

Metadata generation for intangible heritage: the role of crowdsourcing and AI to address bias risk

Authors:

- Valentina Bachi (Photoconsortium)
- Rosa Cisneros (Coventry University)
- Antonella Fresa (Photoconsortium)
- Allison Kupietzky (The Israel Museum, Jerusalem)
- Alexandru Stan (IN2 Digital Innovations)
- Frederik Truyen (KU Leuven)

1. INTRODUCTION

While tangible heritage has a compelling message of its own, as it is a visible and often imposing remnant of past cultures, intangible heritage is often, when it is not actively lived through persistent traditions, buried in documents and testimonials typically held in archives. The problem of the representation of the intangible heritage of communities raises many issues for the archivist, particularly when minoritised communities are involved. First of all, the notion of “representation” deserves scrutiny, even if it does not seem to have a direct relation to the mission of an archive. How does it happen that many discussions are taking place in the context of archives when it relates to minority communities? What do we mean by “representation” in this context and why is this an important issue in today’s archival research? At face value, the original documents in an archive are supposed to speak for themselves: safeguarded, unaltered, untampered with, uncensored. People trust the document in an archive to be the original, to be complete, so what is the problem? Unfortunately, history has already learned that sometimes this is not the case, that archives have been pressured to censor or make documents unavailable. But that is not a new issue. Everything changed with the digital era, with the digital copy, and the idea that archives should provide open access to their contents in the digital space. In the digital world, original documents are *represented* by a digital copy - and so here we are with this “alien” notion of representation. Because by all means, the digital copy *cannot be* the “untampered” original. It will never be. When archives turn an archive fonds into a digital collection, it becomes a *representation* not only of the documents involved but at the same time of the people, places, practices and communities involved. It becomes a window to a past reality. While normally a historian is situated in between the public and the archive, now the digitised archival collection is directly accessible. It starts to tell its own story, its own narrative. This is compounded by the fact that you cannot just digitise a document and

put it online: you need to *describe* it first, at the bare minimum by adding metadata that makes retrieval and discovery easier.

And here's the catch. These metadata have been added through the times at different moments, often in different eras than the original document. So suddenly we are faced with a multitude of often conflicting narratives.

But the problems run deeper. Archives have an officialized mission, a government remit. They are bound to safeguard official documents, and also receive private donations. Further, storage space is limited, and so is processing time. This means there is considerable bias in what is kept and processed in the archive. This bias is both partly by design, namely dictated by the government remit, but also partly by circumstance, e.g. available workforce. But also access to citizens can be different, in particular in regard to minorities. In the worst case, the minorities are just ignored in the archival process, and so are not "represented" in the collections. In many other cases, they are misrepresented, or only mentioned in specific, often negative contexts. Compounding all these problems is the sheer fact that giving open access on the Internet to collections for a wide audience, which is the essence of Open Access, opens many different perspectives on the very same documents at the reception side so that messages in documents can get "lost in translation".

But "representation" can also have an active meaning: people or communities can be "represented" by others. So it is not only the "representation" of the cultural practices and values of the communities in the published documents and their metadata which is at stake, but also if these communities are somehow "represented" in the internal organisation of the heritage institution by *representatives*. This is, in particular, an issue for communities that are not directly represented by the government which supervises the heritage institution, such as e.g. the Gypsy Roma Traveller (GRT) community.

To summarise, we have identified, in regard to minority representation of intangible heritage, problems with:

- the selective process of what has entered the archive
- the priority given to the processing
- the description of the contents in the metadata
- the history of these descriptions through different time frames and social contexts
- opening up the contents in an accessible and understandable way to a wide audience
- connection with representatives of minority communities.

Photoconsortium, as a multidisciplinary association focussing on photographic documentary heritage, tries to find approaches that can offer professional solutions to these issues, through the initiation of several targeted projects. The WEAVE project¹ has the ambition to improve

¹ WEAVE Project: <https://weave-culture.eu/>

The project aims to enrich Europeana (<http://www.europeana.eu>) - the European portal for Cultural Heritage - with high-quality audiovisual and 3D content of tangible and intangible heritage coming from a wide range of European cultural communities. New digital tools will facilitate the aggregation, sharing and showcasing of the content online, while a strand of capacity-building activities will support the digital transformation of Cultural Heritage institutions as well as the know-how for engaging with diverse communities and their heritage. The project is co-financed by the European Union under the Connecting Europe Facility (CEF) Programme.

capacity-building in heritage institutions on the handling and description of intangible heritage, in particular in the case of minority representation.

In this paper, we will derive lessons learned from work on metadata enrichment using both AI as well as crowdsourcing in a complementary strategy, focusing on archival collections. While AI approaches are unavoidable to be able to improve metadata at a sufficient scale, crowdsourcing through annotation tools proves crucial to validate the obtained results. More importantly, involving stakeholders and representative organisations in this process is important as well as necessary to cope with contextual bias and misrepresentation, for which automated mappings cannot easily be implemented. To this regard, the WEAVE project demonstrates best practice approaches that were tested and validated during the project.

In this paper, we will discuss:

- Participatory approaches for inclusiveness and diversity
 - WEAVE
 - REACH
 - INCULTUM
- Metadata enrichment based on crowdsourcing annotation
- Multilingual and Multicultural Biases in Artificial Intelligence
- Concluding recommendations

2. PARTICIPATORY APPROACHES FOR INCLUSIVENESS AND DIVERSITY

This section is articulated on the analysis of three complementary experiences that aim to illustrate how different forms of participatory approaches can enrich the digital cultural content in terms of *representation* of inclusiveness and diversity aspects.

Participation comes in many ways. It can originate in institutional initiatives or be born in community actions. It may involve a variety of beneficiaries, from large audiences to small and specific groups of citizens and stakeholders. Participation can use digital platforms or can take place in physical environments. However, all modes of participation in Cultural Heritage demonstrate the increasing interest in democratising access to culture and, especially in our post-pandemic situation, to open up the fruition, management, preservation and *interpretation* of heritage to ensure an active and effective collaboration of communities, neighbourhoods and individuals.

Three case studies are discussed as complementary ways to cope with the need for participation and how it can be realised in different contexts, namely the case of WEAVE for the use and re-use of Europeana's content, the case of REACH for the creation of a social platform focused on participatory approaches and social innovation in culture, and the case of INCULTUM for the participation of local communities in the development of sustainable cultural tourism.

WEAVE Labdays and Europeana events

The core activities of EU-funded project WEAVE were varied, but for the purpose of this paper we will focus on the process of aggregating high-quality collections to Europeana and on efforts of providing new ways for users to explore the content, e.g. through virtual exhibitions among other formats. Intangible Cultural Heritage (ICH) is of deep importance to several communities from differing socio-economic backgrounds, especially those which are at the fringes of society and at risk of being excluded.

To showcase a variety of relevant intangible heritage content, the project aggregated and published in Europeana over 5,000 new high-quality records that depict the rich and invaluable Cultural Heritage of minority cultural communities, and showcase these collections in a set of engaging editorials and a virtual exhibition². For the scope of this paper we will focus on the dance and GRT-related content aggregated by Coventry University's Centre for Dance Research (UK), as an entry point to discuss the LabDay methodology and its connection to the content collections. By explicitly selecting collections in the project that deal with such communities (e.g. Romani, Gypsy and Traveller), we aimed to increase the inclusiveness of the digital Cultural Heritage and of Europeana, where GRT heritage was rather underrepresented. In addition, the content provided by the Early Dance Circle (EDC) and other dance artists based in the UK and with a strong connection to the Centre for Dance Research, includes a demographic that is varied: in particular, the EDC works mainly with 65+ years of age adults that have different mobility and mental health and wellbeing issues. The initiative envisaged a two-pronged approach of curating and generating content and disseminating to Roma and non-Roma, to educate, promote and celebrate the communities' arts and culture. The inclusion of the co-created Yellow Couch Convos podcast series³ offered a form of collective and creative activism which allowed communities that are outside of mainstream ICH contexts to enter into conversation and directly contribute to the outputs of the project, thus enhancing the content available via Europeana. However, this process is neither straightforward nor linear: it requires a methodology to explore how tangible and intangible Cultural Heritage can be made accessible to citizens from a varied background. This end goal - to make content findable, searchable and respectfully representative of the communities being included under the WEAVE umbrella - aims to make ICH across Europe more accessible to its citizens and provide marginalised communities a sure way to preserve their rich and multi-faceted Cultural Heritage. This process that the WEAVE project underwent relied on presenting information to those outside the community, leading to a better understanding of the complex fabric that makes up our shared European identity.

The WEAVE White Paper⁴ investigates how new relationships between the tangible heritage mediated by institutions and the intangible, "unmediated" heritage that forms its *raison d'être* can be restored and strengthened. While intangible heritage is often alive and kicking in

² The content providers who worked in WEAVE include ARCTUR (Slovenia), CRDI (Catalonia), COVUNI (Centre for Dance Research) (UK), the Early Dance Circle (UK), ERIAC (Germany), PédeXumbo (Portugal) and TopFoto (UK).

³ Yellow Couch Convos Series: <https://soundcloud.com/user-566749993/sets/eriac-yellow-couch-convos> The project is discussed within a forthcoming publication for the Sonic Engagement book (2022) edited by Sarah Woodland and Wolfgang Vachon for Routledge. *Yellow Couch Convos Podcast series: Navigating identity politics through collective voices and counternarratives*, By Rosemary 'Rosa' Cisneros

⁴ <https://drive.google.com/file/d/1DdA6Dz9xbmsdWwxRCRBNhjnxbWNOpryN/view?usp=sharing>

bottom-up citizen organisations celebrating their culture, there is a role for archives to play in nurturing the connection between their collections and the cultural practitioners and practices who help make the archival holdings heritage in the first place. This is particularly important when this relation has been historically lost or distorted. As the White Paper points out, the digital offers a lot of possibilities; however, “the mere fact of placing collections online does not necessarily lead to deeper connections with and between diverse communities”⁵. For this curation and contextualisation of the digital content through storytelling is key and for this reason one of the digital tools made available by the project focuses on these aspects. Using WEAVEx⁶ people can mix digital content that they own with existing content from Europeana, to explore topics from a personal perspective and share their stories and interpretations digitally.

The WEAVE LabDay framework informing the LabDay capacity-building events with communities is underpinned by Communicative Methodology, a sociological method that aims to cross social, cultural and linguistic boundaries. This framework enables an open, egalitarian dialogue between researchers and participants; it is a collaboratively-held ‘brave space’ where all voices are acknowledged and valued, and stakeholders can together reflect on their needs, desires and various forms of participation. This particular methodology enabled communities to engage with project activities and select content and collections to be aggregated. In such a way, the bottom-up approach enabled cultural communities to themselves become a driver for how their digital heritage is presented and the design of the WEAVE Toolkit⁷, developing from their bespoke needs concerning the management and promotion of both their intangible and tangible heritage.

As the WEAVE project aimed to investigate the representation and transformation of intangible Cultural Heritage (ICH) through digitisation, the consortium also examined the ethics and responsibilities of content providers to communities, individuals and other stakeholders, paying special attention to the issues of content selection, identification and description and including sensitive topics such as the relation to identity politics and the issues of (virtual) repatriation/restitution. It was thus important to also help CH professionals develop their known-how on the topic of inclusion. Moderated by external experts in Diversity & Inclusion participative events were organised, giving the CH professionals the chance to learn through fictional case studies and open discussions⁸. Additionally, community members and professionals were both involved in LabDays which were open or closed events, depending on the specific needs of the cultural communities participating. The recordings were placed online and sit on the WEAVE website⁹ alongside other materials produced for the LabDays. Also readily available on the project’s website are a series of documents that further explore the

⁵ WEAVE White Paper, p. 9

⁶ <https://experience.weave-culture.eu/>

⁷ WEAVE also delivered several digital tools to support the documentation, management and sharing of digital content: <https://weave-culture.eu/weavetoolkit/>

⁸ For more information on these events see: <https://weave-culture.eu/capacity-building/europeana-events/>

⁹ WEAVE LabDay recordings available online: <https://weave-culture.eu/labdays/>

tensions and questions around representation in relation to the cultural communities involved in the project (see WEAVE documents¹⁰).

The REACH Project

REACH¹¹ is the Social Platform for participatory approaches and social innovation in culture. The project received funding from the European Union's Horizon 2020 research and innovation programme, running for three years (2017-2020). The project developed an integrated model for resilient European Cultural Heritage *milieux*. By Cultural Heritage *milieu* we mean a unit of synergy between different components, a place of memory, not necessarily geographical or physical, capable of linking tangible, intangible and natural cultural aspects together with people. In the ambit of their relationship with connected communities, Cultural Heritage *milieux* may be suitable for resilience, adaptation and survival, revealing a capacity to renew and reorganise themselves after 'disturbance'. They can offer a mitigation against the risks connected with the management of change and inspire strategies to cope with the consequences of social and economic development. Cultural Heritage regimes are characterised nowadays by widening the definition of what counts as 'heritage'. In this light, participatory approaches become a key factor for Cultural Heritage to react to fast-changing surrounding environments. With this aim, the REACH Social Platform established a space for meeting, discussion and collaborations that addressed the question of participation in culture ranging from four complementary definitions of 'heritage': small towns' heritage, rural heritage, minority heritage and institutional heritage. For each cultural definition, a series of pilot experiences were organised by the partners in many European contexts, representing the basis for further exploitation. The project produced methodological instruments and practical examples of good practices made available to Cultural Heritage managers, administrators and civic associations to decipher how to cope with changes via participation.

INCULTUM: participatory approaches to sustainable Cultural Heritage and tourism

INCULTUM¹² is the Innovation Action also funded by the European Union Horizon 2020 Programme. The project started in 2021 and will run until 2024. Its focus is to experiment with, and to validate, novel approaches to urban and regional development through cultural tourism and participation of local communities. Connecting tourism, Cultural Heritage and participation becomes a way of furthering sustainable social, cultural and economic development of lower-rated areas. The project aims to explore how to unlock the full potential of marginal and peripheral areas when managed by local communities and stakeholders.

¹⁰WEAVE Resources open access materials (e.g White Paper, LabDay Framework paper) <https://weave-culture.eu/capacity-building/resources/>

¹¹ <https://www.reach-culture.eu/>

¹² <https://incultum.eu/>

Tourism is more than travelling and consuming: it is a way to learn and improve oneself, to enrich the vision of the world and to improve the understanding of the others; it is a discovery that satisfies our inherent curiosity and, in some ways, provides us with a romantic sense of adventure, the need to explore and go beyond. From the economic point of view, tourism is one of the most important ‘industries’ in the world, with impacts on culture, environment and social relationships. However, tourism can generate negative impacts to residents, at different levels: touristification is one of the negative consequences in many historical cities; widespread growth and gathering of people cause problems for the local population, such as gentrification, insecurity of employment, social changes, and massive urbanisation. Additionally, touristification reduces the quality of the visitor experience: consumerism generates critical offers in rural areas and natural environments; visiting crowded monuments and historical city centres result in long waiting times. Participation in culture is the key factor of innovation that INCULTUM is experimenting in ten pilot experiences, where local communities are transformed into protagonists, adopting digital and engaging solutions. The selection of the pilot sites to be included in the innovation experiment was based on prioritising deprived, remote, peripheral and deindustrialized areas or cultural-natural heritage not usually taken into account by the mainstream tourism processes. Three levels of actions are validated in the INCULTUM project. The principal validation takes place at the local/regional level related to the ten INCULTUM Pilots, in the framework of the activities undertaken therein. Secondly, the project encourages a cross-fertilisation between the Pilots. This means that – whenever possible – each Pilot serves as the first testing ground of the solutions that the other pilots implement. Naturally, such cross-fertilisation is based on a relevance criterion, even if the partners expect that many aspects of innovation could be relevant to various other pilots. Thirdly, the pilot solutions are spread among the wider network of interested stakeholders created through a dedicated networking activity that is carried out in the project. These links will facilitate re-use of good practices and duplication of experiences in new geographical contexts, preparing for further replications in other places beyond the ten locations of the INCULTUM pilots and translated into strategies and policies. Project’s findings and results are shared through the project’s blog¹³ hosted on digitalmeetsculture magazine. A programme of open workshops, a final conference in Granada, the project’s website and a book target the project’s audiences, addressing stakeholders in the tourism and in the Cultural Heritage sectors, as well as the local communities and stakeholders.

3. CROWDSOURCING FOR METADATA IMPROVEMENT

The work for archives and Cultural Heritage institutions in general to maintain and always improve their collections and metadata is a big effort and a demanding task, that requires careful planning, specific knowledge and available resources. A help to institutions in this effort can come by leveraging the knowledge and participation of citizens, students, culture lovers who are willing to collaborate in preserving and maintaining Cultural Heritage collections. Digital tools can support participatory events with working groups of volunteers

¹³ <https://www.digitalmeetsculture.net/projects/incultum-blog/>

that improve and enrich existing metadata, thus enabling crowdsourcing and citizen science to play a role in Cultural Heritage curation.

For example, in the context of Erasmus+ project CitizenHeritage¹⁴, students in Bulgaria and Hungary were involved in updating metadata of Europeana collections by using the CrowdHeritage Tool¹⁵. CitizenHeritage is a citizen science project which aims to bring together Higher Education and Heritage institutions to collaborate on the engagement of citizens in research. CrowdHeritage, the open platform developed by NTUA National Technical University of Athens, allows for the importing of collections from Europeana and enables users to add descriptive tags selected from controlled lists derived from existing thesauri such as Getty AAT, Geonames and Wikidata. The process is easy and user friendly and is well suited to running crowdsourcing annotation campaigns on existing Europeana data, using controlled vocabularies that enrich the existing metadata. In 2021, Photoconsortium realized two metadata crowdsourcing campaigns with students of Sofia and Budapest universities, in collaboration with Europeana content providers NALIS (Bulgaria) and OSZK (Hungary), which are both Photoconsortium members.

From controlled lists of subjects and places prepared by NALIS and OSZK staff, students selected relevant keywords to describe the collection content items. In such a way, the students were introduced to Europeana and gained an understanding in how the process of aggregation and enrichment of digital collections works. For CitizenHeritage, it showcases that crowdsourcing can be a relatively easy way to enable citizen participation in digital Cultural Heritage co-creation and research. For Photoconsortium, being an accredited aggregator for Europeana, it was a test case to demonstrate how students of Cultural Heritage or media courses can do much to improve the descriptions and searchability of records in online collections and particularly to improve the quality of legacy records in Europeana.

This exercise was followed by a two-step validation action coordinated by Photoconsortium and Europeana in the context of DSI4, the EU-funded project for the maintenance of the Europeana platform. In this framework, the annotations collected from Bulgarian and Hungarian students were carefully reviewed, validated and prepared by Photoconsortium for aggregation and republication in Europeana.

The validation process unfolds in two steps: the first step is itself participatory as any person who adds annotation can also upvote (or downvote if applicable) the annotations created by others, in a peer-review approach; the second step of validation is performed by reviewers from Photoconsortium, who review and can accept or reject each single annotation. The entire process is designed to obtain the highest confidence level of the the crowdsourced annotations, which derives from: engaging curators/experts from content providers who know the collection's content well; engaging students of cultural or media courses about the local Cultural Heritage, who are prepared on the subjects of the collection, due to their studies; and

¹⁴ <https://www.citizenheritage.eu/>

¹⁵ <https://crowdheritage.eu/en>

finally a ‘human-in-the-loop’, two-step validation approach to review the enrichments and prepare for re-publication in Europeana.

4. MULTILINGUAL AND MULTICULTURAL BIASES IN ARTIFICIAL INTELLIGENCE

Cataloguing large collections is a time consuming activity. In this section we examine if and how much AI can assist in museum and archive cataloguing. Clearly there are many benefits to automating the process of cataloguing large collections, however, are there hidden biases in AI?

The narrative of textual descriptions for images are contextualised by the historic period when they were created. These textual descriptions are often in conflict with the current narrative for the same image: for example, where a minority was textually identified by its “coloniser”, who wrote labels for images presenting a bias about the people portrayed.¹⁶ A similar concern has been voiced with regard to how AI describes images using texts. Is it possible to identify bias with regard to AI catalogued images?

To answer these questions, research was conducted at the Israel Museum and its finding presented at Photoconsortium and Citizen Heritage Conference “The role of photographic heritage in empowering communities participation in Cultural Heritage”¹⁷. Multiple AI tools were chosen to be tested and the metadata generated by AI was analysed to see how they actually *represent* the visual content. The tools tested were:

- Google Images – OCR, search and retrieval, AI keywording and cataloguing, searching word alternates and automatic translation (including Hebrew), AI facial recognition of artist and other prominent persons, visual recognition of objects and subjects
- Google Lens - Translation of texts
- Adobe Acrobat Pro / OCR – transcription of images of printed books, journals, pamphlets
- Europeana transcription tool - using manual transcription and crowd-sourcing
- Cambridge UK Transkribus – AI transcription uses training tools for handwriting (HTR) found in manuscripts
- PI video transcription tool - Transcribes video audio, creates links of keywords in WikiData, suggests relevant materials based on keywords
- YouTube - automated transcription
- Every Pixel API - AI matches faces and identifies if same person and their age
- MyHeritage / Deep Nostalgia [D-ID] - Animation and facial recognition

¹⁶ John Balean (TopFoto). “Leveraging photographic heritage to support community engagement and virtual reappropriation of heritage”. *The Importance of Context*, Photoconsortium and Citizen Heritage Conference, Pisa, June 2022. [PDF](#). The presentation recommended that when re-writing labels written in the past for images portraying minorities, it is advisable to work with community members and have their input on the images to correct biased descriptions.

¹⁷ Pisa, June 2022, “Cataloguing large photo collections – is artificial intelligence unbiased?” [PDF](#).

Testing AI for museum and archival cataloguing purposes includes:

- Visual recognition – objects, subjects
- Facial recognition – artists, staff, other prominent persons
- OCR text recognition – typed archival material
- HTR (Handwritten text recognition) – Archival letters
- Number recognition in images – registration numbers
- Audio to text – transcription of videos
- Search and retrieval – semantics (Linked Open Data)

AI cataloguing methods were examined to determine how they can help catalogue large art and photographic collections found in archives and museums. Within the tests, in a Cultural Heritage environment, it was found that there were successes and failures including an interesting discovery – the use of current popular AI tools risks perpetuating biases.

The research findings discovered that, for visual recognition, the metadata created by AI was mostly correct for objects and subjects. Regarding number recognition, it was quite accurate. For audio text and transcription, when languages other than English were found in the image, the AI had more difficulty transcribing the texts. For search and retrieval, terminologies are mostly found for Linked Open Data in English and this created a gap when searching with other languages. When testing Google Lens in a museum environment and “viewing” labels in the showcases, AI was capable of translating the language from Italian to English allowing the English-speaking viewer the ability to “read” the Italian label. When working with Adobe OCR on old books, journals and pamphlets, there were many mistakes when dealing with languages other than English. When using automated transcription tools from YouTube, the languages automatically generating the texts were limited. When using Google Image to identify text within an image in a language other than English the tool was capable of searching for a word in Hebrew and finding the alternate translation in English and vice versa.

In conclusion, images catalogued using AI cataloguing methods were more easily retrieved than images without textual labels. The AI metadata included object names and subjects as well as identified persons. A positive impact of AI was the ability of the computer that was “taught” a face to review millions of images and find the same face among multiple archives and mediums. Most impressive was an artist's face that was found in a painted self-portrait as well as in photographs of him individually and within a group of people, even though the artist was of different ages. Best results were found after a staff member reviewed and made corrections where needed.

However, the computer is only as smart or as biased as the person who trained it and therefore thought should be given to updating AI cataloguing tools. This was most present with regard to facial recognition. In our test environment, AI recognized or partially identified female faces versus male faces. The females, in many cases, were grouped together as one person while the men were identified as individuals. With regard to subjective terms - adjectives which might be perceived as positive or negative descriptions - like pretty or ugly or skinny or fat or tall -

the computer had its own opinions as to how to visually present these topics, oftentimes presenting a bias.

The analysis of AI and image cataloguing in using currently accessible tools in a Cultural Heritage environment created access to images that had not been fully and textually described and allowed searches across multiple languages demonstrating the strengths. Yet there is room for the improvement of AI, including the ability to edit the metadata, so as to be more diversified and inclusive of genders and minority groups across communities.

5. CONCLUDING RECOMMENDATIONS

From these different projects and researches, recommendations can be distilled regarding metadata management for intangible heritage collections and participatory approaches to minority communities. In the “WEAVE Best Practices and Guidelines for Community Management”¹⁸, a series of takeaways have been formulated. These involve documenting and contextualising the existing metadata of collections, but also - very importantly - safeguarding the metadata version history: what changes have been made to the metadata in what period and for what reason? The metadata that we deem correct and publishable today might become obsolete in the future, but on the other hand their existence is a mirror and historical record of the time in which they were conceived. Open access to community and intangible heritage also means that institutions need to develop a clear publication strategy for these contents. From the projects discussed in this paper that aimed specifically toward participatory practices, we learned the importance of engaging the stakeholder groups and local communities in the selection, metadata enrichment and curation activities. Also, we encourage the enhancement of cataloguing and archival strategies, by supporting accessibility, sustainability and visibility, and being open to revise used vocabularies. Finally, as many of the metadata efforts require automation to be feasible, given the amount of data to be processed, there is an unavoidable recourse to artificial intelligence. This paper recommends taking into account the possible bias that these technologies engender, in particular in multilingual and multicultural settings.

¹⁸ <https://weave-culture.eu/capacity-building/resources/>