Beate Löffler; Tino Mager

Minor Politics, Major Consequences: Epistemic Challenges of Metadata and the Contribution of Image Recognition

2020

https://doi.org/10.25969/mediarep/21893

Veröffentlichungsversion / published version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:
Löffler, Beate; Mager, Tino: Minor Politics, Major Consequences: Epistemic Challenges of Metadata and the Contribution of Image Recognition. In: *Digital Culture & Society*. The Politics of Metadata, Jg. 6 (2020), Nr. 2, S. 221–238. DOI: https://doi.org/10.25969/mediarep/21893.

Minor Politics, Major Consequences

Epistemic Challenges of Metadata and the Contribution of Image Recognition

Beate Löffler, Tino Mager

Abstract

Metadata is part of our knowledge systems and, so, represents and perpetuates political hierarchies and perceptions of relevance. While some of these have come up for scrutiny in the discourses on digitization, some 'minor' issues have gone unnoticed and a few new mechanisms of imbalance have escaped attention as well. Yet, all of these, too, influence the usability of digital image collections.

This paper traces three fields of 'minor politics' and their epistemic consequences, both in general and in particular, with respect to the study of architecture and its visual representation: first, the intrinsic logic of the original collections and their digital representation; second, the role of support staff in the course of digitization and data transfer; and, third, keywording as a matter of disciplinary habitus. It underlines the 'political' role of metadata within the context of knowledge production, even on the local level of a single database, and connects to the implementation of contemporary technologies like computer vision and artificial intelligence for image content classification and the creation of metadata.

Given the abundance of digitally available (historical) images, image content recognition and the creation of metadata by artificial intelligence are sheer necessities in order to make millions of hitherto unexplored images available for research. At the same time, the challenge to overcome existing colonial and other biases in the training of AI remains. Hence, we are once again tasked to reflect on the delicate criterion of objectivity. The second part of this paper focuses on research done in the ArchiMediaL project (archimedial.eu); it demonstrates both the potentials and the risks of applying artificial intelligence for metadata creation by addressing the three fields mentioned above through the magnifying glass of programming.

Keywords: architectural history, interdisciplinary, machine learning, image recognition, visual data bases, epistemic challenges, metadata

DOI 10.14361/dcs-2020-0211

Introduction

Metadata is part of our knowledge systems. It, therefore, represents and perpetuates political hierarchies and perceptions of relevance. Some of this comes up whenever curators discuss corpora for digitization, or the means and limits of access, for example. It becomes apparent when the humanities and computer science negotiate the hierarchies of cooperation or the technical parameters of interfaces, ontologies, database design, and data storage. Hence, some factors have already been scrutinized in the discourses on digitization; their consequences are part of the everyday intellectual processes of digital humanities. This decision-making process represents the 'major politics' of metadata. At the same time, many 'minor' issues go largely unnoticed, or are only discussed in small circles of specialists. Yet, they influence the usability of digital image collections and the reliability of their content as well, and thereby have a significant impact on the epistemic meaning of the resultant research.

This paper is inspired by more than two decades of work for—and with—digital image collections between cultural studies and architectural history. It initially traces three fields of 'minor politics' of discussing and attributing metadata and their epistemic consequences, both in general in cultural studies, and in particular with regard to the study of architecture and its visual representation. The discussion underlines the critical role of metadata within the context of knowledge production, showing localized and/or particular issues as part of the overall challenges of metadata. Based on this, it argues for an additional, conceptionally non-textual level of metadata creation, as represented by the implementation of contemporary technologies like computer vision and artificial intelligence for image content classification.

The paper reflects on the motivations and insights of our work within the ArchiMediaL project (archimedial.eu) to demonstrate both the potentials and the risks of applying artificial intelligence (AI) for metadata creation by addressing the epistemic and epistemological challenges from a programming perspective. Here, it becomes evident how the seemingly minor decisions of programming become crucial for the 'major politics' of metadata.

Epistemic challenges of metadata production

When digitization of visual media became a feasible and fundable issue during the 1990s, it aimed to enable access to important cultural objects and materials, and to customize teaching and exchange in research. The digital items were tools that were thought to facilitate processes, and by no means to replace the original collections and the work with those. With the further development in storage capabilities, digital photography, and the introduction of Web 2.0, the framework changed considerably. The available material grew exponentially, and,

soon, the lines between digitized and originally digital material blurred. Over time, ideas arose not to keep the analogue sources but to save archival space by substituting them with their digital representations. With this notion out, and in discussion, the epistemic specifics of both analogue and digitized databases, the intrinsic mechanisms of digitization, and the allocation of meta-text acquired crucial relevance (cf. Matyssek 2009). This is even more so, given the increasingly different work experiences, research approaches, and the expectations of different generations of scholars in interacting with different types of visual databases.

We address three select phenomena from the digitization process to point towards the complexity of 'metadata politics' at a localized and field-specific, and, so, 'minor' level. We reflect on epistemically relevant decision-making from the angle of architectural history: the intrinsic logic of the original collections and their digital representation; the role of support staff in the course of digitization and data transfer; and keywording as a matter of disciplinary habitus.

The intrinsic logic of the original collections and their digital representation

Analogue collections of visual material arose from the most diverse contexts, and went on to carry in themselves the resultant specifics. There are, for example, materials compiled for teaching and collections of visual material generated in the course of research that are primarily of epistemic interest today since they make it possible to trace the theories of knowledge throughout history. There is original material resulting from the research processes themselves, such as sketches, photographs, or diagrams, which are both irreplaceable sources and parts of an order of thought; and there are repositories, including visual and textual materials of different types, and even artefacts, the heterogeneity of which challenges the classification systems of librarians and archivists. Transferring these collections into digital representations should actually mean understanding their intrinsic logic and finding ways to represent it in the ensuing meta-text. Otherwise, the historic and epistemic relevance of the collection will disappear: the images remain in their digital reincarnation but their meaning and context will be lost.

The points of decision-making depend on the specific lot. Teaching collections usually contain secondary materials; their digitization is often a matter of convenience first, and becomes a research topic only in retrospect. Yet, the structure of a slide magazine—the order of motifs—is relevant. Ethnologists talking about a human's transition through life or about a religious festival consisting of a series of rituals need to reproduce the correct sequence in their narration. The sequence of events carries meaning in the biography and the ritual as well as for the rehearsal of research methods and analysis. The meta-text needs to reproduce the timeline represented in the slide archive in an appropriate way, enabling later users to grasp the idea of the original collection. In contrast, the temporal

dimension is secondary for teaching slides concerning artefacts, such as architecture or the inventories of museum collections. The genesis of forms over time or the mechanisms of construction or production play a role as well. Yet, the structures of knowledge evolve not from processes but rather around objects, places, or actors. The order of images aims to provide a general picture of the object in question and follows an established pattern of approach and appropriation from long distance to close-up view, from exterior to interior, from general to specific. Here, the single image gains meaning through its relationship with other images of the same object or the same artist, the linking done in the mind of the experienced scholar, or helped along by the catalogue of the slide collection. Digitization needs to represent not only direct connections but secondary and tertiary levels of order as well. Despite this, teaching collections adhere to very ordered systems of knowledge, mirroring basic training in the respective fields. Hence, the transfer of the underlying reference system is comparatively easy since the database structures of digitization follow similar ideas and encourage the most common kinds of interlinking. Here, the minor policies of handling a specific collection align with the overarching policies of metadata.

This becomes different as soon as original research collections are taken into account. Here, the logical order of the collection meets with the intrinsic developments of research to make one file packed to the bursting point with content or even overflowing to another, while the next remains entirely empty. Original material mixes with copies and reproductions or redrawings. The visual material is easy enough to extract and digitize, but the layers of collecting, sorting, and ordering are nearly impossible to reproduce. There is a haptic and epistemic difference between a dozen photographs illustrating a specific building, glued individually on numbered index cards on the one hand, and the same number of photographs stuffed together in a manila envelope with a shared number on the other. After digitization, all the images become equal 'individuals'; their former belonging—their 'invisible meta-text'—is usually reduced to the citation of reference numbers alone. The crucial information to understand the historical formation of scholar and material, the 'becoming' of the original collection with its shifts and changes, its notes and crossed-out sections, remains but readable in the analogue material.

Such challenges are even more extensive when the collections have an actual spatial character, such as the *Dokumenten-Kabinett europäischer Geschichte, Gegenwart und Zukunftsplanung* [Document cabinet of European history, present and future planning] of German legal expert Alexander Dolezalek (1914–1999), or the Friedrich Achleitner Archiv in Vienna, and have to be removed from their original locations, restored, or reassembled to provide appropriate preservation (Kellner 2008). It would be possible to create a digital copy of the collections to 'describe' their spatial dimensions, as we are able to build 3D reconstructions of archaeological sites or crime scenes by now. Today, however, the economic and personnel expenditure for such an endeavour is limited largely to collections of

relevance compared to World Heritage sites or to diverse kinds of pilot projects, exploring the possibilities or recent technological developments.

These brief thoughts point towards digital databases as not being 'identical twins' of the analogue collections. The two kinds of collections are autonomous entities with their respective strengths and weaknesses, for which we can but partially compensate by meta-text. Even more so, when material or spatial qualities undergo translation into visual or textual information, first into images and second into descriptions or keywords. In consequence, the meta-information on the origin and intrinsic order of an original collection of visual material is largely a black box, depending on the awareness of the acting curators, the aim of the specific digitization, and the local policies (cf. Kohle/Locher 2019). The latter is of critical importance for many visual databases containing digitized material since they depend on how much funding, time, and staff are available for the digitization itself and the transfer of metadata from one medium to the other.

The role of support staff in the course of digitization and data transfer

Digitization means bulk production. The curators, computer scientists, and funders create the general framework of digitization; they do not themselves digitize. Except for material of the utmost importance, support staff usually do the entire process of scanning, storing, labelling, keywording, and meta-texting. The work is understood as being a menial one for its repetitive character and low level of complexity. The largely anonymous staff members with their sparse and/or performance-related payment enable the amount of digitization we experienced during the past few decades. At the same time, they create another black box with their decisions during the work processes: experience shows a number of situations where digitization is not only a practice of copying from one medium to another for the benefit of accessibility and convenience, but as a practice of actual information production. Yet, changes in content might remain unnoticed; data might be created, lost, or changed, unintentionally modifying the content of the source material. This is part of the natural way of collecting and ordering knowledge for which scholarship developed the advanced checks of source criticism. These, however, do not necessarily translate themselves into the digital era.

A slide might get scanned mirrored; the image editing might alter the image significantly or even beyond recognition; identical or seemingly identical images might both be scanned, or not. In a series of images stored together, an image or two might miss the notes on the backs that all the others have. Does this mean that these images are irrelevant, or is it sensible to transfer the information from the others to these as well since they are stored together anyway? Here, a broad field of minor decisions has the potential to impact the metadata. Are spelling errors in the original metadata transferred or corrected? How do we handle place

names or territorial allocations that have changed over time? How about terms that are perceived as racist, militarist, otherwise politically incorrect, or only so old-fashioned that they have become incomprehensible to most of the audience?

For some of these issues, approaches to solutions are provided for in guidelines for digitization, often based on the experiences of libraries and archives. Concerning the location, for example, the linking to GPS coordinates today provides a second level of entirely digital spatial allocation. Other issues depend on the day-to-day decisions taken by the support staff, which, in fact, carries responsibility for the consistence of the data material and the quality of the database. There might be images, for example, that are evidently mislabelled, or unlabelled images, the content of which can be clearly identified, since they depict sights such as castles or famous practices such as, for instance, the Munich Beer Festival. Here, the competences of the support staff can make an important contribution. Their knowledge of places and objects depicted in the images might add contemporary meta-text to historical meta-text, thereby extending the epistemological depth of the database—if the work regime provides economic and intellectual leeway to do so.

Yet, three issues remain unsolved, no matter how carefully the transfer of data—visual and textual—proceeds. First, it is impossible to apply source criticism—the bread and butter of the humanities—to a dataset the digital authorship of which is, in fact, unknown. Second, mistakes such as spelling errors occur, and might make datasets undiscoverable in some means of filtering. Third, information already missing in the original corpus cannot be created in the course of digitization. Images sourced during ethnological research or travel are often full of content but short on detailed information; they only note the place or the main motif. Even if the keywords encompass the entire image content, including costumes or means of transport and architecture, in the background, the metadata remains insufficient.

The first two issues underline the necessity of using the analogue corpus and its digital representation in parallel. The third might be helped along if we succeed in utilizing AI and image recognition to interlink such datasets and enhance them beyond their initial information content and handling, as we will discuss in the second part of this paper. Yet, even then, image description and keywording remain crucial elements of meta-texting, and, therefore, of the usability of visual databases.

Keywording as a matter of disciplinary habitus

Keywording is a crucial element of data storage for analogue corpora and for databases of digitized images, both thematically and in respect of image content. While the thematic order is often part of the basic structure of the collection, the keywording of image content is an additional layer that provides significant

shortcuts in the use of digital collections. There is, however, another black box of metadata related to this coding, another point where decision-making on a small level has an important impact on the overall character of meta-data: The system chosen to describe the images depends on the curatorial context of the original collection and the digitizing institution. The keywording itself might be the task of the general support staff or a specific responsibility of the curators. In any case, the ensuing knowledge production links to the disciplinary qualification and *habitus* of the actors and, so, is neither neutral nor easily accessible for source criticism.

Fig. 1 Farm in Trebendorf, about 1890 to 1897, Sorbian Cultural Archive at the Sorbian Institute Bautzen, Karl Schmidt (052420)

To give an example, there are two historical images depicting rural scenes at the end of the nineteenth century, held by the Sorbian Institute in Bautzen (Germany) (Fig. 1 and 2). They show similar environments of residential and farm buildings with actors in local attire going about their daily business in a more or less staged manner. However, the descriptions as transferred into meta-text differ in accordance with the position of the images within the field-specific order of the original collection. While one is titled *construction—farm buildings and farmsteads*, the other reads *traditional costume and hairstyle—folk costume*. Interestingly, the keywording followed the logic of the original allocation of the image title as either building-related or clothing-related, and not necessarily of the overall image content. In the first case, it listed the general construction method (timber construction, block

building) and roofing (reeds) along with architectural details (balcony, gallery) while the clothing was not mentioned. In the other case, the costume is listed but information on the building missing but for the term *farm building*, though the roofing and construction method is the same as before and easily recognizable. Sadly, as the keywords are a reflection of the language of the original titles rather than the actual content, neither the costumes in the first image nor the building construction in the second one are detectable in the visual database.

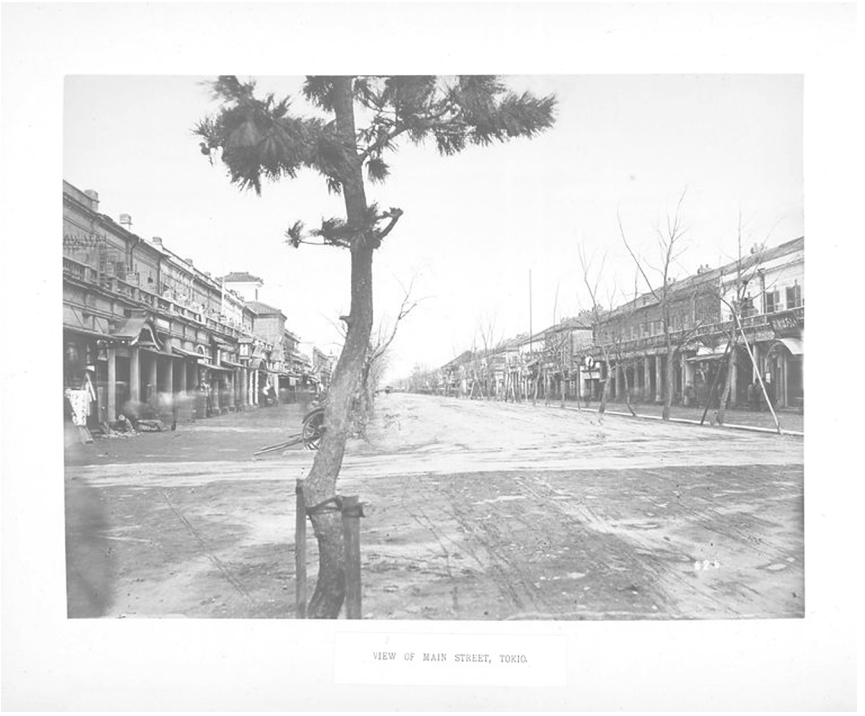
Fig. 2 Farmers at work (in Bórkowy/Burg-Spreewald), 1900, Sorbian Cultural Archive at the Sorbian Institute Bautzen, Steffen (Burg) (053710)

These observations should not be taken as a reproach to any of the actors involved in the digitization but to reflect on the conditions and consequences of meta-texting. They aim to gain a clearer view of the benefits of easy access to formerly hidden source materials in the course of digitization; there are challenges involved, despite—or even because of—digitized visual materials. Digitization makes visual source material accessible and invites one to delve into hitherto unknown or barricaded-off collections. It does neither solve the intrinsic shortcomings of existing analogue corpora, reduce errors, nor free us from the day-to-day business of source criticism. The process of meta-texting is riddled with possibilities of the mistransfer of words and ideas, the results both alluring and imperfect, as with the following example.

The image titled *View of Main Street, Tokio* (Fig. 3) is kept at the New York Public Library (cf. Löffler/Hein/Mager 2018). The metadata does not provide

address, time of day, date, or photographer, as can be observed for many of similar souvenir pictures of the late nineteenth and early twentieth century. The descriptive keywords are consequently sparse: *trees*, *rickshaws*, *row houses*, and *streets*. Yet, the image might be of significant interest for the study of architecture and urban environment of Japan as soon as we succeed to assign the location and to narrow down the period. In this case, experts' cross-references with other digitized holdings makes it possible to recognize Ginza, a famous business district in downtown Tokyo, rebuilt after a fire in 1872 and planted with trees. Even further research unearths a largely identical newspaper photograph dated 1874 (Mainichi Shimbun 1960: 11).

Fig. 3 View of Main Street, Tokio, Still image (albumen print), [Date Unavailable], The New York Public Library, The Miriam and Ira D. Wallach Division of Art, Prints and Photographs: Photography Collection (MFY 96-4255)



This approach, however, asks for a much larger workforce and specialized expertise than most institutions can provide and afford. However, it also leads to the question of whether computer technology is also capable of providing solutions to these problems associated with digitization: Is it possible to use image content recognition as a support to circumvent these weaknesses of metadata, and

to support or replace the many time-consuming minor politics in favour of an overarching technological solution?

Beyond digitizing: automatic image content recognition

It seems that computer technology has the potential to be helpful in solving, or even avoiding, various problems associated with the creation of metadata. The field of computer vision plays a particularly important role here. Computer vision is a branch of artificial intelligence and deals with the understanding of images by computers. In combination with crowdsourcing and linked open data, it will not only be possible to automatically create metadata, but also create references between the content of different images from different collections.

Today, computers can reliably recognize faces or visually understand their surroundings so that they can e.g. steer a car through them. They also outperform humans in detecting cancer cells (Savage 2020). Applied to big data, computer vision can help us study large collections of visual information that reach a global scale. It can contribute to identifying objects or elements, making digital images searchable through content indexing. Researchers from the University of Heidelberg demonstrated the power of computer vision to identify objects and gestures in medieval miniatures (Bell/Schlecht/Ommer 2013). In their argumentation, they point out that this enables a visual scaling of queries that can hardly be defined linguistically and, so, can lead to new research questions and findings (Bell/Ommer 2018: 68). In the meantime, the automated classification of architectural standard elements in images (e.g. windows, doors, or roofs) is well advanced, and aspects such as partial occlusion or perspective distortion are not an obstacle to the assessment of the image content (cf. Nishida/Bousseau/Aliaga 2018; Kapoor/Larco/Kiveris 2019).

However, the automatic generation of metadata remains a challenge (Ioannides/Davies 2017: 176). It is not only necessary here to recognize certain classes of image content (e.g. reeds, gallery, street), but also explicitly identify objects. This means not only recognizing a house or a street, but also providing information about which house and which street it is. In the case of buildings, determining their location can also help one identify them and thereby generate meaningful metadata. The PlaNet model, for example, based on a deep network trained with millions of geotagged images, is able to predict the location of photos comparable to the performance of humans and partially even beyond (cf. Weyand/Kostrikov/Philbin 2016). Where previous work has concentrated on limited subsets, such as certain types of buildings, the availability of street-view images or locations with dense image coverage, PlaNet can locate photographed locations without restriction. However, the result is not a specific geolocalization, but an estimate within a larger region. For a playful competition with AI, users can estimate the location of any image and compare their guess with the result of

the algorithm in an online tool of the Leibniz Information Centre for Science and Technology (Technische Informationsbibliothek).

Artificial intelligence and historic images

Precise identification is being researched by the ArchiMediaL project (TU Delft/ VU Amsterdam/ TU Dortmund). It investigates the use of computer vision to automatically identify buildings in a large number of historical images (ArchiMediaL). This might be particularly useful for accessing great quantities of unannotated images and for identifying previously little-researched architectures. Millions of images featuring built structures are lost to research because the buildings in such a large number of images cannot yet be adequately identified. Unfortunately, most of these structures belong to under-represented architectures outside the canon, and their unavailability also contributes to biases in architectural historiography. In addition, the project examines the vision of the computer itself: What does a computer see in architecture if its perception and processing is not based on human senses and language? How can a computer successfully recognize buildings in images if it knows nothing about columns, windows, or roofs as concepts? ArchiMediaL tracks these questions by applying Grad-Cam technology to multiple convolutional neural networks (CNN), trained to distinguish imagery from different cities, and analyses the resultant heat maps that highlight the most important architectural areas for recognition and differentiation (Shi/Khademi/ van Gemert 2019).

The project's main focus, however, is on the automated identification of buildings in historical photographs. The identification of buildings can be solved by their location: a building has only one address/geolocation. A limiting factor here is that buildings change over time, or are demolished or replaced by other structures. The location can be determined if a computer recognizes that the image content (building) of a historical image is identical to the image content of a geo-referenced image (e.g. from Google Street view or Mapillary). This allows for the identification of the building in the historical image by its location. Here, there is a number of challenges: 1. A suitable algorithm can only be created if there is a sufficiently good data situation. This concerns a large number of historical photographs showing buildings that still exist in the cityscape. 2. The algorithm must be robust against image data from different domains. In contrast to image data for e.g. face recognition, historical photographs are very varied: coloured or black and white, blurred or sharp, taken from a variety of perspectives and with different focal lengths and light situations. 3. A sufficient number (~1,000) of buildings must be recognized by humans in both the historical and the geo-referenced images to generate valid image pairs that serve as a training and verification set for deep learning.

The first challenge limits the application of current AI solutions for image content recognition to collections that meet these criteria. ArchiMediaL has selected the Beeldbank collection of the Amsterdam City Archives (Stadsarchief Amsterdam). Here, on more than 400,000 photographs from the nineteenth and twentieth centuries, buildings of Amsterdam can be found, some of them with geographical information in varying degrees of detail. In most cases, they still exist, and are also visible in the georeferenced images provided on Mapillary. The collection contains image material from a broad variety of sources. Like online image material found through queries, there is hardly an intrinsic logic. The purposes for the creation of the images and their inclusion in subcollections are manifold. It was not a scope of the project to examine their nature but to make use of their diverse nature in respect of the investigation of the built environment of the past.

The second challenge concerns the design of machine learning. For this purpose, ArchiMediaL developed a novel age-invariant feature learning CNN (Wang/Li/Khademi/van Gemert 2019).

The third challenge reconnects AI performance to human knowledge and experience. This is also the entry point for biases and prejudices that are embedded in human thinking. These shortcomings can find their way into AI solutions, as research shows (Koene 2017, ALGB-WG 2017). Specifically, this means that when creating the training sets, not only will errors be incorporated into the training of the CNNs, but, furthermore, only the knowledge that has been acquired beforehand can be integrated. Here, dominant canons, professional habits, or prevailing cultural perceptions are of great weight, which makes it clear that the use of AI can hardly be considered a truly objective method. Hence, acknowledgement of the minor politics in the processes of digitization are vital. ArchiMediaL's strategy of harvesting the human knowledge required to build the training set was focused on crowdsourcing. This made it possible to get input from a variety of people with different cultural and educational backgrounds. This is because we were able to harvest image-specific knowledge provided by people with many different backgrounds—for example, architecture students, historians, computer specialists, as also local residents and interested lay people. They all have a different approach to urban scenes captured in the image. An online tool enabled invited users to view a historical picture and a Mapillary street view from a nearby area on a split screen (Fig. 4). The Mapillary screen allows for navigation along the streets and camera pan and tilt. The user's task is to navigate the camera to a position that roughly corresponds to the historical image on the other side of the screen. In this way, the location of the building can be captured using Mapillary's geodata, which enables identification, and the resultant image pair becomes usable for CNN training. Moreover, this changes the abovementioned role of the support staff, as now large and diverse crowds contribute to the extraction of knowledge from image material. They still operate within a frame of requested details but are able to contribute their own observations and interest via a response form.

In order to keep the resultant errors to a minimum, it was necessary to verify the results individually. An architectural historian reviewed the identified image pairs and also verified possible pre-set comments (e.g. building or parts of it covered/building removed/building added/building not accessible/no street view scene). It became clear that many situations are ambiguous: participants with little architectural knowledge have confused similar buildings with one another, experts have recognized completely changed structures on the basis of neighbouring buildings, residents have verified the correct location without the structure still being there, and so on. These and other cases led to matches that were useless for training the algorithm. But they have provided two crucial insights. On the one hand, they underscored the complexity of human image recognition and related knowledge management. On the other hand, they showed the need to describe the task precisely in order to give very different participants the opportunity to contribute meaningful results. Scholars of the humanities will need to invest a lot of time and supervision in preparing new technologies if they expect these to lead to useful methods. However, in this case, the investment in basic research will pay off when the automated recognition of content in millions of images becomes a reality, which is already a vision within reach.

Fig. 4 A screenshot from the online annotation tool showing a historical image from the Beeldbank collection (middle), the same building today in navigable Mapillary street-level imagery (left), and the locations of the building and camera on the map of Amsterdam. Ronald Siebes: ArchiMediaL annotation tool (http://archimedial.eu/beeldbank/marker-clustering-geojson2.php)

Towards alternative metadata for architecture

Without analysing the process and the results in detail, it becomes clear that research into computer methods for generating metadata for image material requires both profound knowledge of the humanities and visual sciences as well as intensive and smooth communications between the scientists involved in the various disciplines. Instead of just adding digital aspects to current research or using computer scientists as contributors to solve IT-related problems, it is crucial to further develop mixed method approaches. On the one hand, this does justice to the common roots of humanities and science (Mager/Hein 2019). On the other hand, it is the best way to secure a place for the interest of the humanities in the technology-driven development of the future: Computer science, with all its funding opportunities both from science organisations and industry, will continue to develop without great dependence on the interests of the humanities and state-funded organizations like universities (Shamir 2020). If the humanities want to participate in this development and are interested in the applicability of future IT methods for their own research, they must actively participate in this development. This can only happen if, for example, architectural historians and IT scientists develop common research interests or research questions and challenges, with benefits for both disciplines. Progressive mixed-method research can arouse further interest from computer science to contribute to historical and cultural research and supports the creation of a basis for a future-proof orientation of the humanities. In the case of ArchiMediaL, the training of CNN is currently underway, and initial results still show low accuracy. Yet, they demonstrate that the approach is working.

What would it mean if AI was able to recognize the same building in different images from different eras?

It would then be possible to reconstruct the visual representation of architecture over time and locate even less-researched, or not-yet-unidentified, sites worldwide. It could also become a tool for questioning structural parts of metatexts—illuminating these minor politics—by pointing out epistemological connections that we are not yet aware of. Concerning the issues of keywording as disciplinary habitus, as outlined above, the identification of image content based on visual comparison raises questions about the interconnection of image and text, and leads to considerations about the character of metadata, which are 'strictly related to semantics' (Ioannides 2017: 178). When novel technologies allow more direct access to visual information, such as for comparing image content or visually defined queries (e.g. Bell/Ommer 2018, p. 68), this bypasses textual information/metatags to the extent that meanings of image content do not have to be explicitly formulated in order to establish references or relationships with other visual content(s). Here, linguistic inadequacies resulting from the translation of visual into textual information can be circumvented and new connections created. The designation of the information in the image is not restricted any longer to the

interests of a specific discipline, as visually defined queries operate independently of semantic restrictions. On the other hand, this possibility also entails the risk of a loss of control if things no longer have to be named precisely, and are rather associatively linked, based on visual congruences. The automatically created associations and relations could also slip into the realm of decisions by artificial intelligence—because of the sheer number of objects, it might be difficult to keep control on this process, which could lead to another black box.

Conclusion

AI-based image content recognition has the potential to establish novel interlinkages between image material. It can provide text-based mechanisms and their partly known, partly unknown weaknesses with an alternative and seemingly neutral data access that bypasses the meta-text and ultimately enriches it at the same time. The neutrality of technology—which requires a lot of training data that is as unbiased as possible and must, therefore, be highly diverse—could, in contrast to the acceptance of subjectivity in the humanities, be something like a guide behind the narratives of metadata.

Which of the issues of minor politics outlined above will this alleviate, which will it worsen, and what are the new ones that could emerge?

Rather than keeping and reinstalling the intrinsic logic of the original collections and their digital representation, AI may find and propose novel logics that even interconnect multiple collections across disciplinary borders. The potential to recognize similarities and make classifications between millions of visual objects goes far beyond human capabilities. AI may not be very helpful in exploring small collections and the personal interests of their collectors. But it can help make these small and rather unknown collections accessible to a worldwide audience of researchers and give the material of these collections a meaningful and complementary place in the global visual sphere.

In the course of digitization and data transfer, support staff will continue to play a role. Even if computers seem able to compare and analyse without being biased and without having to fall back on hegemonic connections and categories, they are still made by humans. This means that biases embedded in our communications and perceptions of the world become part of the algorithms and artificial intelligences that we create (Hao 2019). The sheer size of the datasets used to train these intelligences can be useful in overcoming bias, as data from very different sources can be used. Nevertheless, there is a need for research in the humanities that critically reflects on and questions the way new knowledge is produced and that accompanies every step of the process—even beyond its own original interests.

The use of AI to investigate image content has very different disciplinary reasons—even within the humanities. Hence, the applied solutions and algo-

rithms are not independent of the disciplinary habitus. Rather, they depend on the epistemological interest of the subject. A strong contribution to the humanities can be seen in the possibility of establishing cross-disciplinary connections between all kinds of visual objects, including objects that might be new as objects of interest for certain disciplines. Spatial and cultural boundaries do not initially play a role in this observation of objects or their integration: ultimately, the global stock of visual material can become the subject of investigation for research that is open to an expansion of its disciplinary horizon.

Bibliography

ALGB-WG—Algorithmic Bias Working Group (2017): "P7003—Algorithmic Bias Considerations", September 29, 2020 (URL: https://standards.ieee.org/project/7003.html).

ArchiMediaL: "Enriching and linking historical architectural and urban image collections", September 5, 2020 (http://archimedial.eu).

Bell, Peter/ Björn Ommer (2018): "Computer Vision und Kunstgeschichte – Dialog zweier Bildwissenschaften", In: Kuroczyński, Piotr, Peter Bell, Lisa Dieckmann (eds.), Computing Art Reader – Einführung in die digitale Kunstgeschichte, Heidelberg: arthistoricum.net, pp. 61-75.

Bell, Peter/ Joseph Schlecht/ Björn Ommer (2013): "Nonverbal Communication in Medieval Illustrations Revisited by Computer Vision and Art History", In: Visual Resources: An International Journal of Documentation 29/1-2, pp. 26-37.

Hao, Karen (2019): "This is how AI bias really happens—and why it's so hard to fix", In: MIT Technology Review, 4 February 2019, URL: https://www.technologyreview.com/s/612876/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/ Access: 30 September 2020.

Inoue, Shuhei (2010): „Strategien gegen Kulturverlust durch Katastrophen in Deutschland und Japan: Das Historische Archiv der Stadt Köln und das Shiryo-Net", In: Japanisch-Deutsches Zentrum Berlin (ed.), 3. Deutsch-japanisch-koreanisches Stipendiatenseminar, 2.-3.10.2009, Berlin: JDZB, pp. 99-108.

Ioannides, Marinos/ Davies, Rob et al. (2017) "3D Digital Libraries and Their Contribution in the Documentation of the Past", In: Ioannides, Marinos/ Magnenat-Thalmann, Nadia / Papagiannakis, George, "Mixed Reality and Gamification for Cultural Heritage", Cham: Springer, pp. 161-199.

Kapoor, Amol/ Hunter Larco/ Raimondas Kiveris (2019): "Nostalgin: Extracting 3D City Models from Historical Image Data", In: arXiv:1905.01772v1 Access: 28 September 2020.

Kellner, Marcel (2018): Die Musealisierung einer Privatsammlung. Provenienzforschung zur Sammlung Alexander Dolezaleks am Deutschen Historischen

Museum (DHM), presentation at the conference *collecting loss*, Weimar, 16.11.2018.

Kohle, Hubertus/ Locher, Hubert: "The Digital Image", Priority Programm of the German Research Foundation, https://www.digitalesbild.gwi.uni-muenchen.de/, 2019, Access: 28 September 2020.

Koene, Angar (2017): "Algorithmic bias: Addressing growing concerns". In: IEEE Technology and Society Magazine 36(2), pp. 31–32.

Löffler, Beate/ Hein, Carola/ Mager, Tino (2018): "Searching for Meiji-Tokyo. Heterogeneous Visual Media and the Turn to Global Urban History, Digitalization, and Deep Learning", blog entry, 20. March 2018, https://globalurbanhistory.com/2018/03/20/searching-for-meiji-tokyo-heterogeneous-visual-media-and-the-turn-to-global-urban-history-digitalization-and-deep-learning/#more-4038 [18.09.2020]

Mager, Tino/ Hein, Carola (2019): "Mathematics and/as Humanities – Linking humanistic historical to quantitative approaches", In: D'Acci, Luca (ed.): "The Mathematics of Urban Morphology", Basel: Birkhäuser.

Mainichi Shimbun (ed.) (1960): Nihon no hyaku nen. Shashin de miru fuzoku bunkashi, Tokyo: Mainichi Shimbun.

Matyssek, Angela (2009): „Der Verlust der Spur. Über die Halbwertszeiten in Bildarchiven", presentation at the conference *Depot und Plattform. Bildarchive im post-fotografischen Zeitalter*, Köln, 05.06.2009

Münster, Sander/ Terras, Melissa: "The visual side of digital humanities: a survey on topics, researchers, and epistemic cultures", In: Digital Scholarship in the Humanities, fqz022, 5 May 2019, p. 2 f., URL: https://doi.org/10.1093/llc/fqz022 Access: 1 October 2020.

Nishida, Gen/ Adrien Bousseau/ Daniel G. Aliaga (2018): "Procedural Modeling of a Building from a Single Image", In: Computer Graphics Forum 37/2, pp. 415-429.

Prescott, Andrew (2015): "The Future and the Digital Humanities", In: Medium, 27 July 2005, URL: https://medium.com/digital-riffs/the-future-and-the-digital-humanities-6c6b3f8a3295 Access: 1 October 2020.

Ruhl, Carsten (2014): "Autobiographie und ästhetische Erfahrung. John Soanes Künstlerhaus in Lincoln's Inn Fields", In: Salvatore Pisani; Elisabeth Oy-Marra (ed.), Ein Haus wie Ich. Die gebaute Autobiographie in der Moderne, Bielefeld: transcript, pp. 129-156.

Savage, Neil (2020): "How AI is improving cancer diagnostics – Artificial intelligence can spot subtle patterns that can easily be missed by humans." In: Nature, 15 March 2020. URL: https://www.nature.com/articles/d41586-020-00847-2 Access: 27 September 2020.

Shamir, Lior (2020): "A Case Against the STEM Rush." In: Inside Higher Ed, 3 February 2020. URL: https://www.insidehighered.com/views/2020/02/03/computer-scientist-urges-more-support-humanities-opinion Access: 15 January 2021.

Shi, Xiangwei/ Khademi, Seyran/ van Gemert, Jan (2019): "Deep Visual City Recognition Visualization", In: arXiv:1905.01932, 6 May 2019, URL: http://archimedial.net/wp-content/uploads/2019/06/1905.01932v1.pdf Access: 28 September 2020.

Stadsarchief Amsterdam: "Beeldbank", September 28, 2020. (https://archief.amsterdam/beeldbank/?mode=gallery&view=horizontal&sort=random%7B1601407054472%7D%20asc).

Technische Informationsbibliothek: "Geolocation Estimation", September 7, 2020 (https://labs.tib.eu/geoestimation/).

Visual Narrative Initiative: "Urban Panorama", September 27, 2020 (https://www.visualnarrative.ncsu.edu/projects/urban-panorama/).

Wang, Ziqi/ Li, Jiahui/ Khademi, Seyran/ van Gemert, Jan (2019): "Attention-Aware Age-Agnostic Visual Place Recognition", in: arXiv:1909.05163v1, URL: https://arxiv.org/pdf/1909.05163.pdf Access: 28 September 2020.

Weyand, Tobias/ Kostrikov, Ilya/ Philbin, James (2016): "PlaNet - Photo Geolocation with Convolutional Neural Networks", In: arXiv:1602.05314, URL: https://arxiv.org/abs/1602.05314 Access: 25 August 2020.