo diary, which is only available for the period from 1940 to 1945. In it, he himself recorded all photographs or photo series with motifs, often including personal names and locations, picture numbers and dates. In this context, it was an essential aid for labelling. This data should be able to be assigned to the individual images or image series using AI and consequently be included in the image indexing.

Another aspect of facial recognition is determining the age of the people depicted. These are to be used for the automated allocation of blocking periods, which - according to the objective - should be correct with a very high probability. At this stage of the project, it is not yet possible to say whether this will be achieved with the necessary high probability and therefore legally justifiable.

Further AI models will be used to obtain additional information. So-called CLIP models (Contrastive Language-Image Pretraining) are to be used to identify similar images, image motifs and image series. The benefit lies in the recognition of duplicates, similar locations and the linking of series that are close in time. In addition, the Fraunhofer IAO suggested the possibility of colourising the images, for which the necessary tools would also be available. However, this was rejected by the archive as an intervention going beyond a restoration measure. In the recent past, two colourised films about the so-called old Rhineland were shown in cinemas and also on a streaming provider. Due to their great success, the possibility of AI-supported colourisation is at least mentioned here with regard to public relations work.

As part of the evaluation cycles already mentioned, regular joint discussions are held on the results obtained, at which clear progress can be seen, but also open problems and challenges are discussed. A few translated examples are presented here to illustrate the development:

1. KAISER WILHELM MONUMENT AT THE DEUTSCHES ECK

LHA Ko archives group 700,391 No. 4987 Sub-No. 3 [4]

[4] The official german designation is LHA Ko Best. 700,391 Nr. 4987 UNr. 3.

Status 02.04.2025:

A detailed, black and white image of an equestrian statue on a high pedestal. The statue shows a person on a horse, both figures are equipped with wings. The pedestal is decorated with architectural details and extends high into the sky.

Status 30.04.2025:

Description: Deutsches Eck - Imperial Monument

Details: An equestrian statue is enthroned on an imposing column.

Criticism: The improvement of the indexing performance can be seen in the transfer of the title (here still - in deviation from the indexing guidelines - named as description) from information compiled by Herbert Ahrens. However, the AI model has still failed to adopt the dating ‘vor 1934’ (before 1934) as well. It can be assumed that this was due to the formulation deviating from standardised dates. The more precise indexing under Details is not correct in terms of content. Firstly, the definite article ‘the’ and not the indefinite article ‘an’ should be used. The Deutsches Eck does not include several equestrian statues. Similarly, the statue does not stand on a column, but on a pedestal. The perspective and the visitors to be seen are not mentioned.

2. German soldiers in front of the Arc de Triomphe, Paris 1940

LHA Ko archives group 700,391 No. 2431 Sub-No. 3[5]

[5] The official german designation is LHA Ko Best. 700,391 Nr. 2431 UNr. 3.

Status: 02.04.2025:
Soldiers in uniform stand in front of the Arc de Triomphe in Paris. In the foreground are several military vehicles.

Status: 30.04.2025:
Date: 28.07.1940
Description: Cityscape in Paris - Arc de Triomphe
Details: In front of the Arc de Triomphe are soldiers and vehicles.

Criticism: The Arc de Triomphe was already known at the first attempt due to the data known in advance to the AI model, the local version of Gemma 3, that is, without training by the project participants. The improvement in the indexing performance can be seen in the transfer of the recording date from information compiled by Herbert Ahrens. However, it is problematic that it changed the general series name ‘Städtebilder Paris’ (Cityscapes Paris) to ‘Städtebild in Paris’ (Cityscape in Paris), which represents a qualitative deterioration.

As part of the evaluation on 30th of April 2025, which included a total of 20 photographs with the respective indexings, it was discussed to provide the AI model with further training data to increase quality. To this end, it was firstly agreed to index a package of images from the Ahrens legacy exactly according to the specifications of our indexing guidelines. These images were to be widely distributed, even though it was of course not possible to provide a cross-section of all types of motifs found in the approximately 66,000 photographs. Within a short space of time, 80 images were catalogued as examples and the indexing data was sent to the Fraunhofer IAO for training purposes. Secondly, it was decided that the data from the archive database Dr.Doc to the actual image collection of the State Archives Koblenz, namely those for which digitised images and indexing data – both – are available, are exported and made available to the local AI model at the Fraunhofer IAO for training purposes. Thirdly, a digitised copy of the NSDAP handbook, which is now in the public domain, was also included for training purposes, as it contains, among other things, precise illustrations and classifications of the uniforms and insignia, which should make it easier for the AI model to classify them.

Repeated evaluation cycles are foreseeable for the future in order to achieve a result that fulfils archival requirements. Further possibilities will be offered by increasing computing power, which has already been addressed by the Fraunhofer IAO. The broadening and improvement of the training data also promises qualitative progress for the project.

As soon as the desired quality levels have been achieved from an archival point of view, the next task will be to create an overall processing workflow for all the images in the Ahrens legacy on Fraunhofer IAO servers. All additional information, such as the list that Herbert Ahrens created after 1945, his photo diary from 1940 to 1945, the exemplary indexing information, which means the information created using images from the legacy and the additional information imported from the archive database of the Rhineland-Palatinate State Archives Administration and obtained during the evaluations, etc., must be integrated into the selected overall AI model. In addition, this overall model must be able to incorporate other context data identified as relevant that originates from other AI models, such as the recognition of duplicates, image series, identical motifs and facial recognition. It must also be taken into account that this model must be able to function for other cases, that is other, future image collections, donations with extensive image material, pure photographic legacies, etc.

At the end of the project, Fraunhofer IOA will provide the Rhineland-Palatinate State Archives administration with knowledge and skills. This should enable the Rhineland-Palatinate State Archives Administration to carry out previously implemented processing procedures independently and within the framework of its own IT infrastructure. This can not only be done for archivists, but must also fully involve the archive's IT staff. The Fraunhofer IAO will provide a comprehensive and detailed explanation of the procedures for processing the Herbert Ahrens legacy as well as for future photo collections or partial photo collections. This will also include the implemented codes for image processing, including support through discussion and documentation as well as further information material for the Rhineland-Palatinate State Archives Administration.

Table of Contents/Inhaltsverzeichnis

TABLE OF CONTENTS / INHALTSVERZEICHNIS 5

FOREWORD.....7

VORWORT.....9

Bettina JOERGENS
Impulsvortrag / Keynote speech. Open Data versus Black Box, or: How can AI Fulfill Archival Tasks and Professional Requirements?.....11

Laura DRECHSLER
The EU's legal framework for public archives: from personal data to artificial intelligence 27

Klaas VAN GELDER, C. Annemieke ROMEIN, Xavier. GILLARD
When archives go digital…! Tools, Practices, Opportunities, and Challenges 41

Daniel HEIMES
The legacy of a Nazi photographer in Gau Moselland - workshop report on AI-supported indexing in collaboration with the Fraunhofer IAO 71

6668

ISBN 978-94-6391-661-5

9789463916615

Umschlagbild : Archives de l'Etat à Namur ©AENamur.

belspo