

Gender in the digitisation of cultural heritage

Evidence, cases and practical steps for heritage institutions

Training webinar | DIGICHer with the Citizen Science Association in Lithuania | 2026

Dr. Monika Mačiulienė

VILNIUS TECH | DIGICHer Gender Equality and Sensitivity Officer
Director, Citizen Science Association in Lithuania



PILIEČIŲ MOKSLO
ASOCIACIJA



This project has received funding from the European Union's Horizon Europe Research and Innovation programme under grant agreement No. 101132481 (DIGICHer).



Gender gaps in three digital collections



Wikipedia, US art museum catalogues and an online citizen science platform

ENGLISH WIKIPEDIA

20%

of biographies are about women.

June 2024. Wikidata Human Gender Indicators.

18 MAJOR US ART MUSEUMS

87%

of the artists in their collections are men.

Topaz et al., PLOS ONE, 2019.

ZOONIVERSE CITIZEN SCIENCE

30%

of volunteer classifications with a known gender came from women.

Ibrahim, Khodursky and Yasseri, 2021.

In all three cases, women are under-represented among the people recorded or among the contributors.



Funded by the European Union · Horizon Europe grant agreement No. 101132481 (DIGICHer)

Misattribution in the Smithsonian collections

Smithsonian Institution, natural history collections

WHAT THE PAPER RECORDS SAID

S. Ripley

01 Suzanne Ripley, primatologist

Her research, written as “S. Ripley” on paper records, was commonly credited to someone else once the metadata was typed into the database.



02 Sidney Dillon Ripley, ornithologist

For a time Secretary of the Smithsonian. The database gave him her work.

WHAT THE RECORDS SHOWED

40

women identified with specimen collections or identifications

120,000+

specimens attributed to them

“Barriers to accurate specimen attribution disproportionately affect women collectors and determiners.”

Dikow et al., 2023

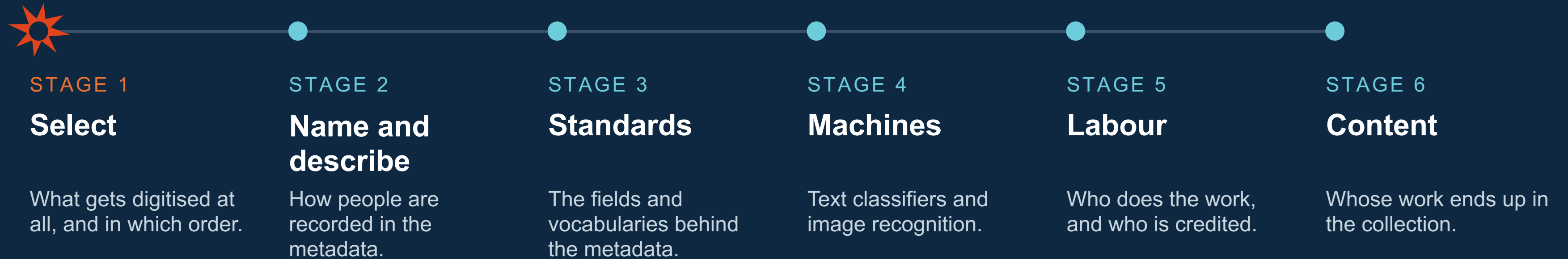
The transfer of paper records into the database credited part of Suzanne Ripley’s work to another person.



Funded by the European Union · Horizon Europe grant agreement No. 101132481 (DIGICHer)

Six places where gender enters the pipeline

One question per stage. The questions become a checklist at the end.



IN SHORT

Gender gaps can be reproduced or corrected at each of the six stages.



Funded by the European Union · Horizon Europe grant agreement No. 101132481 (DIGICHer)



What gets digitised

How institutions decide which material is digitised first

Money, staff and politics decide

Selection follows funding, staff time, research priorities, user demand, preservation needs and institutional politics. The gaps of the analogue collection are repeated online, and often magnified.

Use-driven selection is self-fulfilling

The material used most in the past is digitised first. The existing audience is served again. A community that never used the archive may never learn that it holds something about them.

Flat and scannable goes first

Material that is hard to scan waits. Objects, textiles and oral material, where much of women's heritage sits, are the last in the queue.

Naming the gap in public

The UK National Archives' women's histories finding aid states that women's voices are marginalised in its records.

- The gap is larger for women of colour, working-class women, disabled women, and lesbian and bisexual women.
- The finding aid then points to the records that do exist: small curated sets of digitised documents, each introduced by a specialist.
- Few institutions write this down. Most digitisation strategies say nothing about who is missing.



2

STAGE 2 · NAME AND DESCRIBE

How women are named in metadata

Three archive cases from the UK and the US

TALE ARCHIVE · EDINBURGH

Story credited to the husband

One storyteller is recorded only as Bean Sheumais, wife of James.

Another record credits “Andrew Stewart and family”. The document itself names Mrs Andy Stewart, who told the story about her own experience with a ghost.

TREVELYAN ARCHIVE · NEWCASTLE

Catalogued under the husband's name

The papers of Sir Charles and Lady Mary Trevelyan are catalogued as the Charles Philips Trevelyan Archive.

A National Archives grant now funds a machine-learning search of the catalogue for other cases.

MRS. HIS NAME PROJECT · NEVADA

A published method

The University of Nevada, Reno went through its finding aids to identify married women listed only under their husbands' names.

The archivists published the method and call the finding aid a living document.



This work is called reparative description. It is mostly done by hand.



Funded by the European Union · Horizon Europe grant agreement No. 101132481 (DIGICHer)

Correcting names without losing information

A dilemma raised by heritage professionals in Havens' study at Edinburgh

IF YOU DELETE THE OLD FORM

Information that can be lost

- “Mrs. John MacDonald” is all the record has. Remove “Mrs.” and her identity goes with it.
- One participant traced a woman through her husband’s documentation, then learned that she had brought the money that built the couple’s art collection.
- Gendered terms are often the only lead to a woman’s role.

ADD, DO NOT DELETE

Recommended practice

- Keep the historical form as it was written.
- Add the restored name beside it.
- Record who added it, when, and on what evidence.
- Mark every later change, so the catalogue tells its own story.





The gender field

Name authority records in library cataloguing. RDA rule 9.7, MARC field 375.

2013

A field is added

The new cataloguing standard RDA asks cataloguers to record a creator's gender in name authority records.

The Library of Congress allows three values: male, female, not known.



2016

A study looks inside

Of the records with a gender value, 65 percent revealed that a self-identified transgender author had changed name or gender.

The catalogue was outing people.



2022

The guidance reverses

New guidance for cataloguers: do not record gender.

Delete existing values whenever you edit a record for any other reason.



The field was introduced to improve description. In practice, it exposed personal information.





Can a model find biased descriptions?

Lucy Havens, University of Edinburgh, PhD 2024, on the university's own archive catalogue

01

Write a taxonomy

What counts as gendered language, and what counts as biased.

02

Label by hand

Thousands of catalogue descriptions, labelled by archivists and researchers.

03

Train classifiers

Models learn from the labels to recognise the same patterns elsewhere.

04

Flag for review

The model points. A person decides what to do with the description.

WHAT THE MODELS CANNOT DO

- They point to descriptions that need a human. They do not remove bias, because the catalogue is a historical document.
- Debiasing methods built for social media or Wikipedia do not suit heritage collections.
- The taxonomy and the labelled datasets are published openly, so other archives can build on them.



Gender bias in image recognition

Image recognition on heritage collections

COMMERCIAL TAGGING SERVICES

Stereotype labels

Services trained on web photos attach “cooking” and “shopping” to images of women, “golfing” and “snowboarding” to images of men. Training can amplify the pattern beyond the data it learned from.

Barlas et al., 2021

MUSEUM TEST · TWO PORTRAITS

Same pose, different tags

Two Chuck Close portraits, a man and a young woman, each with a cigarette and the same expression. The man was tagged “funny” and “crazy”. The woman was not.

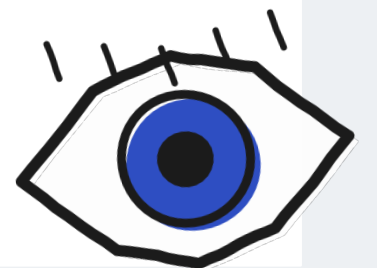
Ciecko, Museums and the Web, 2020

DIGITALTMUSEUM · 74,000 PHOTOGRAPHS

Inconsistent existing tags

1930s photographs from Swedish museums. Before any model ran, the existing “woman” and “man” tags were inconsistent between institutions. A separate audit traced CLIP’s gender bias across a century of archival photographs.

Visual Studies, 2026; Springer, 2025



Keep machine-made tags separate from human ones, and say which is which.



Who does the work

Volunteers, editors, staff, and the people who typed the first catalogues

30%

of Zooniverse classifications with a known gender came from women.

Data 2009 to 2013. Wildlife projects: over half. Astronomy: under 20%.

80%

of active Wikipedia editors were men in 2024.

Wikimedia community data.

61%

of museum professionals are women.

Mellon Foundation estimate, 2022.

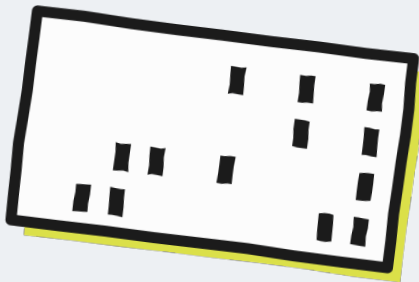
43%

of US museum directorships are held by women.

Association of Art Museum Directors.

THE KEYPUNCH HISTORY

When museums first moved their catalogues onto computers, the data entry was done by women hired as typists and keypunch operators, often on short contracts, and not counted as knowledge production. The same feminised labour sat behind the Index Thomisticus of the 1950s, usually named as the first digital humanities project.





Whose work gets in

Topaz et al. scraped the online catalogues of 18 major US museums: 10,000 artist records, 2019

87%

of the artists are men.

85%

of the artists are white.

01 Below 8% women

Detroit Institute of Arts, the Met, MFA Boston. Leaders: LA MOCA at 24.9%, the Whitney at 22%. Museums with the same collecting mission had very different numbers.

02 11% of acquisitions

Works by women as a share of US museum acquisitions between 2008 and 2020. Burns Halperin Report, 2022.

03 A second filter

On English Wikipedia in 2017, 41% of biographies nominated for deletion were about women, when women were 17% of all biographies.

The study was possible because the catalogues were available online.



Initiatives that have changed the numbers

Most of these are run by volunteers, at low cost

WIKIPEDIA · WOMEN IN RED

About 175,000 articles in five years

Women's share of English Wikipedia biographies rose from 15% in 2015 to 20% by December 2024. The project is one of the main reasons.

LIBRARIES AND MUSEUMS · ART+FEMINISM

Edit-a-thons since 2014

By 2018: more than 7,000 people at more than 500 events had created or improved over 11,000 articles on women in the arts. The format is a room, a date and a tutorial.

EVIDENCE · BIOLOGY FACULTY

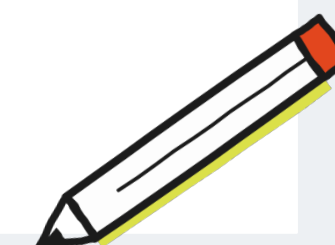
Women now more often included

Among biology faculty at top US universities, 9.4% of women now have a Wikipedia biography, against 7.5% of men. Men led until 2018. The pattern reversed in 2022, after years of campaigns.

AGGREGATOR · EUROPEANA

Women's history every March since 2016

Editorials with partner institutions, the Pioneers exhibition of 2019, and Transcribathon runs on letters and diaries written by women.



Three formats a citizen science association can run

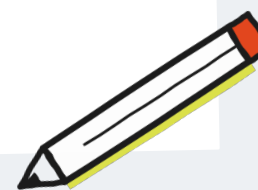
Each one has a tested model. Each one produces something a heritage institution can use.

MODEL · ART+FEMINISM

An edit-a-thon

One day, one room, a tutorial for beginners and a list of missing women. Participants write or expand Wikipedia and Wikidata entries on women connected to a collection.

Needs: a library or museum as host, a Wikimedia trainer, the list.
Produces: articles, and a count of how many.



MODEL · TRANSCRIBATHON

A transcription run

Volunteers transcribe letters, diaries and documents written by women, so that the texts become searchable and the names become findable.

Needs: scanned documents with a licence that allows it, a platform, a few weeks.
Produces: full text, and names for the metadata.



MODEL · MRS. HIS NAME

A name audit

Volunteers go through finding aids and catalogue entries and list every woman recorded under a husband's name, as "wife of", or not at all. The archive decides what to add.

Needs: an archive willing to open its catalogue, a simple form, a reviewer.
Produces: a list of corrections with evidence.



In each format, volunteers are credited by name.



Funded by the European Union · Horizon Europe grant agreement No. 101132481 (DIGICHer)

Archives dedicated to women's history

Three examples of separate digital archives

FOUNDED 1995 · UNITED STATES

Jewish Women's Archive

Founded to use the internet for Jewish women's stories. Its editors wrote that women were absent from accounts of the Jewish past because male experience served as the measure of historical significance.

REVISED 2021

Shalvi/Hyman Encyclopedia of Jewish Women

About 350 entries updated and more than 180 added, with attention to Sephardi and Mizrahi women, Jewish women of colour, LGBTQ Jews and women with disabilities.

FOUNDED 2016 · LUSAKA

Women's History Museum of Zambia

A museum that exists only online. More than 5,000 digital items on African women's history and indigenous knowledge, curated from an African perspective.



Gender-based access rules

Community protocols in Indigenous archives

MUKURTU · WARUMUNGU COMMUNITY, AUSTRALIA

Access based on community protocols

- Men do not view women's rituals. Family members do not view images of the deceased.
- Access can be limited by gender, age or family. A user sees only the content whose protocol matches their own.
- The platform is open source and now used by Indigenous communities and institutions worldwide.

SÁMI CRAFT ARCHIVES · ÁJTTE MUSEUM AND THE SÁMI ARCHIVES

Makers involved in digitisation

- Archives of duodji makers were digitised in Jokkmokk, Sweden, and Inari, Finland.
- The makers and artists took part in the digitisation itself, from donor agreements to database registration.
- The stated aim was to end the institution's role as the only expert and manager.



Open access and community control over access can conflict. In such cases, a gender-sensitive approach may mean restricting access.



Five open dilemmas

Questions without a single correct answer

01 Record gender, or not?

A field helps discovery and makes counting possible. It also outs people. The library world added the field in 2013 and withdrew it in 2022.

02 Repair the name, or keep the record?

Restore her name, and keep the historical form that led you to her. Deleting the old form can delete the only clue.

03 Machine tags at scale, or human description?

A million images described in a week, with the web's stereotypes attached. Or a few thousand a year, described by people.

04 Open access, or community rules?

Some knowledge is gendered by design. For the communities that hold it, openness can be a harm.

05 How do you count?

Zooniverse guessed gender from first names with software. The museum study asked crowd workers to guess. Both methods are questionable. Both were the only way to get a number.



One question per stage

A checklist for a heritage team, before publishing a digitised collection

1

SELECT

Who is missing from what we chose to digitise, and did we say so in public?

2

NAME

Does every person have their own name, wherever the source allows it?

3

STANDARDS

Which fields describe people, and who could be harmed by them?

4

MACHINES

Is every machine-made description marked as machine-made?

5

LABOUR

Who does the work, and are they credited by name?

6

CONTENT

What share of the creators in our collection are women, and do we know the number at all?

Most institutions do not have data to answer the last question.



Funded by the European Union · Horizon Europe grant agreement No. 101132481 (DIGICHer)

What would you change first?

In your own collection, catalogue or dataset

- 01 Which of the six stages is the weakest where you work?
- 02 Which of the six questions could you answer this year?
- 03 What could a citizen science project in Lithuania take on first?



Use the chat, or unmute.



Sources (1 of 2)

- Dikow, R. B. et al. (2023). Let the Records Show: Attribution of Scientific Credit in Natural History Collections. *International Journal of Plant Sciences* 184(5), 392–404.
- Smithsonian American Women’s History Museum. The Challenge of Metadata in Uncovering Women’s History (womenshistory.si.edu).
- Banks, C. (2023). Decoding Hidden Women: Feminist Digitisation Practices in the Tale Archive. Gaelic Algorithmic Research Group, University of Edinburgh.
- Newcastle University (2024). Finding Her Name in the Archive. National Archives Research and Innovation Grant.
- Olson, E. A. The Mrs. His Name Project. *Journal of Western Archives* 14(1).
- The National Archives (UK). Women’s Histories Finding Aid.
- Science Museum Group, Congruence Engine (2023). Data Is Never Raw: Ethics and Biases in Digital Cultural Heritage Collections.
- Billey, A., Drabinski, E. and Roberto, K. R. (2014). What’s Gender Got to Do with It? A Critique of RDA Rule 9.7.
- Thompson, K. J. (2016). More Than a Name: Name Authority Records for Authors Who Self-Identify as Trans. *Library Resources & Technical Services* 60(3).
- PCC Ad Hoc Task Group on Gender in Name Authority Records (2022). Guidance summarised by NISO, May 2022.
- Havens, L. (2024). Recalibrating Machine Learning for Social Biases. PhD thesis, University of Edinburgh.
- Havens, L., Alex, B. and Terras, M. (2025). Confronting Gender Bias in Heritage Catalogues. In: *The Routledge Handbook of Heritage and Gender*.
- Havens, L. (2025). Data Bias and the Natural Language Processing of Metadata. In: Gooding, P. and Terras, M. (eds), *Library Catalogues as Data. Facet*.
- Barlas, P. et al. (2021). To “See” Is to Stereotype: Image Tagging Algorithms, Gender Recognition, and the Accuracy-Fairness Trade-off. *PACM HCI* 4 (CSCW3).
- Ciecko, B. (2020). AI Sees What? The Good, the Bad, and the Ugly of Machine Vision for Museum Collections. *Museums and the Web 2020*.
- All Computer Vision Models Are Wrong/Useful: Reflections on Automatic Detection of Gender in Historical Photographs. *Visual Studies* (2026).



Sources (2 of 2)

- Revisiting How We Access Historical Archives: Auditing Gender Stereotypes and the Division of Labour in Historical Photography Collections. Springer (2025).
- Disentangling Algorithmic Bias from Archival Artifacts: A Controlled Audit of Vision-Language Model Valuation in Metropolitan Museum Archives. arXiv (2026).
- Ibrahim, K., Khodursky, S. and Yasseri, T. (2021). Gender Imbalance and Spatiotemporal Patterns of Contributions to Citizen Science Projects: The Case of Zooniverse. *Frontiers in Physics* 9.
- Rayburn, A. (2024). *Pink-Collared Computing: Examining Gendered Work in Museum Collections*. University of Michigan Museum Studies.
- Nyhan, J. *Hidden and Devalued Labour in the Digital Humanities: On the Index Thomisticus Project 1954–67*. Routledge.
- Smyth, H., Nyhan, J. and Flinn, A. (2020). Opening the “black box” of digital cultural heritage processes: feminist digital humanities and critical heritage studies.
- Topaz, C. M. et al. (2019). Diversity of Artists in Major U.S. Museums. *PLOS ONE* 14(3). Burns Halperin Report (2022).
- Wikipedia: Gender bias on Wikipedia; Women in Red; Wikidata Human Gender Indicators (Klein and Konieczny). Science (AAAS) report on the Proceedings of the Royal Society B study of biology faculty biographies.
- Art+Feminism (artandfeminism.org).
- Europeana. *Women’s History; Pioneers: Trailblazing Women in the Arts and Sciences* (2019); Transcribathon Women’s History Run.
- Jewish Women’s Archive (jwa.org). *The Shalvi/Hyman Encyclopedia of Jewish Women* (2021 revision).
- Women’s History Museum of Zambia (whmzambia.org).
- Christen, K. (2012). Does Information Really Want to Be Free? Indigenous Knowledge Systems and the Question of Openness. *International Journal of Communication* 6. Mukurtu CMS (mukurtu.org).
- Westman Kuhmunen, A. and Linkola-Aikio, I.-A. (2025). Samifying Archival Practices and Digitising Sami Duojár Archives with the Slow Archive Method. In: *Digital Indigenous Cultural Heritage*. Palgrave Macmillan.



Thank you

If you want the checklist or any of the sources, write to me.

Dr. Monika Mačiulienė

VILNIUS TECH, Faculty of Creative Industries · DIGICHer Gender Equality and Sensitivity Officer
Director, Citizen Science Association in Lithuania (Piliečių mokslo asociacija)

monika.maciuliene@vilniustech.lt

DIGICHer · Digitisation of cultural heritage of minority communities for equity and renewed engagement

The DIGICHer Gender Equality Plan (D8.4) is a public deliverable. It includes gender-sensitive language guidelines and harassment safeguards.

Organised by VILNIUS TECH for the DIGICHer project, together with the Citizen Science Association in Lithuania.



PARTNERS



This project has received funding from the European Union's Horizon Europe Research and Innovation programme under grant agreement No. 101132481 (DIGICHer).



DIGICHer