


ARTIFICIAL INTELLIGENCE IN ARCHIVAL PRACTICE: focus on data curation and enrichment

Valentine Charles, Data Services Director

EUscreen Symposium

A bronze sculpture of a mermaid, known as 'Havfrue', designed by Marie Carl-Nielsen. The sculpture depicts a woman with a fish tail, sitting on a pedestal. The background is dark with a grid pattern.

MERMAID (DANISH: HAVFRUE) IS
A BRONZE SCULPTURE DESIGNED
BY THE DANISH SCULPTOR ANNE
MARIE CARL-NIELSEN,
DEPICTING A MERMAID.

PLACE: SMK - STATENS
MUSEUM FOR KUNST

LICENSE: CC0

1921



A **machine-based system** designed to operate with varying levels of **autonomy** and **adaptability**, capable of **inferring**, from the **input it receives**, how to **generate outputs** such as predictions, content, or recommendations (...)

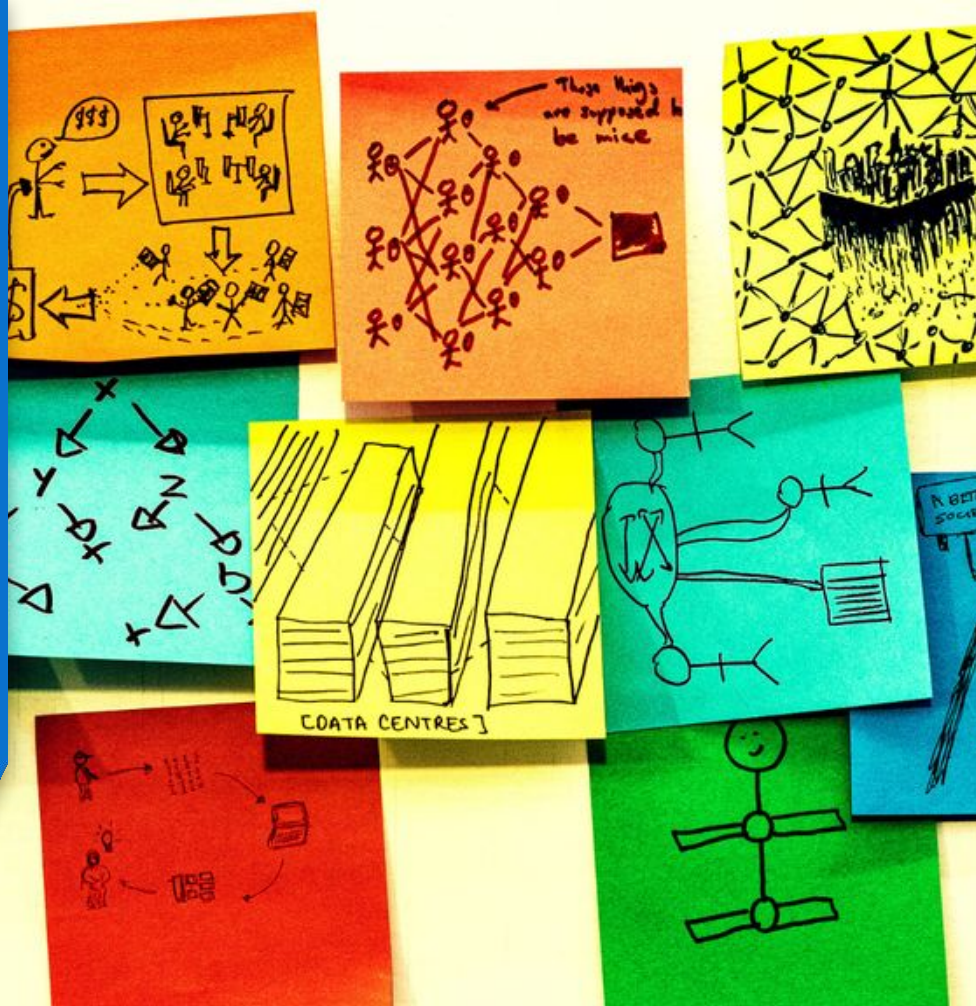
EU AI Act, 2024, European Commission

AI system

AI for cultural heritage



Learning by doing: AI Experiments at Europeana & the data space





The issue:

Thousands of records in Europeana.eu have no description, just a title



The solution:

We used Data-to-Text methods to convert information from a structured format into natural language

Metadata enrichment



The process:

We applied and evaluated different D2T methods and their combinations

Rule-based Data-to-Text system (ChatGPT)

Input:

dc:type=**fashion collection**, dc:type=**womenswear**,
dc:type=**spring-summer**, mrel:dsr=**Barbara Bui**,
dcterms:created=**2013**, mrel:pht=**Etienne Tordoir**,
edmfp:model=**Ava Smith**

Output:

The objects are fashion collection, womenswear, spring-summer. *They were designed by* Barbara Bui. *The creation date was* 2013. *They were photographed by* Etienne Tordoir. *The fashion model was* Ava Smith.

Data-to-Text with generative AI (ChatGPT)

Input:

```
{ "model" : "text-davinci-003", "prompt" : "make a short,
factual descriptive text with following key-value pairs:
object=fashion collection, object=womenswear,
object=spring-summer. fashion designer=Barbara Bui.
date=Spring-Summer 2013. photographer=Etienne Tordoir.
fashion model=Ava Smith. Use dry business language.",
"temperature" : 1, "max_tokens" : 73, "top_p" : 0,
"frequency_penalty" : 0, "presence_penalty" : 0 }
```

Output:

The Spring-Summer 2013 fashion collection from Barbara Bui featured womenswear photographed by Etienne Tordoir and modelled by Ava Smith.

Rule-based + generative AI

Input:

```
{ "model" : "text-davinci-003", "prompt" : "Rewrite this  
text: The objects are fashion collection, womenswear,  
spring-summer. They were designed by Barbara Bui. The  
creation date was 2013. They were photographed by Etienne  
Tordoir. The fashion model was Ava Smith.", "temperature" :  
1, "max_tokens" : 61, "top_p" : 0, "frequency_penalty" : 0,  
"presence_penalty" : 0 }
```

Output:

*Barbara Bui's Spring-Summer 2013 fashion collection,
womenswear, was photographed by Etienne Tordoir and modeled
by Ava Smith.*

Rule-based + ChatGPT + Wikidata

Input:

```
{ "model" : "text-davinci-003", "prompt" : "Rewrite this  
text: "Womenswear created by the French fashion designer  
Barbara Bui (1956) for the Womenswear Spring-Summer 2013  
collection. The creation was presented by the American model  
Ava Smith (1988). The photograph was taken by the Belgian  
fashion photographer Etienne Tordoir.", "temperature" : 1,  
"max_tokens" : 61, "top_p" : 0, "frequency_penalty" : 0,  
"presence_penalty" : 0 }
```

Output:

The Women's Spring-Summer 2013 collection featured womenswear designed by French fashion designer Barbara Bui (born in 1956). The collection showcased a creation modeled by American model Ava Smith (born in 1988) and captured in a photograph by Belgian fashion photographer Etienne Tordoir.



The results:

Identified the combination of methods that yielded best results



The learnings:

Evaluation of the capabilities and results of each method is key

What we're learning



Next steps:

Apply at scale & explore translations options



The issue:

Cultural heritage catalogues often contain outdated descriptions that don't reflect evolving language and societal values



The solution:

AI-powered tools to analyse millions of records to spot and correct outdated or harmful terminology

Detect and correct bias



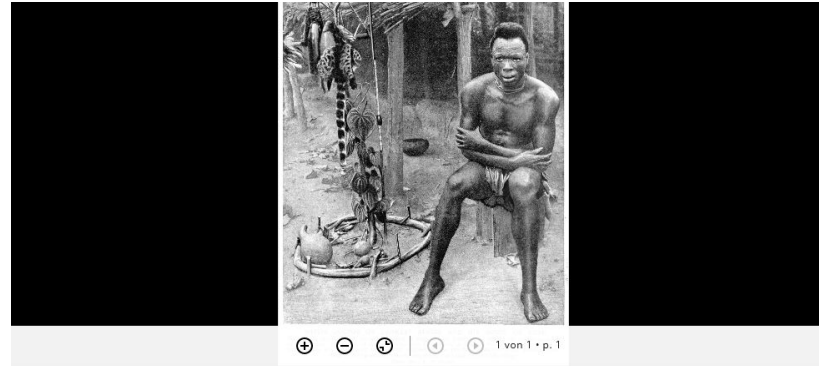
The process:

DE-BIAS is analysing 4.5M+ records on Europeana.eu to flag and update problematic metadata and has co-curated a vocabulary of terms with communities

The DE-BIAS project

Example of a Eurocentric, supposedly enlightened view of 'the other'

The title contains a whole series of **discriminatory terms and stereotypes**, and also the used key words



© CC BY



TEILEN

HERUNTERLADEN

Witch doctor of darkest Africa & his house of fear

Dieses Objekt wird von Wellcome Collection bereitgestellt und gepflegt

Auf der Website des Anbieters ansehen

Gut zu wissen Alle Metadaten

Betreff

Practitioner and equipment ; Non-European medicine ; Primitive knowledge: witch doctor



The results:

+ 180.000 terms detected in 5 languages
co-curated vocabulary with 600+ terms published as a linked open data knowledge graph



The learnings:

Importance of high accuracy for sensitive terms; limitations of text analysis algorithms; importance of bottom-up approach; human validation and evaluation campaigns is key

What we're learning

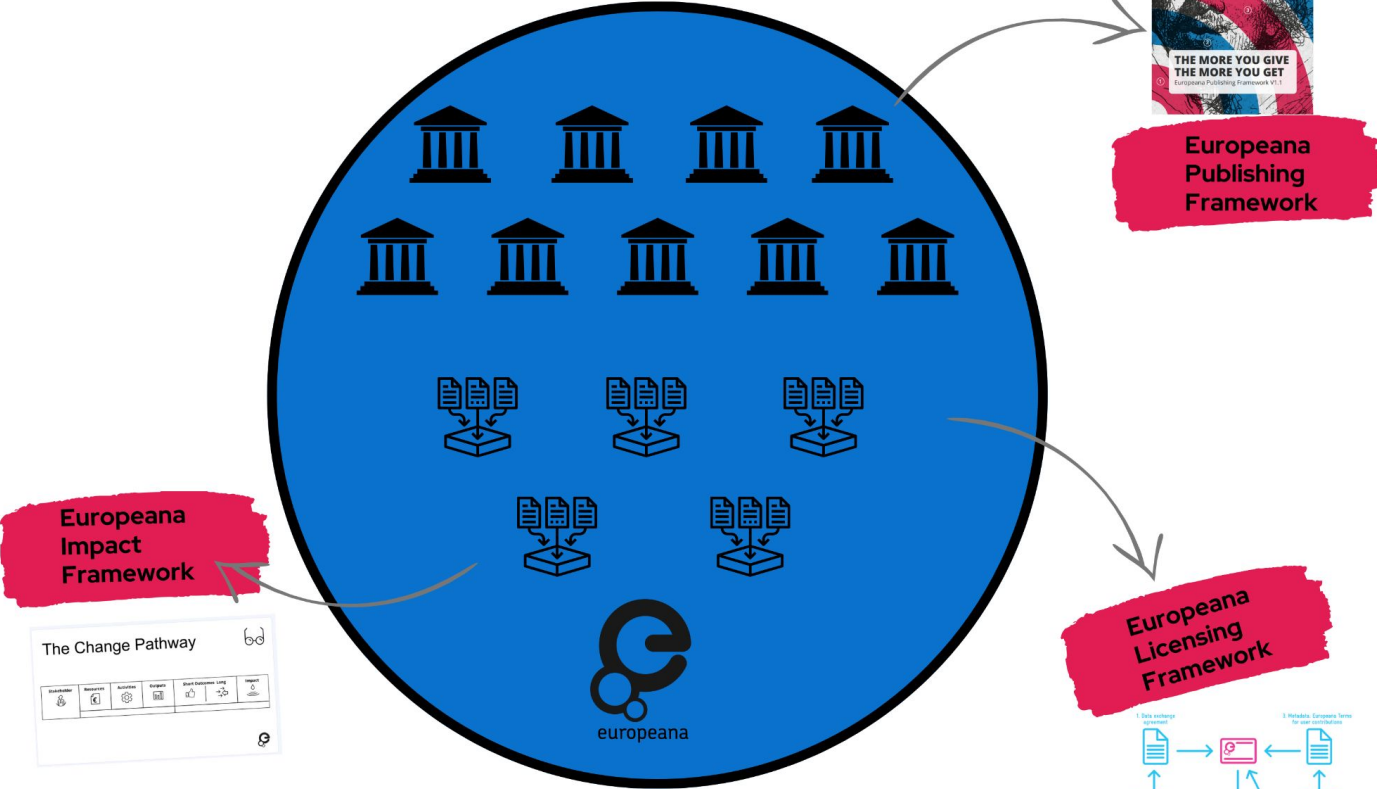


Next steps:

Maintaining and expanding vocabularies to more languages and evolving sensitivities; focus on community engagement and capacity building

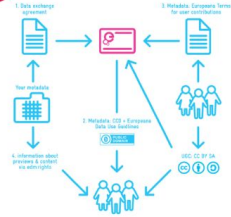
Cultural heritage for AI

CHIs are 'the original data set'



The Change Pathway

Knowledge	Research	Activities	Engage	Short Duration Life	Impact



Some challenges we face as a sector...



Some challenges...



TECHNICAL

Lack of transparency & traceability of information

Prone to errors & 'Hallucinations'

Environmental sustainability of LLMs



ETHICAL

Perpetuation of negative bias

Homogenisation

Power dynamics in AI development & deployment



LEGAL

Privacy concerns

Reuse of in-copyright data in text & data mining

Risk of illegal copying in genAI outputs



SKILLS

Lack of clear vision & long-term strategy in AI use by CHIs

Need for new foresight & technology management skills

Need for continuous upskilling as tech evolves



FUNDING

Need to invest in training for AI literacy & technical expertise

Financial resources to invest in AI projects

Costly tools & infrastructure

What's next?

Capacity building


The AI4Culture project

Focus on scenarios with **high interest from CHIs**


Customised, extended, and re-packaged versions of **existing tools** developed in previous projects

Can be applied either independently from the **Europeana platform** or in interconnection with it.


AI4Culture


|| TOOL 

AI4C Color detection tool




The color detection tool is a web based tool that can detect the type and the color of objects depicted on...


 Creator Datable

 License MIT


AI4Culture


|| TOOL 

OCCAM



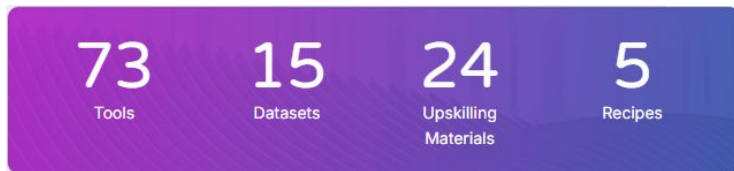
Tool for automatically transcribing and translating printed or handwritten documents and automatically...

 Creator CrossLang NV (Bel...

 License MIT

Explore and contribute to the platform!

<https://ai4culture.eu/>



ai4culture

[Explore](#) [Mission](#) [Contribute](#) [Contact Us](#)

[Search](#)

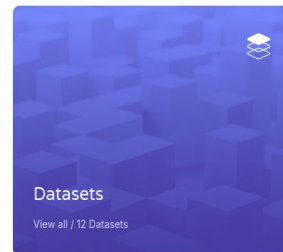
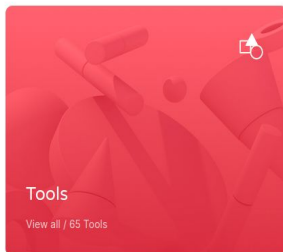
[Sign In](#)

Empowering Cultural Heritage through Artificial Intelligence

AI4Culture offers access to a pool of curated AI software tools, datasets and capacity building material.

[Learn More](#)

or [Contribute to AI4Culture](#)





The results:

73 Tools
15 Datasets
24 Upskilling Materials
5 recipes



The learnings:

Commitment to open-source principles is key, need to empower organisations to customise and fine-tune tools, importance of a hands-on approach

What we're learning




Next steps:


We look forward to enhancing the early results of this ongoing project with new tools and datasets contributed by users at ai4culture.eu

Thank you!

COMMON EUROPEAN
DATA SPACE FOR
CULTURAL HERITAGE



 pro.europeana.eu

 @EuropeanaEU