IMATGE I
RECERCA

I&R
CRDI 25 ANYS
Centre de Recerca i Difusió de la Imatge

17th International Conference
Girona, 16-19 November 2022

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

Antonella Fresa, Photoconsortium International Association
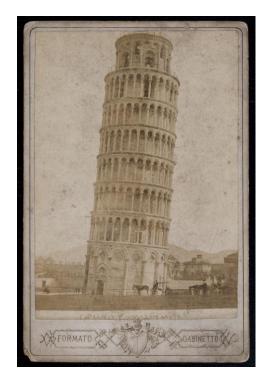
Tangible, intangible and digital heritage

photo courtesy Promoter Digital Gallery

photo courtesy CRDI Ajuntament de Girona

While **tangible heritage** has a compelling message of its own, as it is a visible and often imposing remnant of past cultures, when **intangible heritage** is not actively lived through persistent *traditions* of communities, it is often buried in documents and testimonials typically held in archives.

In the **digital world**, original documents are represented by a digital copy - which is expected to be a (re)presentation not only of the **documents** involved, but at the same time of the people, places, practices and **communities** involved.

Representation of intangible heritage

Opening access to intangible heritage collections in digital form adds layers to documental representation.

Cultural Heritage Institutions have big responsibilities:

- To take decisions on the **selection of content** to be digitised (what to share and what not to share..?) to allow harmonised progresses
- To add description and **metadata** (what is this object about..?) to enable retrieval and discovery
- To represent not only object and metadata, but also **cultural practices and values** (what is embedded in the digitised intangible heritage?...), to give voice to the communities involved

PARTICIPATORY APPROACHES FOR INCLUSIVENESS AND DIVERSITY

How different forms of participatory approaches can enrich digital cultural content in terms of *(re)presentation*

➜ democratising access to culture
➜ open up the management, preservation, presentation and interpretation of heritage

weave

Develop and provide a framework for linking and presenting the connections between **tangible** and **intangible** heritage of **cultural communities** *(also minority and under-represented or mis-represented communities)*, and bringing them to the center of attention by making them **accessible in Europeana**.

Key challenges

With regard to (**re)presentation of minority heritage, with particular regard to intangible heritage**, WEAVE identified the following questions:
- Which was the process that selected what to enter in the archive?
- Which are the cultural references used for the description of the contents in the metadata?
- Which is the history of these descriptions through different time frames and social contexts?
- Which is the connection of these contents with the represented communities
- How can we open up the contents in an accessible and understandable way to a wide audience

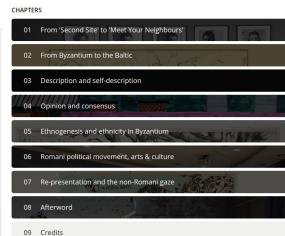photo courtesy Promoter Digital Gallery

WEAVE Approach

- The project worked closely with communities to aggregate over 8,000 new high-quality records that depict Cultural Heritage of minority cultural communities (i.e. Romani, Gypsy & Traveller, Portugues rural groups)

- The collections were showcased through a wide range of editorials

- The virtual exhibition named "The Renaissance of Romani Re-presentation" was written by a community member (as opposed to the authoring of "outside" people, aas it usually happens)

A DANÇA PORTUGUESA
A GOSTAR DELA PRÓPRIA

≡ 𝒆 europeana

Exhibitions

Explore the exhibitions

The Renaissance of Roman...

Gypsy, Roma, Traveller communities, images and identities over a millennium

CHAPTERS

01	From 'Second Site' to 'Meet Your Neighbours'
02	From Byzantium to the Baltic
03	Description and self-description
04	Opinion and consensus
05	Ethnogenesis and ethnicity in Byzantium
06	Romani political movement, arts & culture
07	Re-presentation and the non-Romani gaze
08	Afterword
09	Credits

https://www.europeana.eu/en/item/600/urn____eriac_223
https://www.europeana.eu/en/item/575/urn____pedexumbo_35

Labdays & capacity building

Replicable and lightweight formats to enable an open an egalitarian dialogue between researchers, professionals and communities.

"the mere fact of placing collections online does not necessarily lead to deeper connections with and between diverse communities" → WEAVE WhitePaper

LabDays with Communities

September-December 2021

In presence and online events organised with cultural communities and institutions to engage the participants in co-creation exercises

Diversity and Inclusion Workshops

February-March 2022

Four online events for cultural heritage professionals, to gain a better understanding and to get inspiration on how cultural institutions can support more diverse and inclusive experiences

All the events' recordings of the are available in the project's website and on its YouTube channel

The REACH Social Platform on participatory approached in culture and social innovation

Funded by EU Horizon 2020 Programme

- In the context of radical social changes taking place at global levels, Europe faces the need for its citizens to live together in peace and mutual respect and to value and enjoy diversities

- Cultural Heritage plays an important role in contributing to social integration in Europe and a more detailed picture of the impact of participatory methodologies associated with culture, can enhance its potential for social good

- Four specific pilot actions were carried out, located in different EU countries, addressing participation in different declinations of cultural heritage

Key challenges

photo courtesy of C-Dare Coventry University CC-BY SA

- Compare and contrast regional examples of practice, including the case of small towns in heritage marketing and tourism

- Engage with staff members from different types of cultural institutions to support transformations and to host participatory practices

- Collect and share examples of good practices, encouraging further replications

- Promote participatory mediation processes between local stakeholders, communities and regional/ national administrations and institutions

- Create co-governance initiatives for territorial safe-keeping

REACH Pilots

Minority heritage

Small towns heritage

Rural heritage

Institutional heritage

INCULTUM Visiting the margins Participatory approaches to sustainable Cultural Heritage and tourism

Visiting the margins
INnovative CULtural ToUrisM in European peripheries

Tourism is one of the most important industries and economic activities, a key element in the globalization processes and a development opportunity for territories.

It great potential when it comes to culture, nature, knowledge, and personal experiences.

It is a way to learn and improve oneself and it enriches our vision of the world and helps to improve mutual understanding.

However …

Key challenges

INCULTUM

NEGATIVE IMPACTS EXIST AT DIFFERENT LEVELS
- Touristification, gentrification, massive urbanization
- Insecurity of employment, social changes
- Reduction of the quality of visitor's experience
- ……
-

INCULTUM REPLY
- Adopting participatory and collaborative approaches
- Transforming local communities in protagonists
- Engaging stakeholders as dynamic agents of transformation

photo by Pietro Masi CC-BY SA

INCULTUM Pilots

10 experimental experiences in 9 European countries:
Campina de Faro, Altiplano de Granada, Bibracte in Morvan, San Pellegrino in Alpe, Montains of Trapani, Vjosa/AOOS Valley, Central Slovakia, Sweden Archipelago, Historic Graves in Ireland

INCULTUM

CROWDSOURCING FOR METADATA IMPROVEMENT

Leveraging the knowledge and participation of citizens, students, culture lovers who are willing to collaborate in preserving and maintaining Cultural Heritage collections

Citizen Heritage Engaging citizens in education and cultural heritage curation

- Cultural Heritage Institutions are constantly looking for new ways to involve citizens in their activities.
- Focus on technological innovations to manage digital cultural heritage can support and enhance Citizen Science participation, both from a pedagogical and heritage perspective.
- We investigate what needs to be added to move from crowdsourcing to active citizen science engagement.
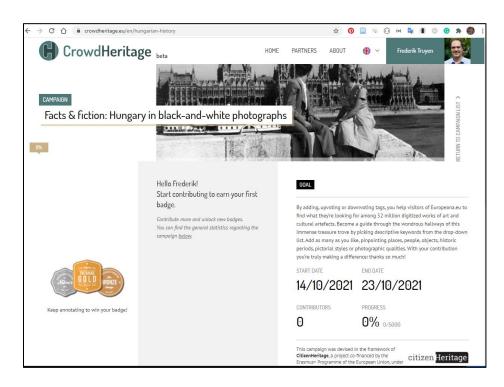
citizen Heritage

Erasmus+

Enrichment through crowdsouring

We used the CrowdHeritage platform asking citizens to enrich metadata of cultural collections:
- Create volume via a low-threshold platform and easy-to use tools
- Add a game and a community element
- Install peer review to optimize results
- Take a human in the loop for final validation
- Feed back to the source

AI
FOR METADATA
IMPROVEMENT

How can AI assist in cataloguing material found in museums and archives?

AI Annotators used in WEAVE

On the basis of **existing thesauri** (Wikidata, Getty AAT) and a selected **keywords list** about tangible and intangible heritage, two tools for automated metadata enrichments were applied to WEAVE collections:

- **NERD Annotator**
  performs a Named Entity Recognition and Disambiguation (NERD)

- **Linked-Data Annotator**
  Link text to Thesaurus/Vocabulary terms

After the enrichments, the records were validated for granting accuracy and then the enriched collections were re-aggregated in Europeana.

WEAVE enrichments - human knowledge

The approach in WEAVE was a hybrid one: the bulk of the enrichment work was allocated to the machine, but human knowledge was deeply embedded in the process. This achieved a good balance in having a massive enrichment of data and granting the highest confidence level of enriched records.

- **WEAVE keywords list**
  The terms collected reflect different terminologies, relating to the different types of cultural heritage that were part of the WEAVE collections.

- **Enrichment validation with human-in-the-loop**
  A round of validation of the automatically generated enrichments was then provided by the content partners on their respective collections, so to accept or reject annotations that the machine recognized for each record.

Making sure that sufficient, precise, correct, multilingual, rich, descriptive, authoritative and interlinked metadata are available is a task that is equally monumental, and can't be completely taken over by AI (yet).

What we propose is **a path in between manual metadating and automated enrichment.**

Manual vs Automatic strategies

Manual

- ☑ High quality enrichments
- ✖ Labour-intensive process

Automatic

- ☑ Large scale
- ☑ Easy to execute
- ✖ Prone to errors

Crowdsourcing

- ☑ Large scale
- ✖ Lack of expert knowledge
- ✖ Prone to errors

HOW AI WORKS FOR MUSEUMS

Are there hidden risks and biases in using AI?

Image cataloguing using artificial intelligence

J. Mączyński, Lexicon Latino-Polonicum, Königsberg 1564, Digital Library polona.pl, Poland, Public Domain

AI IN RELATION TO GLAMS TASK FORCE

Report and recommendations

Co-financed by the Connecting Europe
Facility of the European Union

europeana
network
association

https://pro.europeana.eu/files/Europeana_Professional/Europeana_Network/Europeana_Network_Task_Forces/Final_reports/AI%20in%20relation%20to%20GLAMs%20Task%20Force%20Report.pdf

How can AI be used in museum and archive cataloguing?

- Visual recognition – Objects, subjects

- Facial recognition – Artists, Staff, VIP

- 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 Data)

How can AI be used in museum cataloguing?

Some faults include:

Gender, racial, and language algorithmic biases

Algorithmic biases due to coding based on assumptions and computer misunderstandings

→ Because of these issues, human intervention is necessary.

Visual recognition – Objects, subjects

Identifying other items within an image

Searching by identifying traits

Unknown photographer
Chana Orloff, Petah Tikva
1906
Postcard
Gabriel Talphir Archive, Information Center for Israeli Art, Israel Museum, Jerusalem

Facial recognition – Artists, Staff, VIP

Facial recognition in IMJ's collection
Identifying Avigdor Arikha's face across collections

Thu, Nov 10, 2016

Sun, Jul 26, 2015

Mon, Jan 16, 2012

Avigdor Arikha, Israeli, 1929-2010
Canvas with Self Portrait
1976
Oil on canvas
Purchase, Ayala Zacks-Abramov Fund
B77.0620
Israeli Art Department

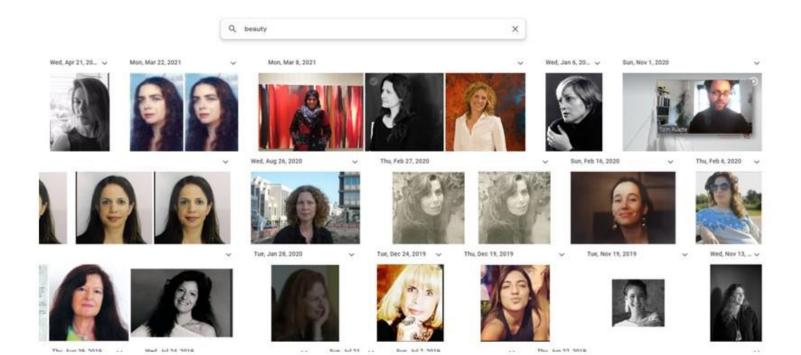"Avigdor Arikha" exhibition opening, 1998
IMJ Photography Laboratory

Avigdor Arikha
Photographer: Stanley I. Batkin
From the Archives of the
Information Center for Israeli Art
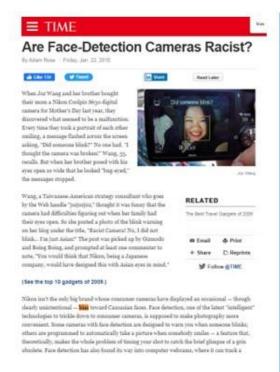
Facial recognition – Artists, Staff, VIP

Facial recognition – Artists, Staff, VIP

CONCLUDING REMARKS

Recommendations can be distilled regarding participatory approaches to minority communities and metadata creations and management for (intangible) heritage collections

Digitization and cultural heritage

- Digitization is not only concerned with preservation of cultural heritage, but it relates also to the issue of giving access to collections to an ever-changing user base.

- There is need for new approaches in terms of bottom-up, participatory and co-creative methodologies, whereby constituent communities are empowered to take ownership for the ways in which their heritage is managed, safeguarded and (re)used.

- Ensuring that digital collections allow for the capture of intangible heritage and cultural practices requires new approaches to metadata. These can not be limited to physical properties of the digitized objects: they need to be embedded in narratives that tell the story of the underlying cultural practice.

Using the machines: room for improvement

- AI and machine-based metadata generation practices are the future in supporting the cataloguing work of cultural collections, but CHIs should build capacity about these tools and new workflows
- There should be awareness regarding AI inaccuracies and bias in cataloguing
- Machine-generated tags should be marked as such
- AI shortcomings can be improved with further training
- Human intervention, knowledge and validation is still needed despite the growing abilities of the machine

This presentation is joint effort of the following authors

photo courtesy Promoter Digital Gallery

Valentina Bachi (Photoconsortium): valentina.bachi@photoconsortium.net

Rosa Cisneros (Coventry University): ab4928@coventry.ac.uk

Antonella Fresa (Photoconsortium): antonella.fresa@photoconsortium.net

Allison Kupietzky (The Israel Museum, Jerusalem): allison@imj.org.il

Alexandru Stan (IN2 Digital Innovations): as@in-two.com

Frederik Truyen (KU Leuven): fred.truyen@kuleuven.be

Many thanks for your attention

Antonella Fresa
Photoconsortium
www.photoconsortium.net