



(Mis)matching Metadata: Improving Accessibility in Digital Visual Archives through the EyCon Project.

Katherine Aske

Northumbria University and Loughborough University

Marina Giardinetti

Université Paris Cité and Université Paris Nanterre

Discussing the current AHRC/LABEX-funded EyCon (Early Conflict Photography 1890-1918 and Visual AI) project, this article considers potentially problematic metadata and how it affects the accessibility of digital visual archives. The authors deliberate how metadata creation and enrichment could be improved through Artificial Intelligence (AI) tools and explore the practical applications of AI-reliant tools to analyse a large corpus of photographs and create or enrich metadata. The amount of visual data created by digitisation efforts is not always followed by the creation of contextual metadata, which is a major problem for archival institutions and their users, as metadata directly affects the accessibility of digitised records. Moreover, the scale of digitisation efforts means it is often beyond the scope of archivists and other record managers to individually assess problematic or sensitive images and their metadata. Additionally, existing metadata for photographic and visual records are presenting issues in terms of out-dated descriptions or inconsistent contextual information. As more attention is given to the creation of accessible digital content within archival institutions, we argue that too little is being given to the enrichment of record data. In this article, the authors ask how new tools can address incomplete or inaccurate metadata and improve the transparency and accessibility of digital visual records.

Keywords: metadata, accessibility, visual archives, photography

The possibilities for conducting research through digitised and born-digital materials are greater than ever. More is possible now than even a few years ago thanks to the global efforts of cultural heritage organisations and the developments of innovative technologies to make digitised and digital records more accessible to their users. In the age of digital technologies, record data or metadata are crucial, yet Kimmo Elo has suggested that there is still too much focus on creating digital material, rather than improving record data.¹ However, endeavours to improve the accuracy and accessibility of metadata, and its potential compatibility with other cataloguing systems, are now becoming more of a priority for archival institutions. There are major projects across the cultural heritage sector, particularly in the UK and US, focusing on the improvement of digital metadata and accessibility, such as *Towards a National Collection* and *Living with Machines*, as well as projects specifically on image archives, including *Access and Discovery of Documentary Images (ADDI)*; *CAMPI* (Computer-Aided Metadata Generation for Photo archives Initiative); and the Frick Collection's Photoarchive.² Many of these projects are developing open source data processing through tools such as the International Image Interoperability Framework (IIIF); Wikidata; and OpenRefine, to create a more centralised, linked data and interoperable approaches to collection records and object descriptions for cultural heritage organisations. Such approaches are enabling, and will continue to enable, larger datasets to be consulted in the development of Artificial Intelligence (AI) and machine learning algorithms.³ However, there remains an unequal investment in archival institutions for image-based projects, as such projects are often expensive, time consuming, and technically intensive.

In this article, the authors, a digital humanist and a computer scientist, address the challenges in the creation and enrichment of metadata within digital photograph archives through AI and machine learning technologies, including object detection, and explore the current findings of the EyCon (Early Conflict Photography 1890-1918 and Visual AI)⁴ project, jointly funded by the Arts and Humanities Research Council (AHRC) in the UK and LABEX in France. One of the primary objectives of the EyCon project is to collect a vast corpora of conflict images, with a particular focus on non-European theatres of conflict, to create an open-access database. The project partners include several French⁵ and UK⁶ institutions that hold photographic collections and photo albums that have been digitised. The EyCon collection is being assembled

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

1556-4673/2023/1-ART1 \$15.00

<http://dx.doi.org/10.1145/3594726>

through manual approaches, where partnering institutions have directly provided their relevant collections, and via search engines and APIs of digital libraries. The large dataset includes not only photographs, but also images from newspapers, magazines and periodicals, which are being extracted with their captions and contextual information to automatically create metadata. The combination of image formats and sources is being used to train computer vision and machine learning tools for object detection, and to help identify the distribution of conflict images. Similar techniques have been successfully applied in the *ADDI* project (2019–2022), where Lauren Tilton and Taylor Arnold have conducted object-based computer vision techniques to help improve image metadata on five twentieth-century image collections from the Library of Congress.⁷ The EyCon project is furthering the research into object detection in historical photographs, as well exploring ways to identify and process sensitive and violent conflict photography from the late-nineteenth and early-twentieth centuries, much of which is held across disparate archival collections. More broadly, the project addresses issues about access and discovery for historical photographic collections that share a thematic coherence. Ultimately, the team hopes to enable research and broaden accessibility to these materials by delivering an intuitive interface that gives users access to an entire database of collected images, which will in turn help image archives to automate the processing of their historical collections and make them more discoverable for users.

With a focus on the EyCon project, this article addresses the current concerns for digital image archives in the cataloguing and reviewing of their metadata, particularly for sensitive or contested records. It considers the bigger picture approaches to the application of AI for visual digital records, while also exploring select examples of archival photographs that have problematic metadata, and some of the ways the EyCon project is addressing these issues.⁸ Considering the growing concerns for photographic record metadata, we ask how digital visual archives can deal—and are dealing—with incomplete or problematic metadata for historical photographic records; the transparency of their algorithmic processing for digital visual records; and the automated creation of record data.

In the first section, the authors examine the importance and common approaches to image metadata creation for visual archives, considering wider issues with out-dated semantic language and the potential biases in the recording of a photograph's provenance. The second section considers the ways AI tools and other digital technologies have been and can be implemented to address these factors, to recognise similarities between photographic images, and metadata including specific times, events, people, photographers, places, or content.⁹ The third section offers working examples of AI tools in relation to the EyCon project, and the use of a range of image sources with text (newspapers, magazines and illustrated periodicals) to train the visual identification software used within the project.

1) CREATING, CLEANING, AND CORRECTING(?) METADATA.

Image archives held by various institutions across the globe are experiencing several and often distinctive difficulties in making their photograph collections available to users. The collection histories of photographic records, including their origins, donors, acquisition details, commentaries, or institutional histories, are frequently incomplete, or missing entirely.¹⁰ But what information is recorded is also determined by the type of institution, and the different approaches and practices employed. For example, libraries might prefer to focus on their content rather than broader contexts, and metadata will often work around a structured catalogue, while for archives, contextual information can be buried within the materials, often as part of a larger collection as opposed to individual items, found through finding aids.¹¹ On the other hand, museums may focus on the provenance of an individual object, rather than its part in a broader context, or even use metadata, as Jenn Riley has argued, 'as a way to interpret collections'.¹² Focussing on museum collections, Elizabeth Edwards and Christopher Morton have suggested that photographs 'are increasingly being understood as knowledge-objects in their own right', and yet few photograph collections held by museums are even inventoried, let alone catalogued. But often, researchers and other users of photographic records may need to work across these sectors, and they will not only have to face the disparity of photographic collections spread throughout archives, museums, libraries, and galleries, but also issues with accessibility, incomplete cataloguing, and the invisibility of some archival records. The varied approaches to record data

have been addressed by Karen Smith-Yoshimura in the OCLC report (2020), who suggests that for image collections, depending ‘on the nature of the collection and its users’, there are questions that will arise, such as ‘identification of works, depiction of entities, chronology, geography, provenance, genre, subjects (‘of-ness’ and ‘about-ness’), which should be included in image metadata.’¹³ However, the first hurdle many institutions are facing is knowing what they have, in order to process, review and categorise image collections, and to identify what information is potentially missing from a record.

Thanks to digitisation, issues with missing information or incorrect data are becoming more visible. As Gross et al. have suggested, metadata, while performing a valuable role, is not always ‘formed with the rigor necessary to be actionable [in a] machine-readable environment’.¹⁴ The large amount of data created by mass digitisation efforts is not always followed by the generation of very precise metadata concerning the technical aspects of an image, such as its dimensions or the digitisation information; or, importantly, a reflection on the categories with which the names of places, persons, or events are classified. A lack of consistency in the processing of metadata is a problem for many institutions and their users, as inaccurate or very thin data can affect the accessibility and discoverability of digitised visual records. However, the scale of digitisation projects can limit the capacities of institutions to individually tailor metadata, especially for potentially problematic or sensitive materials. This is particularly true for image records, where little context may be transferred with the acquisition, and holding institutions may not be able to fully assess the nature of a photograph or collection, or their contexts. This is a critical consideration for the EyCon project when digitising images that represent violence or sensitive scenes, and cultural experiences outside of the host organisation’s nation.

There is a need to pause and reflect about what data is really required of a visual record. The problem with addressing the accuracy or extensiveness of a record is that there are several interpretations of what that might be. For example, there is actual *inaccuracy*, such as the careless recaptioning of images by companies like Alamy.¹⁵ And then there is the instability of image interpretation by different spectatorships, reflected in unfixed captions, particularly for historical photographs. For the latter, tools to identify visual similarity and natural language processing (NLP) might help to trace changing contexts. But accuracy is too loaded a term to reflect these needs. Therefore, the aim to create relevant or consistent record data for current and future users, is our approach.

Demand for digitisation and online accessibility is putting extra pressure on archives holding photographic collections. In the age of mass digitisation initiatives, as well as the transferring of digital files between institutions, the possibility of corrupting an image’s historical record during processing, is ever present.¹⁶ Moreover, the aims to digitise as much as possible for many cultural heritage organisations—while providing greater accessibility for public users and researchers who have internet connections—mean that many digitised collections have underdeveloped metadata or contextual details, as curatorial processes are purposefully reduced to help meet targets. These digitised photographic records then enter large digital interfaces, discoverable only through keyword searches or predetermined filters.¹⁷

As Steven Verstockt et al. have highlighted, many cultural heritage organisations that hold digital photograph collections are facing issues with their metadata, which are impacting the ‘interpretation, exploration and exploitation’ of these visual records.¹⁸ When problematic metadata is attached to digital photographs, it not only impacts the accessibility of photographic collections, but creates further space for cultural biases, misinterpretation, misappropriation, and a prejudiced or nationalistic recording of historical events. Of equal concern is that guidance for how to tackle potentially problematic or missing metadata remains far from straightforward. This becomes further complicated when image collections are of a sensitive, or contested nature, such as conflict imagery. Yet, there are still no universally accepted definitions for what makes good quality metadata.¹⁹ Concerning image archives, the earlier work of projects like the European Visual Archive Project (EVA) and Safeguarding European Photographic Images for Access (SEPIA) at the turn of the millennium, attempted to address the inconsistencies between archival cataloguing approaches to describing photographic collections.²⁰ More recently, scholars such as Anne J. Gilliland, María Montenegro,

and Anna Näslund Dahlgren have outlined and scrutinised the current guidance for processing archival metadata and increasing accessibility.²¹ While more is now being done to recognise and even address inconsistencies in record data processing, the issues of missing or incorrect contextual information, particularly for visual records, remain deeply embedded within archival and collecting practices. These practices are underpinning the data and records that current AI and machine learning developers are using to train their image identification algorithms.

Unlike text-based records, which can, with varying levels of accuracy, be read through automated programmes to generate metadata or enable searchability, photographs require far more complex approaches to automate the creation of contextual metadata. Over the last few decades, great strides have been taken to automate image content extraction through machine learning using digital image archives.²² Digitised image records have inspired a range of innovative approaches and solutions, including 3D scanning, OCR, and interactive digital imaging with AI, enabling users to zoom in or manipulate images. However, while the technologies to aid in the digitisation and online viewing of archival photograph collections are being improved, such solutions come with their own set of challenges, especially in the choice of what can, or should, be made available online.²³ For historical photographs, and particularly those of a sensitive nature (included enslaved persons, conflict or warzone images), if record data is missing, inaccurate, or potentially misrepresenting a period of national trauma, or if the data is just not available, it is incredibly time consuming, if not impossible, for the host organisation, or anyone wishing to use these resources, to contextualise the image. And a single photograph is one of potentially thousands, if not millions held by a single institution. The capacity to undertake such research without the help of automated systems, including sensitivity review, is just too far beyond individual organisations, many of which already have huge backlogs of records to process. Archives without the financial means to employ digital experts, will find it difficult to keep up with the advances in digital technologies. With so many elements to consider, we must first examine the cautions and practical solutions to approaching problematic metadata for digital visual archives.

1.1) Connecting the Data

The problem with missing or thin record data is not just affecting historical photographs. The visual records of our current history are at risk of being poorly recorded too, if recorded at all. Most of our photographs are now born-digital, as many of our mobile phones double as digital cameras and the quality of film-based photography has been far surpassed by digital capabilities. According to Anderson Almeida Firmino et al., hundreds of millions of photographs are added to Facebook daily, as well as 80 million photos and videos on Instagram, but few of these visual records are annotated by their creators.²⁴ Most smart phones will now record basic metadata when their users capture audio or visual images, such as the time and location; some will recognise faces in photographs. But how do these technologies translate within a complex archival context? What can the advances in AI do to help improve the accessibility and the data records of historical photo archives?

Efforts to digitise archival collections are often done in silo. The disconnect between visual archives, where collections or similar images may be scattered between numerous institutions, or even countries, means their usability as historical objects becomes confounded. While this problem is being recognised and addressed by several current projects, the broader issue remains that many historical records, in all formats, have been affected by contemporary prejudices, nationalistic interpretations and/or racist discrimination, and specific contextual details may have been lost as a result. Without researchers being able to access and critically approach these images, the relevance of the metadata will continue to remain a sticking point.

For example, when searching for First World War photographs on the Imperial War Museums UK (IWM) website, users are given a search bar, or the initial options of filtering results by type, 'photograph' and period, 'First World War'.²⁵ These two selections give over 135,000 results. Subsequent filters, including 'Format', 'Creator' and 'Keywords', provide users with predetermined search terms, or the search bar can be used for more specific keywords. Searching for 'colonial' within these filtered records, the results are reduced to 507. Each of these images are labelled with the category 'photographs', and yet the record 'Journée de

l'Armée d'Afrique et des Troupes Coloniales [African Army and Colonial Troops Day]' is an illustrated poster, not a photograph.²⁶ The mis-categorised record is therefore not included in the search for 'Posters' and 'First World War', potentially skewing the search results. Such issues can be corrected by direct communication with the host organisation, but they also remind us that the creation of record metadata will never be a flawless process, even with the use of AI.

The driving force behind the EyCon project is that such disparity in visual records of early conflict, particularly concerning colonial or imperial warfare, has led to the entrenchment of national exceptionalism, and an inability for researchers and public users to cross-examine potentially connected photographic collections. To offer an example of national exceptionalism, and how a photographic record can become distorted when processed in silo, we can examine an image currently held by two different French archives.



Figure 1: 'Blessé français évacué sur l'arrière', 23 octobre 1917, album Valois, La Contemporaine, BDIC-VAL- 006-091.

The black-and-white photograph, dated 23 October 1917, depicts three male soldiers in the First World War, wearing uniform and helmets, standing in a muddy field. One man appears to be injured. The photograph is held by La Contemporaine and the Etablissement de Communication et de Production Audiovisuelle de la Défense (ECPAD). The ECPAD description reads '*Un tirailleur sénégalais est blessé au 'Balcon', position allemande conquise par les alliés près de Soupir*' [A Senegalese rifleman was wounded at the 'Balcon', a German position conquered by the Allies near Soupir]. However, La Contemporaine includes the description handwritten under their image, which is part of an album: '*Blessé français évacué sur l'arrière*' [French wounded evacuated to the rear].²⁷ These different descriptions have automatically become the title of the same image, and demonstrate the necessity for a collaborative approach to the recording of image data and historical context. If users were able to compare these images visually, and analyse their different descriptions together, such potentially problematic metadata and semantic issues would be more easily identified. For EyCon, especially in the inclusion of photograph albums in the database, where different

metadata is held for a photograph or multiple photographs of the same event, all the variant records are being maintained to preserve the contextual labelling of these historical documents for research and discoverability.

2) APPLYING AI TO VISUAL ARCHIVES

The advances in computer vision and multimedia retrieval have been notable over the last two decades.²⁸ Image identification and computer vision technologies are becoming essential within the realms of surveillance, healthcare, text-recognition and even geographic information systems (GIS). For image archives, computer vision is already improving the accessibility of visual records. Image identification algorithms are developed by collecting keyword-described digital images from the internet and using them to train models through machine learning. These processes can be used to create or add metadata; automate archival processes and evaluation; and be used to populate record categories to increase discoverability.

Many recent projects exploring the use of computer vision have focussed on artworks, automating the identification of artists, styles, and subjects.²⁹ Discussing the Frick Art Reference Library's Photoarchive, X.Y. Han et al. have noted the ability for deep neural networks to achieve 'near human-level performance' in the identification of subject matter within digitised images, and such tools have the potential to automate metadata creation and image retrieval.³⁰ While advances continue to be made, technologies to identify subject matter, or what objects are featured in an image, have become so widespread that even a modern iPhone can search a user's personal photos for certain objects, such as a 'dog' or a 'train', albeit with limited success. Such advancements are also making projects like DALL-E—an image generator that works with user-inputted text descriptions—possible.³¹ While the application of AI tools already plays a valuable role in increasing the accessibility and usability of image archives, using existing descriptions and keywords to develop image identification algorithms means the process is not without risk. Such techniques, by necessity, must use already digitised materials to train the algorithms, and are therefore going to be better at identifying images that have been prioritised for digitisation—and prioritised by archives or web-based image databases that have the funds and facilities to digitise and host photographs and make them available to users for computational analysis. For archival institutions, the computer vision tools available are largely pre-trained on colour images of Euro-American visual culture materials, often made up of a huge amount of late twentieth- and early twenty-first-century photographs. For example, COCO³² or ImageNet³³ datasets are renowned for their wealth of information, but often accentuate anachronism and a lack of context in many cases. Such widely used datasets create a noticeable gap when approaching the challenges of creating historical record metadata. Indeed, many existing training datasets are historically biased, and are potentially reproducing racial, social, cultural and economic inequities.³⁴

However, while steps for improvement of automated technologies are heading in the right direction, there are still inherent risks in the application of AI to historical documents. As Lise Jaillant argues, when AI is applied to archives, there is potential to bias the historical record, and consequently affect the records of our collective memories.³⁵ Many archives will be aware of the growing demand to be more transparent about the algorithms that they use to process their digital and digitised materials.³⁶ When automated systems are used in archives for records that are contested and sensitive,³⁷ the ethical implications underlying the use of AI tools are further amplified.³⁸ Researchers and public users alike need to know what decisions have been made and whether a record has been altered within its archival timeline, as such information can dictate how a record is used and its cultural and historical value. As photographic records can be included within other multimedia documents and collections, digitising and categorising an image by itself furthers the risk of lost context. But these are time-consuming and expensive tasks to complete without automated assistance, which in itself can be costly. In this sense, there is a somewhat circular problem developing: to acquire more diverse datasets to suitably train computer vision models and improve automation, carefully curated and diverse training datasets need to be prepared and made available.

This issue has not gone unnoticed.³⁹ The current steps being taken to connect cultural heritage records demonstrate the significance and necessity of AI advancements to the sector, and the conscious need for transparency and collaboration in such approaches. But these steps start small. Take for example The

National Archives (TNA) in London, the home of the UK's government records and a national deposit library. TNA holds one of the largest collections of photographs in Britain, comprising of millions of individual photographs covering the historical timespan since the invention of photography. Yet, while their digital Image Library has over 75,000 keyword-searchable images, most of TNA's extensive photograph collections can only be consulted on site. Moreover, TNA admit that many of their photographs are potentially still undiscovered in volumes buried within their collections, or else remain unidentified or uncatalogued--and this will be true of any large cultural heritage archive.⁴⁰ Where TNA direct their users to other online image libraries, tangible connections between such collections, including the IWM's online database, are not yet available.

The EyCon project, with its collection and labelling of late-nineteenth and early-twentieth-century conflict images, is hoping to improve image recognition tools for historical photographs. By developing new models adapted to a specific corpus and interoperable with other similar archives--a main goal of the EyCon project--we can begin to develop a pipeline that can be appropriated to issues of historical semantics. However, the initial step in this collaborative approach is to ask *how* to improve the creation of record and catalogue metadata. Archivists and other information specialists need to be aware of what is required of this data for research purposes, and what possibilities are realistically achievable through automation. While there is no single approach or answer to these questions, if cultural heritage organisations are better able to share their data and compare records and approach large-scale digital collections with AI tools to help reduce manual labours and spending, this will allow for greater transparency in archives, and a more centralised approach to creating and managing diverse visual record data.

2.1) What is 'problematic' Metadata?

The EyCon team has had to identify what metadata is important to collect for the purposes of the project, and its afterlife. However, assessing the relevancy of a digital image's record is not a straightforward process. But as current approaches are being built around existing records and their metadata, there must be a serious consideration for how this data is approached and applied. Where problematic metadata can be identified in existing records, such as inconsistencies with the object description or the use of outdated or insensitive language--as researchers and users of archival records, we must first ask whether any existing data should be corrected? And if so, by whom?

Many archives are issuing policies and statements regarding offensive language or terminology used in their records. Taking an example from The Keep, an archive centre in East Sussex, UK, partnered with East Sussex Record Office, Brighton and Hove Record Office and the University of Sussex, their 'Inclusive Cataloguing' decolonisation statement outlines that they hold material with direct links to enslavement and British Imperialism, and that the 'catalogue does not always accurately reflect the reality of these relationships'. The statement also acknowledges that such terminology can be a 'barrier to access'.⁴¹ The University of Sussex's policy states that 'Locally, we will identify the use of problematic or incorrect language and take steps to change it, updating terminology and subject headings'.⁴² Many decolonisation statements will include similar approaches in an attempt to improve awareness for users and to break-down accessibility issues. Indeed, the Archives & Records Association (ARA, UK and Ireland) has a recent series of blog posts addressing inclusive cataloguing.⁴³

However, if archives are tackling these issues in isolated and targeted ways within their own institutions, what decisions are being made regarding problematic or incorrect language in catalogue data? And who is making these updates? Without national, or even international guidance and collaborative approaches, are archives not still at risk of furthering cultural and national biases by 'correcting' metadata? An object's metadata is of itself a historical resource. While more is being done to address sensitive or offensive record data through trigger warnings and archival policy updates regarding the decolonising of collections, we cannot simply erase or correct the anachronistic language or potentially misleading descriptions, as they are part of that photograph's history. These details can tell us about how the image was recorded; about the views of those who wrote its original record; and enable analyses of how that data has impacted historical context.

Although there will be controversial decisions made, especially in the choice to remove out-dated and racist language, there is also a strong argument to preserve the semantics of the original record, to keep history accountable.

The 2019 *Le Modèle Noir* exhibition at the Musée d'Orsay, Paris, stands as an example. Orsay's President at the time, Laurence des Cars, made the decision to change some of the titles of the exhibited works to include the names of black subjects, demonstrating a shift in power. The title *Portrait d'une négresse* was changed to *Portrait femme noire*.⁴⁴ While the accusation of *rewriting* history will ring in many ears, the solution to problematic contextual data is to find a balance between retaining original descriptions for research purposes and enriching the metadata to allow modern users to discover images in new lights. The recorded history of an object is what Hannah Turner refers to as 'legacy data', a term that encourages us to think critically about how an object's documentation can embed some narratives while excluding others. Indeed, Turner argues that the most important metadata for an image is tied to its historical or original situation.⁴⁵ The EyCon project is therefore working on the principle that out-dated or mismatching metadata should be preserved, but it should also be enriched by more recent information, and even additional sources of knowledge where contexts may have been lost.⁴⁶

3) EYCON PROJECT IN FOCUS: CREATING DATA FROM VISUAL CONTENT

The very nature of a photo archive makes the relationship between image and metadata problematic: the visual objects in the image need to be represented semantically. The record manager must therefore make a lexicological choice that inevitably creates some distance between the photographer, the photographed, and their description, as well as the images' content.⁴⁷ The art historian Michael Baxandall suggests that such 'description consists of words and concepts in a relation with the picture', but a 'relation' can be both 'complex' and 'problematic'.⁴⁸ By including more generalised descriptions in an image record, in addition to the more specific details such as creator, date, subject etc., a photograph will become more discoverable. However, as Caimei Lu, Jung-ran Park and Xiaohua Hu suggest, 'no single access word, however well chosen, can be expected to cover more than a small proportion' of users' attempts to find what images they need.⁴⁹

Machine learning and computer vision techniques can provide researchers with a shorter bridge between visual content and its semantic data. In replacing and augmenting human eyes and cognitive abilities, such techniques make it possible to focus primarily on the visual content of a photograph, as well as its format and its components, without primitive semantic needs.⁵⁰ Machine learning models, at least on the surface, appear to be impartial compared to the subjective considerations of archivists or other information managers, and researchers. While distant viewing can alleviate the cost and laboriousness of metadata creation, in the end, machine learning tools used to automate metadata are not devoid of subjective semantic concerns, and this is a major concern across the archival sector, and particularly for the EyCon project, considering the sensitive subject matter of the project's materials.⁵¹ As Francois Chollet suggests, 'artificial intelligence isn't about replacing our own intelligence with something else, it's about bringing into our lives and work more intelligence—intelligence of a different kind [a] more augmented intelligence than artificial intelligence'.⁵² This is why automated metadata is not a replacement for human-produced metadata. If AI tools for metadata creation are to be put to good use, a new automated metadata format is needed that allows the archiving of all the information ever created, to retain a history of the record's metadata.⁵³ Computer vision and machine learning can produce information in new formats that, we argue, must be integrated with more traditional metadata forms.

AI-generated metadata cannot be entirely created by an algorithm, as it requires previously annotated records to train the predictive model. For EyCon, the assembled corpus consists of photograph albums and collections, books, postcards, as well as a range of periodicals and magazines containing conflict images from the 1890s to 1918. Extracting images from each medium has allowed us to create a homogeneous database of photographs ready for computational processing, and this collection will be made available for external use. The database will also include a layout analysis for all images extracted from periodicals, magazines and albums, and information that aims to keep the archival records in line with their original material supports.

The prototype database has already allowed us to run similarity searches, which are essential for discovering images reproduced in the press, as well as in common collections between institutions. Moreover, the project is trialling object detection algorithms to identify weapons, landscapes, animals, vehicles in keeping with the themes of the collection, not only to help categorise images, but to enrich descriptive content metadata.

For EyCon's image identification processing we set up a semi-supervised learning algorithm. As we are looking for suitable corpora to help identify sensitive images within the early-conflict period, supervised and unsupervised learning would not be suitable. AI tools respond to the demands made by the trained model designer, i.e., the focus, semantic associations, and what to ignore. These decisions are made clear at the time of the annotation preceding the development of a pre-trained model. As such, the requirement to harmonise the existing information in a way that is effective for machine learning, does not necessarily accommodate a desire to expand and diversify such controlled vocabulary--especially for records that often require item-level descriptions. While machine learning can generalise decisions and apply them to a whole corpus of documents, it is essential to consider that automatically generated metadata can carry just as much bias as manually created metadata. We do not want to categorise images too narrowly and risk limiting their historic realities, but we also need to use controlled vocabularies to fit with the existing ontologies of image archives, and to enable interoperability. Finding a balance is therefore crucial, and we suggest that the future of digital research within photo archives must first look at the images and then the data. In the following sub-sections, we discuss three methods of the EyCon project's metadata creation, each requiring the development of adapted pre-trained models.

3.1) Object Detection and identifying sensitive content

Image recognition and object detection tools can be used to create textual data from the automated visual analysis of an image. Such tools are widely used in the conservation of cultural heritage records, and aim to define a picture by identifying and locating a predefined element during training, to which a semantic label can be attached.⁵⁴ In this sense, computer vision performs the descriptive tagging work normally done by archivists and record managers, which in turn makes it possible to reference and connect the content of a corpus of images, and to carry out searches based on these tags.

One of the aims of the EyCon project is to use descriptive tagging to link, analyse and comment on their curated collections to facilitate discovery and connect these neglected and scattered visual materials. As the corpus includes hugely varying content, the necessary diversity of the created metadata requires the use of several complementary approaches and newly developed tools.⁵⁵ The methodology used by the EyCon project for developing a pre-trained model for object detection includes two main steps. The first prerequisite is the choice of specific classes that correspond to the objects to be detected. The second is to annotate a part of the corpus with all the elements we need and their location to allow automatic recognition. The training is then semi-supervised and the annotations are made for the algorithm to learn from.

However, the largest issue in automated detection is the relationship between the visual content and the vocabulary used to describe it.⁵⁶ For the conflict pictures analysed by EyCon, we have envisaged using several main classes for recognition, which will be integrated into the metadata: weapons, ships, landscapes, and the gender of the individuals depicted. The alleged objectivity of AI tools is therefore continually juxtaposed with the manual choice of vocabulary, which will ultimately constitute the ontology of the image database created through the project. The subjectivity of this choice can be reduced by using existing ontologies that are considered to be as neutral as possible, although any bias in the results will be amplified by the scale of the corpus and will depend on the selection of the descriptors and their interpretation.

Most libraries, for example, will follow one of several indexing systems based on chosen vocabularies, which force the interpretation of certain events simply by selecting one name at the expense of others. One example of the Rameau indexation used by the Bibliothèque Nationale de France is the title 'Boxeurs, Guerre des (1899–1901)', which is used to describe a large variety of realities (such as the dates) that could differ from one ontology to another.⁵⁷ Likewise, the term 'war' could be replaced by 'conflict', 'rebellion' or other

descriptors, depending on the approach, the point of view, and the language used. The ontology used by the EyCon team is based on *Iconclass*,⁵⁸ even though its vision is predominantly Eurocentric. For example, the annotations made during the digitization of the Victor Forbin collection at the SHD (Service Historique de la Défense) using the *Iconclass* categories still must be implemented at scale. We will be testing the possibilities offered by the *Iconclass* test set and apply ImageNet on the classes that make the most sense, keeping only the highest recall rates.

The fact that the images chosen to select the suited vocabulary are only from Europe also poses a problem for a project such as EyCon, which aims to de-Westernise the view of visually archived events.⁵⁹ Where the detection of some features, such as landscapes or objects, might seem unproblematic, ethical issues soon arise when we want to identify and name colonial troops through semantic choices. It is difficult to select vocabularies that will allow us to keep the cultural and geographical context of objects, while avoiding the occidental or Westernised perspective. This is more significant when it comes to out-dated descriptions: for example, the term 'colonial army' does not cover all historical realities. The questions concerning controlled vocabularies are particularly important to consider upstream, as they determine the level of interoperability of the delivered database. A tension therefore surfaces in the decision-making process for the choice of ontology. The universalisation of the vocabulary used by *Iconclass* leads to a harmful over-classification, but an interoperable ontology is necessary for the dissemination and usability of the EyCon database. This is further problematised when one is dealing with populations or events that are unrecorded or even erased from the archive, stunting the emergence of new thinking on controversial issues.⁶⁰

The work of Mrinalini Luthra, Konstantin Todorov, Charles Jeurgens and Giovanni Colavizza on their recent *Unsilencing Colonial Archives* project has approached the issues of problematic archival classification and metadata through automated entity recognition for content-based indexing. An annotation typology was applied to a colonial archive of the Dutch East India Company, and annotations were completed as a shared task, based on state-of-the-art neural network models.⁶¹ While this shared process is beyond the scope of the EyCon project, we share in the wider aims of the dataset creation to train entity recognition models to be more inclusive. The training of new models on more accurate corpora will allow machine learning to target visual content more accurately and reduce digital biases, such as image qualities that are too different, or object-labelling that is too anachronistic.⁶² Pictures depicting colonial conflicts are particularly affected by this lack of semantic precision. As potential bias lies in the choice of vocabulary used by the annotation, we focus on two prerequisites: prior knowledge of the corpus and its content (to define the terms) as well as the connection to a controlled vocabulary linked to a pre-defined and pre-conceptualised ontology.⁶³ The constitution of a corpus is based on thematic, sometimes geographical and temporal criteria on which the visual content depends. The more training a model has on a homogeneous (and therefore, characteristically restricted) dataset, the more accurate it will be when detecting the same type of documents.⁶⁴ However, while such categorisation is necessary, it must be as flexible as possible to avoid locking the image into its network of visual descriptors. For this reason, the wider EyCon team is formed of experts and consultants within archival science, humanities, and computer science, who have been consulted regarding the choices of vocabularies used in the project's training models.

We are also using object detection to help identify the sensitive content of photographs in the database. As the collection is intended to be published online, we will include trigger warnings to identify sensitive content within the collection. However, with a corpus of thousands of images, it is impossible to annotate each sensitive image individually, and this process is also subjective, considering the potentially traumatic nature of the images and their use as research materials. The advantage of automatic detection tools is that they can allow for many images to be recognised as sensitive by an algorithm. For example, Google Vision's 'Detect Explicit content - Safe search'⁶⁵ is being used to help us classify certain photographs in our corpus, most notably those featuring corpses, which have been detected correctly as 'violent', but, on occasion, wrongly as 'racy'. For example, out of a selected corpus of 200 images containing human corpses, only five were recognized as 'likely' to be violent by Google Vision, and only 22 as 'possibly' violent. We are now experimenting with the common errors made by the algorithm and how to keep a constant dialogue with the

shortcomings of existing sensitive content detection tools applied to heritage images. We envision a potential pipeline to improve automated sensitivity classification by using the project's dataset. This pipeline would identify sensitive or contested contents across visual archives using pre-trained models that have been re-trained on high-quality annotations completed by experts from a wide consortium of GLAM sector institutions, who manage potentially sensitive photographic materials. These experts could identify textual captions and visual contents that can be tested and improved by users through a crowdsourcing platform, such as Zooniverse. The *Living with Machines* project is using this platform to check their keyword search results, offering insight into the reliability of automated results.⁶⁶

3.2) Layout Analysis to Extract Metadata from Captions

Another part of the EyCon project involves studying and extracting images from photo albums and the layouts of newspapers to enrich our photo database. We have gathered numerous titles of French and English periodicals and magazines from the late-nineteenth and early-twentieth centuries. This period is host to an outburst of photographic production and mass-produced visual news and constituted the first generation of photojournalists. Comparing this corpus with photographs of the period's conflicts makes it possible to identify the trajectories of several of the images that fall within the project's perimeters.

Walter Benjamin has claimed that the caption to a photograph is equally as important as the photograph itself. Benjamin questions, must we not 'count as illiterate the photographer who cannot read his own pictures? Will not the caption become the most important component of the shot?'⁶⁷ For images in newspapers, a picture makes no real sense without a caption, whether it is used to illustrate a news item or as the basis of a media discourse. Images need context to provide a certain point of view on the events to be disseminated. The captions that accompany such images are therefore intrinsically valuable for describing and identifying them. In fact, these captions are a necessary basis for generating metadata for the image. To process images included in text-based records, we need to extract the images and their related captions, and these two different tasks can be processed by layout analysis technologies.

For example, some relevant datasets for the extraction of periodical layouts are directly implemented in *Layout Parser*⁶⁸, the tool that seems best suited to EyCon's challenges.⁶⁹ The datasets are composed of the image files of the periodicals' pages, as well as files in XML-ALTO⁷⁰ format, indicating each layout element⁷¹, including titles, images, text blocks, maps, graphics, and advertisements. The latter hold a problematic place in our workflow as they are neither intended to be directly extracted and retained as part of our final database, nor the subject of semantic enrichment. When extracting the images from the newspapers without distinction, the associated captions will allow us to sort out those that are thematically relevant and that will be included in our final database.

Several projects concerning layout analysis have already been carried out,⁷² and there are numerous pipelines for document layout analysis.⁷³ The extraction of the photographs from the albums was done with *Layout Parser* without much of a challenge: the images are well recognised, extracted and quickly integrated into the database. Some of the images are more difficult to crop, and these represent almost 20% of the images, but they are still usable for object detection and image similarity, although cropping sometimes results in a loss of information and the image is not as well detected, such as in Figure 2. On the other hand, more than 75% of the photographs in the albums are perfectly detected, even when the albums have complex layouts.



Figure 2: Example of a cropped image from the Valois Album, La Contemporaine.

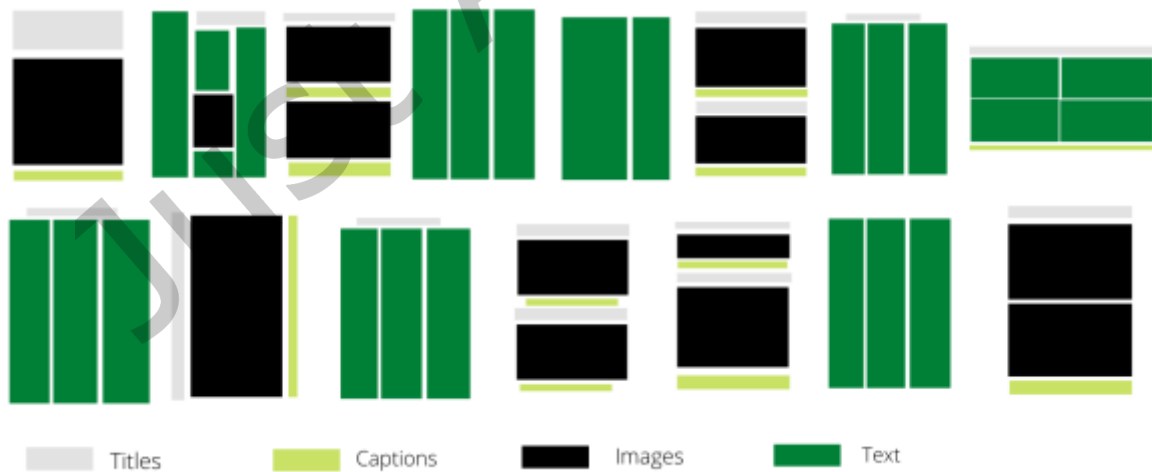


Figure 3: Layout skeleton of L'Excelsior, September 13th, 1914, 16 pages, Gallica, Ref: bpt6k46028367

However, the layouts of periodicals and magazines, especially in the late-nineteenth and early-twentieth centuries, can be very different from one another. One example is the *Excelsior* whose images overlap in a hierarchical jumble (see Figure 3).⁷⁴ Despite the various tools created for the analysis of historical document layouts, few have demonstrated accurate and rigorous predictions when it comes to extracting images and

their captions from periodicals.⁷⁵ It should be noted that, even within a single issue, each page can follow very different editorial rules. This in turn contributes to the inappropriate development of heterogeneous training datasets that are not suitable for applying automated algorithms. We find here the same strategic problem as object detection.⁷⁶ To remedy this, we will continue to improve existing newspaper datasets with other periodicals and magazine titles from different regions and time periods to add variety to the types of document layouts.⁷⁷

That said, providing a layout analysis with a dataset is still uncommon⁷⁸, i.e., XML-ALTO files attached to the images described. Thus, the illustrated periodicals available with XML-ALTO files produced by the Bibliothèque Nationale de France are particularly valuable to the project, as most of the captions are defined and associated with the image they describe.⁷⁹ A dataset of 30,000 images has been created thanks to these annotations, which are converted into the COCO format (JSON file). This format is adapted to the training of visual recognition models, as it contains the dimensions of an object as well as its placement, allowing for the automated creation of metadata. In linking captions directly to an image's metadata, we can avoid severing the image from its physical and editorial context.⁸⁰ Moreover, *Layout Parser* enables OCR for detected text blocks, including image captions. While OCR tools are now widely developed and can provide predictions with a high accuracy rate, it is particularly useful to associate the extracted text directly with its location, as defined by an identifier. *Layout Parser* detects the region of a text block (corresponding to a paragraph) and extracts the text at the same time. Each recognised element is distinguished by an identifier, and by automating the addition of the same identifier to an image and a caption, they can be associated as soon as they are segmented.

To generate the layout analysis for the dataset, considering the diversity of the newspaper and magazine layouts from the beginning of the twentieth century, it was necessary to first make classifications for the pages. The EyCon team have used *Pixplot*⁸¹ to help visualise the different layouts, including cover pages, blank pages, or pages with one or several images, and advertisements. *Pixplot* represents the similarities between images in a whole corpus and vectors them by assigning a numerical value. It then shows the images according to their similarity and can be used to get a first look at an unknown corpus. The aim for EyCon was to provide a pre-classifier of layout pages and isolate particular ones. As the layouts are very different from one page to another, pre-classifying the pages by their layout and developing a model for each type of layout, can be a way of improving accuracy. Even if the regions of the layout are overlapping or messy, the accuracy rate of the images extracted is better thanks to a model trained on more precise features. Such results will enable the study of page layouts and their chronological evolution.

Having visual descriptions of images and their positions in relation to text is invaluable when considering the creation of new metadata from scratch. The evolution of new tools makes it possible to link, in the most precise and objective way possible, an image and its metadata with multimodal (text and vision) analysis tools.⁸² Models such as *CLIP*⁸³ or *Flamingo*⁸⁴ are potential solutions to automatically create very detailed descriptions of images. *CLIP* is trained on pairs of images and text to associate the linguistic concepts and the visual semantics of a picture. It can be used to recreate missing captions for pictures published in newspapers, for example. This bypasses the pre-normalisation of the vocabulary used and avoids the modern and potentially anachronical judgement of the researcher or curator. However, by preserving the original captions, *CLIP* simply repeats the contemporary description. When it comes to navigating potential biases, the same difficulties affect all approaches, from the most manual to the most automated.

3.3) Representing Similarity

The third computational entry point proposed by the EyCon project is to employ visual image searches that do not rely on existing metadata. This appears to somewhat mitigate the dangers of pre-determined vocabularies and semantic biases in the previous approaches. To study the circulation of conflict images in the contemporary press, we are using computer vision to identify the visual similarity between all image types. The pre-trained models do not use the associated metadata but work solely by digital visual matching. The features of each image in the corpora are extracted as vectors, which are used to measure the distance

between them, thus obtaining a similarity score. The higher the ratio, the closer the two images are visually. It is thus possible to identify reuses with identical similarity, but also photographs that are similar and may have been taken at the same event or at the same time. It is important, however, to account for the quality of the images in the collected corpora when considering the similarity scores between two photographs.⁸⁵ Indeed, our photographs extracted from the press are of a far more degraded quality compared to the original prints preserved by institutions and digitised with high-quality cameras.

Different methods have been tested to match images by visual similarity. The poor quality of halftone reproductions of photographs is a hindrance for the AI, so we have enhanced the halftone images with the Jarvis algorithm to preserve the details of the original image.⁸⁶ The imitation of the photographic grain of an analogue camera was simulated using a homogeneous boolean model filter. To increase the robustness of the model, the data can be augmented: this was done in the first similarity tests on a single corpus of thousands of photographs from the First World War⁸⁷. The zoom technique, for example, can be used to simulate variations in the distance between the object and the camera, or the noise technique to simulate defective pixels. A CNN architecture is used (the same as ImageNet) with an unsupervised K-means clustering algorithm that matches the nearest neighbours. It is possible to reach an optimal number of clusters after several training sessions. Thus, thanks to a visualisation (using the t-SNE algorithm), it is possible to identify interesting clusters: a sensitive image cluster is the 'soldiers with facial disfigurement' cluster, for example, which can help to identify scattered photographs that were taken at the same time, or aggregate images of the same location, taken at different times, from slightly different angles and mounted in different albums.

The mixture of photographs and prints within the EyCon project is imperative. A photograph of the Battle of the Somme during World War One kept at the IWM,⁸⁸ which was re-used by *La Guerre illustrée*⁸⁹ (Figure 4), offers a further example of the disparity between visual records held in different institutions. The French magazine captioned the photograph 'The British advance on the Somme: the "London Scottish" going to the trenches'⁹⁰ while it is described in the IWM as 'A Company, 1/14th Battalion, London Regiment (London Scottish) marching to the trenches on Doullens-Amiens road at Pas-en-Artois, 26th June 1916'. To incorporate both examples, the primary metadata needs to be completed. The image metadata from *La Guerre illustrée* for example, can be enriched with an exact date and place, as well as the name of the photographer, as provided in the IWM record. Then, a layer of secondary metadata can be added to both records, providing information regarding the duplication of the image.

As both descriptions need to be retained in the EyCon project's metadata, whether they were created for publication or for conservation purposes, it is necessary to find a conservation format that can differentiate between the layers of information and their origins without removing the additional contexts. Like other types of semi-automated metadata enrichment, the results must be saved in a structured format that the XML requires for their durability and exploitation. The extracted information can then be cross-referenced,⁹¹ allowing places, events, or characters to be identified. The extraction of named entities—ideally with a model trained on its own data—allows this information to be aligned with external databases through controlled vocabulary. This process ensures the veracity of the information cited and its cross-referencing under a single identifier. Continuing with the previous example, the IWM has listed the photographer as 'Brooks, Ernest (Lieutenant)'. By aligning this data, all photos attributed to Brooks can be linked together. This linked construction, through text and image, will allow for the curation of relevant information within the interface, to automate historical photograph analysis and enable images to be collated within subject headings.



Figure 4: La Guerre illustrée, 1 January 1916, p. 14, Gallica, Ref: bpt6k977415x

This brings us to the heart of the issues concerning the reliability of metadata for image records. Most important is the form of the information: firstly, the format in which the metadata is stored, and secondly, in the ways these new computationally developed results are displayed. For similar images discovered between different archives, it is essential to include this information on each image's record, so that users can, regardless of potentially different descriptions, consider these two images together. In this approach, EyCon is building on existing visualisation projects that emphasise visual similarity, such as *Snoop*.⁹² *Snoop* allows searching by visual similarity only, without any need for semantic description, and puts the image at the centre. Additionally, the computer vision tools developed within the *ModOAP*⁹³ project allow the calculation of a similarity score of over 80%. For projects such as *Impresso*⁹⁴, *NewsEye*⁹⁵, *Newspaper Navigator*⁹⁶, or *Chronicling America*⁹⁷, the information extracted from an image makes quantitative analyses possible. Being able to approach a photographic database via its spatial characteristics, for example, is relevant to users studying potentially contested pasts in their direct geographical context.⁹⁸

The improvement of automated metadata creation for image archives will in turn improve the accessibility of visual materials, particularly for humanities researchers. As Dahlgren suggests, humanities scholars comprise a core user group for image collections in cultural heritage organisations, and yet often they feel their needs are not being addressed by existent description practices within image collecting institutions.⁹⁹ Indeed, the demand for historical approaches to be applied to the processing of visual materials—that is, including the necessary information to make photos 'readable' as historical artefacts—has been ongoing, but

often without addressing the practicalities of such labour-intensive endeavours.¹⁰⁰ For EyCon, we will continue to consider what details are needed to ensure a photograph is found, and to avoid it being misinterpreted, or taken out of context, as the way an image is labelled remains an essential part of its discoverability.

CONCLUSION

In digital humanities and computer science projects such as EyCon, there will always need to be an equilibrium between the universality of the tools developed, and the specific issues that tool needs to address. While there must be a constant balance between expectation and practicality, and the knowledge that, as Dalhgren reminds us, metadata ‘can only ever be good enough’,¹⁰¹ by developing more curated, tailored approaches to specific corpora, we can further advance the relevancy of AI to types of archival records. By adapting existing tools and providing curated datasets to other users, we are one step closer to connecting digital visual archives and allowing information professionals to apply computational tools to enrich record data, and potentially apply these tools to address wider issues—such as decolonisation initiatives and the recovery of hidden and marginalised histories. The computer vision and machine learning techniques considered here, including page layout analysis, object detection, and image similarity, can be used to ensure that visual content can be automatically contextualised and made available for large scale analysis. This is significant when records of visual corpora can easily be overtaken by the need for semantic labelling based on human-made decisions. Moreover, these decisions hold more weight when it comes to sensitive and potentially traumatic images, which also need to be preserved and studied. The common purpose of the three approaches presented here is to develop a clear pipeline where the tools need to be understood in their relation to each other. It is difficult to imagine using similarity results if the images and their captions have not yet been extracted, and therefore cannot be related to collections held by other institutions. Likewise, the training of multimodal tools must be run on a dataset containing these same extractions, including images and captions. In this way, the EyCon project intends to bypass contemporary interventionism in semantic data creation, by forming a pipeline assembling extraction from layout detection and the use of multimodal tools, allowing visual descriptors to be generated closer to the original images’ contexts and uses.

Through the EyCon project’s database, we hope to illuminate the ethical problems generated by the application of AI to sensitive collections representing contested pasts, while also allowing hidden records to become more visible, and to make these kind of ethical assessments possible. Such developments are also enabling the improvement of automated sensitivity identification and classification, which will play a significant role in decolonisation initiatives—a next important step for research in this area. Being able to identify and discover the contents of historical visual corpora in more accessible ways will be beneficial to many archival institutions, as well as information professionals and researchers working with historical visual records.¹⁰²

¹ Kimmo Elo. 2020. Big Data, Bad Metadata: A Methodological Note on the Importance of Good Metadata in the Age of Digital History. *Digital Histories: Emergent Approaches within the New Digital History*, Mats Fridlund, Mila Oiva and Patri Paju (Eds.) Helsinki University Press, Helsinki, 103–111. DOI: 10.33134/HUP-5-6.

² Towards a National Collection: <https://www.ukri.org/what-we-offer/browse-our-areas-of-investment-and-support/towards-a-national-collection-opening-uk-heritage-to-the-world/>; Living with Machines: <https://livingwithmachines.ac.uk>; ADDI: <https://www.photometadata.org/About>; CAMPI: <https://github.com/cmu-lib/campi>; and Frick Collection Photoarchive: <https://www.frick.org/library/photoarchive> (all accessed 25 February 2023).

³ See, for example, Eero Hyvönen. 2022. *Publishing and Using Cultural Heritage Linked Data on the Semantic Web*. (Ebook). Springer Nature Switzerland [Reprint of original edition by Morgan & Claypool, 2012]; Ed Jones and Michele Seikel. 2016. *Linked Data for Cultural Heritage*. American Library Association; and Koraljka Golub and Ying-Hsang Lui (Eds). 2021. *Information and Knowledge Organisation in Digital Humanities, Global Perspectives*. Abingdon: Routledge.

⁴ EyCon Project. <https://eycon.hypotheses.org/>

⁵ French Institutions: Gallica, Archives Nationales, Service Historique de la Défense, La Contemporaine, Musée du Quai Branly, Archives Nationales d’Outre-Mer, Établissement de communication et de production audiovisuelle de la Défense.

⁶ UK Institutions: Imperial War Museum, Wellcome Collection, National Library of Scotland

-
- ⁷ See Olivia Dorsey. 2022. Computing Cultural Heritage in the Cloud: Expert Researchers Share Their Outcomes [blog]. Library of Congress, March 21, 2022. <https://blogs.loc.gov/thesignal/2022/03/cchc-researchers-share-outcomes/> (accessed 20 February 2023).
- ⁸ The EyCon project is harnessing AI-reliant tools to analyse a large corpus of early-era conflict photographs. Using images from colonial warfare and pre-1914 conflicts, including the Russo-Japanese war and the Balkans, as well as the many battlefields of the first world war, EyCon is working to improve the metadata of these historical photographs.
- ⁹ See Anssi Männistö, Mert Seker, Alexandros Iosifidis, and Jenni Raitoharju. 2022. Automatic Image Content Extraction: Operationalizing Machine Learning in Humanistic Photographic Studies of Large Visual Archives. *arXiv* (5 April 2022). <https://doi.org/10.48550/arXiv.2204.02149> [accessed 8 September 2022]; and Taylor Arnold and Lauren Tilton. 2020. Enriching Historic Photography with Structured Data Using Image Region Segmentation. In *Proceedings of the 1st International Workshop on Artificial Intelligence for Historical Image Enrichment and Access* (Marseille, France: European Language Resources Association (ELRA), 2020), 1–10. <https://aclanthology.org/2020.ai4hi-1.1>.
- ¹⁰ Elizabeth Edwards and Christopher Morton, eds. 2015. Introduction. *Photographs, Museums, Collections: Between Art and Information*. Bloomsbury, London, 3–23 (pp. 7–8).
- ¹¹ See Jenn Riley. 2017. Understanding Metadata: What is Metadata and What is it For? (2017). Baltimore, National Information Standards Organization (NISO) Primer, p. 5; and K. Megan Gross, Cory Lampert, and Emily Lapworth. 2018. Optimizing Merged Metadata Standards for Online Community History: A Linked Open Data Approach. In Caroline Fuchs and Christine M. Angel, eds. *Organization, representation and description through the digital age: information in libraries, archives and museums*. Berlin, De Gruyter Saur, 206–218.
- ¹² See Riley. 2017.
- ¹³ Karen Smith-Yoshimura. 2020. *Transitioning to the Next Generation of Metadata*. Dublin, OH, OCLC, 19.
- ¹⁴ Gross, et al. 2018.
- ¹⁵ See Michelle Moyd. 2021. Visual Testimonies: Photography, ‘Sudanese’ Soldiers, and Origin Stories in German East Africa’s Colonial Army. *Decolonize the Lens* (2 June 2021), public talk on YouTube. <https://www.youtube.com/watch?v=S49ctzG5ueM> [accessed 8 September 2022].
- ¹⁶ Ayla Stein, Kelly J. Applegate and Seth Robbins. 2017. Achieving and Maintaining Metadata Quality: Towards a Sustainable Workflow for the IDEALS Institutional Repository. *Cataloguing & Classification Quarterly* 55, 7–8 (2017), 644–666. DOI: 10.1080/01639374.2017.1358786.
- ¹⁷ Edwards and Morton. 2015; and Anne J. Gilliland. 2008. Setting the Stage. *Introduction to Metadata* (3rd ed.), Murtha Baca, ed. The Getty Research Institute, Los Angeles, 1–19 (p. 1). See also María Montenegro. 2019. Subverting the Universality of Metadata Standards. *Journal of Documentation* 75, 4 (2019), 731–749. Anna Näslund Dahlgren. 2022. Image Metadata. From Information Management to Interpretative Practice. *Museum Management and Curatorship* (2022), 1–21 (p. 1). <https://doi.org/10.1080/09647775.2022.2073562> [accessed 16 August 2022].
- ¹⁸ Steven Verstockt, Samnang Nop, Florian Vandecasteele, Tim Baert, Nico Van de Weghe, Hans Paulussen, Ettore Rizza, and Mathieu Roeges. 2018. UGESCO - A Hybrid Platform for Geo-Temporal Enrichment of Digital Photo Collections Based on Computational and Crowdsourced Metadata Generation. *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection*. EuroMed 2018 (Lecture Notes in Computer Science) 11196. Springer, Cham. 113–124. https://doi.org/10.1007/978-3-030-01762-0_10.
- ¹⁹ Ayla Stein et al. 2017. p. 645.
- ²⁰ K. Edwin Klijn and Yola de Lusenet. 2000. *In the Picture: Preservation and Digitisation of European Photographic Collections*. European Commission on Preservation and Access, Amsterdam. <http://www.knaw.nl/ecpa/publ/pdf/885.pdf> [accessed 16 August 2022].
- ²¹ Gilliland. 2008. pp. 1–2.
- ²² Anssi Männistö et al. 2022.
- ²³ Temi Odumosu. 2020. The Crying Child: On Colonial Archives, Digitization, and Ethics of Care in the Cultural Commons. *Current Anthropology* 61, S22 (2020), S289–302 (p. S290). <https://doi.org/10.1086/710062>.
- ²⁴ Anderson Almeida Firmino and others. 2022. Automatic and semi-automatic annotation of people in photography using shared events. *Multimedia Tools and Applications* 78 (2019), 13841–13875. <https://doi.org/10.1007/s11042-018-6715-9> [accessed 31 August 2022].
- ²⁵ Imperial War Museums Search. <https://www.iwm.org.uk/collections/search> [accessed 1 September 2022]
- ²⁶ ‘Journée de l’Armée d’Afrique et des Troupes Coloniales [African Army and Colonial Troops Day]’ [IWM Q81339]. <https://www.iwm.org.uk/collections/item/object/205325721> [accessed 1 September 2022].
- ²⁷ Katherine Aske and Lise Jaillant. 2022. Photos of wartime Europe still shape views of conflict – here’s how we’re trying to right the record. *The Conversation* (19 May 2022). <https://theconversation.com/photos-of-wartime-europe-still-shape-views-of-conflict-heres-how-were-trying-to-right-the-record-181880> [accessed 31 January 2023].
- ²⁸ Alkim Almila Akdag Salah. 2021. AI Bugs and Failures: How and Why to Render AI-Algorithms More Human?. *AI for Everyone? Critical Perspectives*. Pieter Verdegem (Ed.) University of Westminster Press, London, 161–179. (p. 172). See also David G. Stork. 2009. Computer Vision and Computer Graphics Analysis of Paintings and Drawings: An Introduction to Literature. *International Conference on Computer Analysis of Images and Patterns*. Springer, Heidelberg, Berlin, 9–24.
- ²⁹ See X.Y. Han, Ellen Prokop, and Vardan Papyan. 2022. Artificial Intelligence and Discovering the Digitised Photoarchive. *Archives, Access and Artificial Intelligence: Working with Born-Digital and Digitized Archival Collections*, Lise Jaillant (Ed.) Transcript Verlag, Bielefeld, 29–60; see also Stork. 2009.

- ³⁰ X.Y. Han et al. 2022. p. 30.
- ³¹ Dalle-E. <https://openai.com/dall-e-2/> [accessed 1 September 2022].
- ³² COCO - Common Objects in Context. <https://cocodataset.org/#home> [accessed 22 September 2022].
- ³³ ImageNet. <https://www.image-net.org/> [accessed 22 September 2022].
- ³⁴ See Thomas Smits and Melvin Wevers. 2021. The Agency of Computer Vision Models as Optical Instruments', *Visual Communication* 21, 2 (2021), 329–349. <https://doi.org/10.1177/1470357221992097> [accessed 8 September 2022].
- ³⁵ Lise Jaillant (Ed.). 2022. Introduction. *Archives, Access and Artificial Intelligence: Working with Born-Digital and Digitized Archival Collections*. Transcript Verlag, Bielefeld, 7–28 (p. 23).
- ³⁶ See *Archives, Access and Artificial Intelligence*. 2022; and Heike Felzmann, Eduard Fosch Villaronga, Christoph Lutz and Aurelia Tamò-Larrieux. 2019. Transparency You Can Trust: Transparency Requirements for Artificial Intelligence between Legal Norms and Contextual Concerns. *Big Data & Society* 6, 1 (2019), 1–14.
- ³⁷ Daniel Foliard. 2020. *Combattre, punir, photographier*. La Découverte, Paris. <https://doi.org/10.3917/dec.folia.2020.01>.
- ³⁸ About the question of ethics and digitization see: Odumosu. 2020. David Mindel. 2021. Ethics and Digital Collections: A Selective Overview of Evolving Complexities. *Journal of Documentation* 78, 3 (2021), 546–563. <https://doi.org/10.1108/JD-11-2020-0193>.
- ³⁹ See Athiya Deviyani. 2022. Assessing Dataset Bias in Computer Vision. *arXiv* (2022), 1–50. <https://doi.org/10.48550/arXiv.2205.01811>; Tianlu Wang, Jieyu Zhao, Mark Yatskar, Kai-Wei Change, and Vincente Ordonez. 2019. Balanced Datasets are not Enough: Estimating and Mitigating Gender Bias in Deep Image Representations. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (2019), 5310–5319; and Adam Zewe. 2022. Can Machine-Learning Models Overcome Biased Datasets. *MIT News* (21 February 2022). <https://news.mit.edu/2022/machine-learning-biased-data-0221> [all accessed 9 September 2022].
- ⁴⁰ The National Archives Guide to their Photograph Collection. <https://www.nationalarchives.gov.uk/help-with-your-research/research-guides/photographs/#2-the-span-of-the-collection> [accessed 31 August 2022].
- ⁴¹ The Keep. <https://www.thekeep.info/inclusive-cataloguing/> [accessed 19 October 2022]
- ⁴² The University of Sussex Decolonisation Statement. <https://www.sussex.ac.uk/library/about/strategy/decolonisation> [accessed 25 February 2023].
- ⁴³ ARA Decolonisation Blog. <https://www.archives.org.uk/news/tag/Decolonising+Blogs> [accessed 19 October 2022]
- ⁴⁴ See Jenny Hughes. 2019. A Sign of Changing Times, 'Le modèle noir; at the Musée d'Orsay Addresses Race in Art. *Frenchly* (26 March 2019). <https://frenchly.us/le-modele-noir-au-musee-dorsay-march-26-july-21> [accessed 9 September 2022].
- ⁴⁵ Hannah Turner. 2020. *Cataloguing Culture: Legacies of Colonialism in Museum Documentation*. UBC Press, Vancouver. p. 7.
- ⁴⁶ Turner. 2020; and Dahlgren. 2022. p. 12; see also Gilliland. 2008.
- ⁴⁷ Schwartz. 2002.
- ⁴⁸ 'Patterns of Intention', *Yale University Press* <<https://yalebooks.yale.edu/9780300037630/patterns-of-intention>> [accessed 30 September 2022].
- ⁴⁹ George W. Furnas, Thomas K. Landauer, Louis M. Gomez and Susan T. Dumais. 1987. The Vocabulary Problem in Human-System Communication. *Communications of the ACM* 30, 11 (November 1987), 964–971.
- ⁵⁰ Taylor Arnold and Lauren Tilton. 2020. Enriching Historic Photography with Structured Data Using Image Region Segmentation. In *Proceedings of the 1st International Workshop on Artificial Intelligence for Historical Image Enrichment and Access* (Marseille, France: European Language Resources Association (ELRA) (2020), 1–10. <https://aclanthology.org/2020.ai4hi-1.1>.
- ⁵¹ Caimej Lu, Jung-ran Park, and Xiaohua Hu. 2010. User Tags versus Expert-Assigned Subject Terms: A Comparison of LibraryThing Tags and Library of Congress Subject Headings. *Journal of Information Science* 36, 6 (2010), 763–779. <https://doi.org/10.1177/0165551510386173>; also, Taylor Arnold and Lauren Tilton. 2019. Distant Viewing: Analyzing Large Visual Corpora. *Digital Scholarship in the Humanities*, Volume 34, 1, (2019), pp. i3–i16. <https://doi.org/10.1093/llc/fqz013>.
- ⁵² Francois Chollet. 2017. *Deep Learning with Python* (1st ed.). Manning Publications, Shelter Island.
- ⁵³ Yalemisew Abgaz Renato Rocha Souza, Japesh Methuku, Gerda Koch, and Amelie Dorn. 2021. A Methodology for Semantic Enrichment of Cultural Heritage Images Using Artificial Intelligence Technologies. *Journal of Imaging*, 7, 8 (2021), 121. <https://doi.org/10.3390/jimaging7080121>.
- ⁵⁴ Francesca Condorelli et al. 2020. A Neural Networks Approach to Detecting Lost Heritage in Historical Video. *ISPRS International Journal of Geo-Information* 9, 5 (2020), 297. <https://doi.org/10.3390/ijgi9050297>.
- ⁵⁵ Ferdinand Maiwald. 2019. Generation of a Benchmark Dataset Using Historical Photographs for an Automated Evaluation of Different Feature Matching Methods. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W13 (2019), 87–94. <https://doi.org/10.5194/isprs-archives-XLII-2-W13-87-2019>.
- ⁵⁶ Arnold and Tilton. 2019.
- ⁵⁷ It could be translated as 'The Boxer war (1895-1900)'.
⁵⁸ Iconclass. <https://iconclass.org/> [accessed 22 September 2022].
- ⁵⁹ The corpus used is described into this paper: Hans Brandhorst. 2019. A Word is Worth a Thousand Pictures - Why the Use of Iconclass Will Make Artificial Intelligence Smarter' (2019). https://labs.brill.com/ictestset/ICONCLASS_and_AI.pdf [accessed 5th October 2022].

- ⁶⁰ ‘Provisional Semantics’. *Imperial War Museums*. <https://www.iwm.org.uk/research/research-projects/provisional-semantics> [accessed 22 September 2022].
- ⁶¹ Mrinalini Luthra, Konstantin Todorov, Charles Jeurgens and Giovanni Colavizza. 2023. Unsilencing Colonial Archives via Automated Entity Recognition. *Journal of Documentation*, vol. ahead of print (2023), 1–24. <https://doi.org/10.1108/JD-02-2022-0038>.
- ⁶² Florian Eiler, Simon Graf, and Wolfgang Dorner. 2018. Artificial Intelligence and the Automatic Classification of Historical Photographs. In *Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality* (presented at the TEEM’18: Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality, Salamanca Spain: ACM (2018), 852–856. <https://doi.org/10.1145/3284179.3284324>.
- ⁶³ Mariana Ziku. 2020. Digital Cultural Heritage and Linked Data: Semantically-Informed Conceptualisations and Practices with a Focus on Intangible Cultural Heritage. *LIBER Quarterly: The Journal of the Association of European Research Libraries* 30, 1 (2020), 1–16. <https://doi.org/10.18352/lq.10315>
- ⁶⁴ Enrique Manjavacas and Lauren Fonteyn. 2022. Adapting vs. Pre-Training Language Models for Historical Languages. *Journal of Data Mining & Digital Humanities*, NLP4DH. Digital Humanities in Languages (2022), 9152. <https://doi.org/10.46298/jdmdh.9152>.
- ⁶⁵ Google Safe Search. <https://cloud.google.com/vision/docs/detecting-safe-search?hl=en> [accessed 22 February 2022].
- ⁶⁶ Living with Machines Zooniverse. <https://www.zooniverse.org/projects/bldigital/living-with-machines/classify> [accessed 22 February 2022].
- ⁶⁷ Walter Benjamin. 1972. A Short History of Photography. *Screen* 13, 1 (1972), 5–26. <https://doi.org/10.1093/screen/13.1.5>.
- ⁶⁸ Layout Parser. <https://layout-parser.github.io> [accessed 22 February 2023].
- ⁶⁹ See Benjamin Charles Germain Lee, Jaime Mears, Eileen Jakeway, Meghan Ferriter, Chris Adams, Nathan Yarasavage, Deborah Thomas, Kate Zwaard, and Daniel S. Weld. 2020. The Newspaper Navigator Dataset: Extracting and Analyzing Visual Content from 16 Million Historic Newspaper Pages in Chronicling America. *arXiv* (2020), 14 pages. <https://doi.org/10.48550/arXiv.2005.01583>; and Zejiang Shen, Ruochen Zhang, Melissa Dell, Benjamin Charles Germain Lee, Jacob Carlson, Weining Li. 2021. LayoutParser: A Unified Toolkit for Deep Learning Based Document Image Analysis. *arXiv* (2021), 1–16. <https://doi.org/10.48550/ARXIV.2103.15348>.
- ⁷⁰ ALTO (Analysed Layout and Text Object): ALTO files allow to describe layouts within the XML standards.
- ⁷¹ Nicolas Gutehrlé and Iana Atanassova. 2021. Logical Layout Analysis Applied to Historical Newspapers. <https://hal.archives-ouvertes.fr/hal-03468972> [accessed 23 September 2022].
- ⁷² Béatrice Joyeux-Prunel. 2019. Visual Contagions, the Art Historian, and the Digital Strategies to Work on Them. *Artl@S Bulletin* 8, 3 (2019). <https://docs.lib.purdue.edu/artlas/vol8/iss3/8>.
- ⁷³ Apostolos Antonacopoulos, David Bridson, Christos Papadopoulos, and Stefan Pletschacher. 2009. A Realistic Dataset for Performance Evaluation of Document Layout Analysis. In *2009 10th International Conference on Document Analysis and Recognition* Barcelona, Spain: IEEE (2009), 296–300. <https://doi.org/10.1109/ICDAR.2009.271>.
- ⁷⁴ Excelsior (Paris. 1910) – 32 Années Disponibles – Gallica. <https://gallica.bnf.fr/ark:/12148/cb32771891w/date> [accessed 23 September 2022].
- ⁷⁵ ‘Eynollah’ (QRATOR-SPK, 2022). <https://github.com/qurator-spk/eynollah> [accessed 23 September 2022].
- ⁷⁶ Melvin Wevers and Thomas Smits. 2020. The Visual Digital Turn: Using Neural Networks to Study Historical Images. *Digital Scholarship in the Humanities* 35, 1 (April 2020), 194–207. <https://doi.org/10.1093/dsh/35.1.fqy085>.
- ⁷⁷ Excelsior (Paris. 1910) – 32 Années Disponibles – Gallica. <https://gallica.bnf.fr/ark:/12148/cb32771891w/date> [accessed 23 September 2022].
- ⁷⁸ Nicolas Gutehrlé and Iana Atanassova. 2022. Processing the Structure of Documents: Logical Layout Analysis of Historical Newspapers in French. *Journal of Data Mining and Digital Humanities* (2022), 1–25. <https://hal.archives-ouvertes.fr/hal-03681657>.
- ⁷⁹ Documents de Presse Numérisés En Mode « article » Api. <https://api.bnf.fr/fr/documents-de-presse-numerises-en-mode-article> [accessed 23 September 2022].
- ⁸⁰ Paul Fyfe and Qian Ge. 2018. Image Analytics and the Nineteenth-Century Illustrated Newspaper. *Journal of Cultural Analytics* (2018), 1–25. <https://doi.org/10.22148/16.026>.
- ⁸¹ Yale DHLab – PixPlot. <https://dhlab.yale.edu/projects/pixplot/> [accessed 23 September 2022].
- ⁸² Alec Radford et al. 2021. Learning Transferable Visual Models From Natural Language Supervision. *arXiv* (2021). <https://doi.org/10.48550/arXiv.2103.00020>.
- ⁸³ CLIP: Connecting Text and Images. *OpenAI*. <https://openai.com/blog/clip/> [accessed 23 September 2022].
- ⁸⁴ Jean-Baptiste Alayrac and others. 2022. Flamingo: A Visual Language Model for Few-Shot Learning. *arXiv* (2022). <https://doi.org/10.48550/arXiv.2204.14198>.
- ⁸⁵ Anil Poudel. 2021. *Face Recognition on Historical Photographs*. Master’s Thesis. Uppsala University. 55 pages. <http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-462551>.
- ⁸⁶ Mohamed Salim Aissi. 2023. Comment retrouver des photographies historiques similaires à une image requête?. Rapport de projet de M2. Master Informatique de Sorbonne Université.
- ⁸⁷ This is the Valois collection, held at the Contemporaine, Nanterre.
- ⁸⁸ The Battle of the Somme, July–November 1916. *Imperial War Museums*. <https://www.iwm.org.uk/collections/item/object/205072238> [accessed 23 September 2022].

-
- ⁸⁹ La Guerrel. *Gallica*. 1916. <https://gallica.bnf.fr/ark:/12148/bpt6k977415x> [accessed 23 September 2022].
- ⁹⁰ In French: 'L'avance britannique sur la Somme: les « London Scottish » allant aux tranchées.'
- ⁹¹ Ziku. 2020.
- ⁹² Olivier Buisson, Jean-Christophe Lombardo, and Alexis Joly. 2019. Snoop (hal-02096036). *HAL Open Science* (2019). <https://hal.archives-ouvertes.fr/hal-02096036>.
- ⁹³ ModOAP – Modèles et outils d'apprentissage profond. <<https://modoap.huma-num.fr/>> [accessed 23 September 2022].
- ⁹⁴ impresso. Media Monitoring of the Past. *Impresso* (impresso). <<https://impresso-project.ch>> [accessed 23 September 2022].
- ⁹⁵ NewsEye. <https://www.newseye.eu/> [accessed 23 September 2022].
- ⁹⁶ Newspaper Navigator. <https://news-navigator.labs.loc.gov> [accessed 23 September 2022].
- ⁹⁷ Lee et al. 2020.
- ⁹⁸ Taylor Arnold, Stacey Maples, Lauren Tilton, and Laura Wexler. 2017. Uncovering Latent Metadata in the FSA-OWI Photographic Archive. *Digital Humanities Quarterly*, 011.2 (2017). See also projects as: Photogrammar. <https://photogrammar.org> [accessed 23 September 2022].
- ⁹⁹ Dahlgren. 2022.
- ¹⁰⁰ Helena Zinkham. 2006. Reading and Researching Photographs. *Photographs: Archival Care and Management*. Mary Lynn Ritzenthaler and Diane Vogt-O'Connor (Eds). Chicago, The Society of American Archivists, 59–77 (p. 59). See also William H. Leary. 1985. *The Archival Appraisal of Photographs: A RAMP Study with Guidelines*. United Nations Educational, Scientific, and Cultural Organization, Paris. Section 2.5.2.; Tim Schlack. 2008. Framing Photographs, Denying Archives: The Difficulty of Focusing on Archival Photographs. *Archival Science* 8, 2 (2008), 85–101; Chassanoff. 2018.
- ¹⁰¹ Dahlgren. 2022.
- ¹⁰² Nicolas Gonthier et al. 2018. Weakly Supervised Object Detection in Artworks. In *Computer Vision – ECCV 2018 Workshops*. ECCV 2018. Munich, Germany. https://doi.org/10.1007/978-3-030-11012-3_53.