



What Got Us Here, Won't Get Us There

15th International Conference of the European Academy of Design
ONLINE and in PERSON in Brazil, Finland, India, Spain and the UK.
16-20 October 2023

The archive of the future: Challenges and possibilities for digital collections

Renata Perim^{a*}, Giselle Beiguelman^b

^a Faculdade de Arquitetura e Urbanismo da Universidade de São Paulo

^b Faculdade de Arquitetura e Urbanismo da Universidade de São Paulo

*Corresponding author e-mail: renataperim@usp.br

Abstract: Cultural contents from museums, galleries and other memory institutions acquire new characteristics when organized to be preserved in a digital context. This content begins to form part of a dynamic and networked space through its links that connect to other information and people. It is proposed in this article to outline the ways in which some objects of cultural heritage have been made available on the web in order to indicate new possible dynamics from the use of Artificial Intelligence (AI). Through exploratory research with a case study of AI in a museum, it was possible to envision ways to build a new digital repository that is formed from the analysis between theory and practice. The possibilities that AI offers, as well as its limitations, will be pointed out to propose alternative ways to create a digital repository as an ecosystem of dynamic information.

Keywords: Artificial Intelligence, Museum Collections, Archives, Network

1. Introduction

The model of cataloguing artistic objects and cultural heritage must be understood from a different logic when we think of the space of the website. This space deals with networked objects that can be connected to other information, making room for collaboration, updating, and adaptation. This dynamic opens a field of possibilities for the management of digital collections, however it also presents difficulties for repositories that do not have functionalities capable of aggregating this dynamic information.

We propose in this paper to outline the paths that Artificial Intelligence (AI) can offer for the field of digital archives and collection management on the web, also observing the challenges presented in the implementation of this system. This research, of an initial nature, is part of a Fapesp Project called *Digital Collection and Research* that has Prof^a Dr^a Giselle Beiguelman as responsible researcher; Prof^a Dr^a Ana Gonçalves Magalhães, Prof^a Dr^a Priscila Almeida Cunha Arantes, Prof. Dr. Dalton Lopes Martins and Prof. Dr. Eduardo Augusto Costa, as principal researchers; and Paula Monseff Perissinotto as Associate Researcher. Within the scope of this project, we seek to understand how the resources of AI can contribute to the elaboration of methods of classification

and cataloguing of artistic works, and thus, contribute to the formation of a critical discourse in the fields of Design and Art History in dialogue with the Digital Humanities.

2. Digitization of collections

The artistic object, when transformed into a digital image, will consist of layers that allow its interrelationship and connection with other media, based on mathematical attributes (Beiguelman, 2021). We must understand the digitized image as a tangled object, likely to be found through links. The digital space then becomes a meeting place for people and content. However, we see some images of artistic works on museum websites in a static way and with a traditional cataloging model – Figure 1.



Figure 1. Screenshot of the digital collection of the Ipiranga museum, highlighting the work *Rancho na Estrada de Sorocaba, 1830* by Adrien Henri Vital Van Emelen. Source: <https://acervoonline.mp.usp.br/iconografia/rancho-na-estrada-de-sorocaba-1830/>

Although the current modeling of the Ipiranga museum page is organized according to web design principles and presents good functionalities, we see in this example what happens in many digital repositories: a static information system; little or no permission for social participation or collaboration; does not allow automatic and semi-automatic documentation by machine learning algorithms. The latter, the part we focus on in this paper, will be presented below based on two museum projects that apply AI in their collections.

3. AI as a way of searching digital archives

According to Jaillant, “artificial Intelligence (AI) is a large concept designating the creation of intelligent machines that can simulate human thinking capability and behaviour (Jaillant, 2022, p. 7). One way to define AI in the context of this work is to understand the possibility that a system has to search, recommend, and generate content from cultural archives. This process takes place through

machine learning which is a subset of AI that allows machines to learn from data without being directly programmed. Moreschi, et al, indicate that “in this process, the algorithms of a given system are trained to learn to recognize patterns, and then be able to reproduce them when presented with new input data” (Moreschi, et al, 2022, p. 206). Part of the machine learning process takes place through image tagging, which consists of selecting elements from a given image, naming, and categorizing these elements. It is through this process that machines are able to “see” images and, based on this, can name and relate large sets of images.

The technology department at the Harvard Art Museums built a dataset using machine-generated descriptions and tags on parts of their digital holdings. For each artwork, the team collected data generated from AI and computer vision services such as Imagma, Google Vision, and Amazon Rekognition. When entering a keyword on the site, we find works that contain a corresponding tag generated by machine, another way is to access the images arranged in thumbnails on the main page. From there, we can go to an individual artistic work to see and compare the tags of the five AI services – Figure 2.

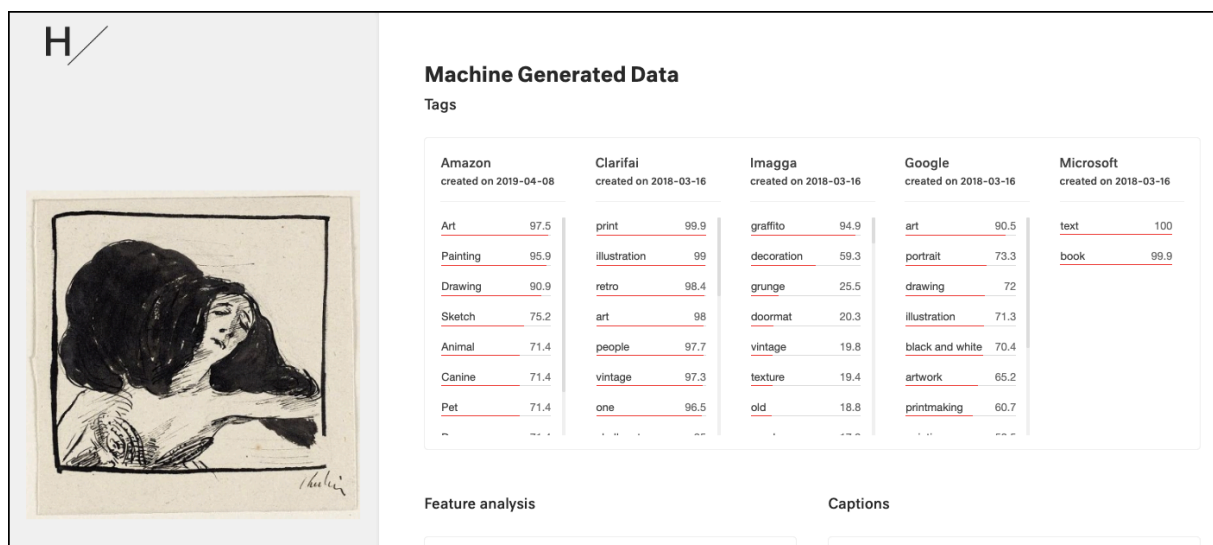


Figure 2. Screenshot of the digital collection of the Harvard Art Museums, highlighting the work *Head of a Woman*, 1910. Do artista Alfred Kubin, and the tags generated by five different computer vision services. Fonte: <https://ai.harvardartmuseums.org/object/223395>

When explaining the proposal, the museum reveals that the use of AI and computer vision offers a different look for curators since the computer lacks context or training in art history and indicates a non-specialist view. Another proposed point of view is precisely to reveal that computer vision services operate in a closed manner, “in a black box”, and do not disclose the algorithms and training sets used in their systems. When using the services and presenting the data, it is possible to notice possible errors or biases that occur in this type of service.

In fact, it is curious to note that, in the research with the work “Head of a woman”, none of the computer vision services indicated the tag “woman” and we see the indication of tags such as “animal”, “pet” and “canine”. It was also possible to see the emphasis on suggestions of materials and categories to which the work could belong, with indications of “print”, “graffito”, “drawing”, “portrait”. Looking at these tags not as misinterpretations, but as ways the algorithm was trained, and combining machine and human generated tagging can enrich the process and make collection items more accessible to the public.

Another example that we can highlight is MosAIC. An algorithm developed by researchers from MIT and Microsoft to discover hidden connections between paintings in the Metropolitan Museum of Art

(the Met) and in the Rijksmuseum in Amsterdam. The algorithm discovers similarities not only in patterns and shapes, but also in meaning and theme, and matches across different cultures and media. When entering the site, some works appear as thumbnails and, when clicking on one of them, it is possible to quickly see the suggested combination – Figure 3.

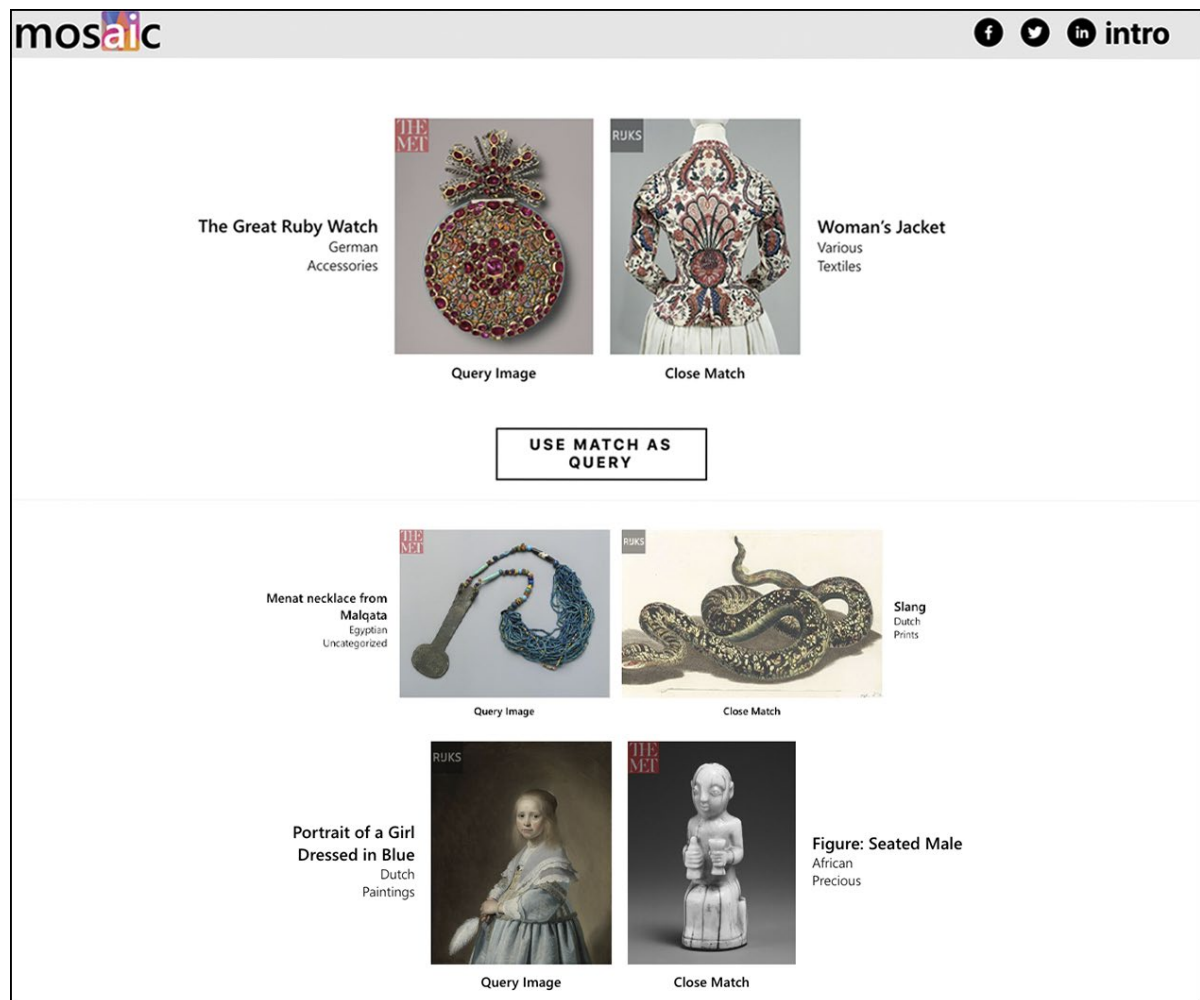


Figure 4. Three pairs of automatic combinations of artistic objects from The Met and Rijksmuseum. Image adapted and edited by the author from the MosAlc application web page. Source: <https://microsoft.github.io/art/app>

To find analogous images across cultures, the team used a classification algorithm, a data structure that groups similar images together in a tree-like structure. To find a close match, they start at the “trunk” of the tree and follow the most promising “branch” until they are sure they have found the closest image (Gordon, 2020).

It is worth mentioning that the starting point for the creation of the algorithm was an exhibition at the Rijksmuseum in Amsterdam that presented the unlikely combination of two works (Francisco de Zurbarán's *The Martyrdom of Saint Serapion* and Jan Asselijn's *The Threatened Swan*) (Gordon, 2020). In this case, the curators of an exhibition inspired the algorithm. Enabling relationships between different works, artists, and periods, is a curatorship effort. Historian Didi-Huberman (2003) indicates that this is a practice of anachronism, that is, offer the viewer the idea that the work of art is seen as an object that transcends its own time. Observing the combinations suggested by MosAlc instigates these relationships. Therefore, AI systems can inspire curators and it is through this path of exchanges, between humans and machines, that our project also seeks inspiration.

Researchers Elena Villaespesa and Oonagh Murphy, in extensive research on projects for the use of AI in museums (2021), indicate that although computer vision projects are seen as a positive contribution, the vast majority have not yet been implemented in official museum websites.

We understand that we are at an interesting point to think about alternative interpretations based on works from digital collections. AI can be a powerful tool if we know how to use it with balance. According to Crawford “classificatory schemas enact and support the structures of power that formed them, and these do not shift without considerable effort” (Crawford, 2021, p. 149). By understanding these systems, making them less opaque, we can provide an alternative practice that brings benefits to users. In the references presented, we saw that computer vision algorithms can suggest, in their classification, works for initial research. When analyzing this dynamic, some questions arise: How to select and classify works for the digital collection? How to insert diversified information about the works? How lesser known artists can be found? Starting the construction of a digital repository from questions like these indicates mobilizing a set of practices that should be part of a system of documentation of cultural objects – Figure 4.

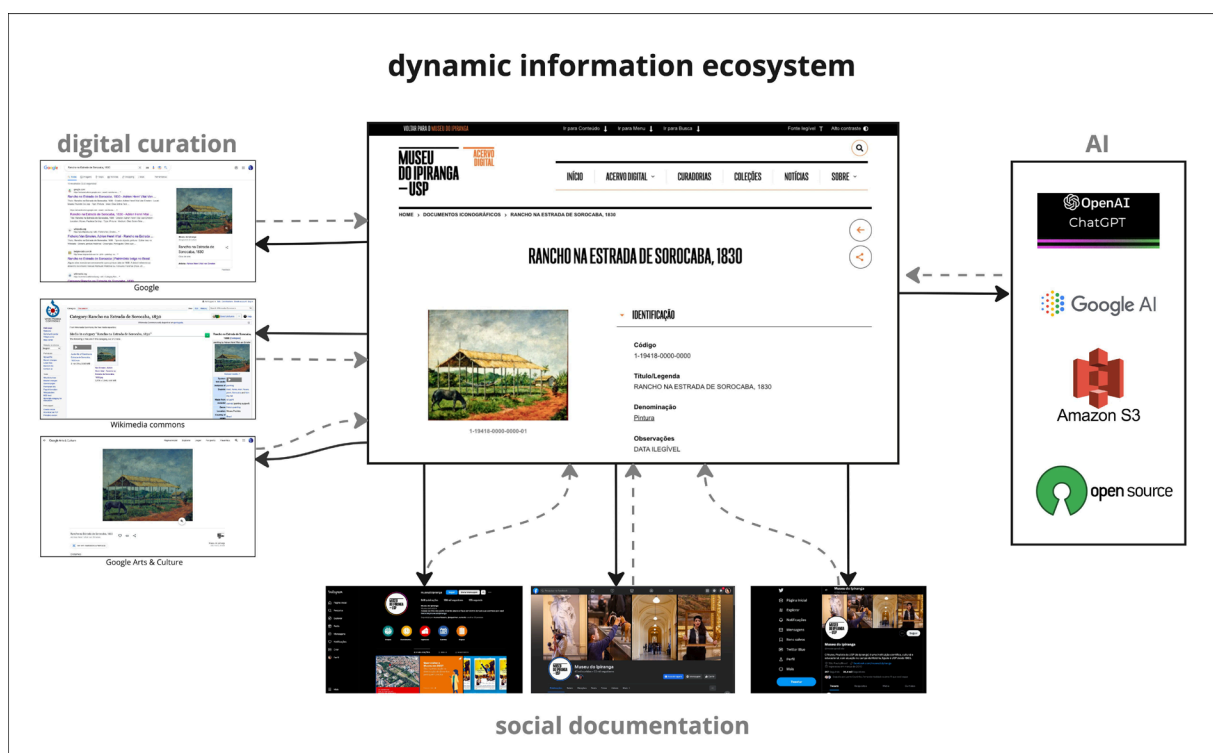


Figure 4. Diagram showing the flow of information between a museum's website and other information channels that can feed it, thus creating an information ecosystem. Source: created by the author.

We propose to highlight some elements that can compose a dynamic information ecosystem:

- Digital curation – allowing the involvement of other actors in the construction of documentation;
- Social documentation – incorporating socialized documentation in different media;
- Generative AI – automatic and semi-automatic documentation production by machine learning algorithms.

4. AI as part of a dynamic information ecosystem

Despite the advantages of using AI for organizing archives and cultural content, there are several challenges to be faced. We note that one of the main ones is data darwinism, which refers to the

process of selecting and valuing data that is processed within AI algorithms. It is necessary to understand how the machine understands and intervene editorially in the dynamic automation system to prevent algorithms from reproducing prejudices and discrimination present in society. It is important to think of tools that can be implemented to identify biased data and correct these distortions, mitigating possible consequences of a “cultural labor” (Velkova & Kaun, 2011, p. 535), such as classifying, evaluating, and making decisions, done by the algorithms.

Another challenge is how to think about the cataloging of documentation to be done. It is possible to think of an archive that starts from the refusal of storage and that privileges dynamism and flexibility, however this means giving up conservation in the management process. This is a principle contested by Hal Foster in “The Archives Without Museums” (2004), who argues that it is important to preserve and conserve archives to ensure historical continuity and cultural memory.

Therefore, it is necessary to find a balance between dynamism and conservation, thinking about flexible and adaptable architectures that allow the customization of the user experience, but that also guarantee the preservation and historical continuity of the contents.

References

- Beiguelman, G. (2021). *Políticas da Imagem: Vigilância e resistência na dadosfera* [Image Policies: Surveillance and Resistance in the Datasphere]. São Paulo: Ubu Editora.
- Crawford, K. (2021). *Atlas of Ai: power, politics, and the planetary costs of artificial intelligence*. New Haven: Yale University Press.
- Didi-Huberman, G. (2003). Before the Image, Before Time : The Sovereignty of Anachronism. In C. Farago, & R. Zwijnenberg (Eds.), *Compelling Visuality*. Minnesota: University of Minnesota Press.
- Foster, H. (1996). The archive without museums. *October*, 77, 97–119.
<https://doi.org/10.2307/778962>
- Gordon, R. (2020, July 29). *Algorithm finds hidden connections between paintings at the Met*. MIT News. Retrieved March 21, 2023, from <https://news.mit.edu/2020/algorithm-finds-hidden-connections-between-paintings-met-museum-0729>
- Jaillant, L. (2022). Introduction. In L. Jaillant (Ed.), *Archives, Access and Artificial Intelligence: Working with Born-Digital and Digitized Archival Collections* (pp. 7–28). Bielefeld: Bielefeld University Press.
- Moreschi, B., Jurno, A., & Beiguelman, G. (2022). Historical continuum and normatization in art collections and datasets: experiments with artificial intelligence at paulista museum. *MODOS*, 6(2), 202–267. <https://doi.org/10.20396/modos.v6i2.8667715>
- Velkova, J., & Kaun, A. (2021). Algorithmic resistance: media practices and the politics of repair. *Information, Communication & Society*, 24(4), 523–540.
<https://doi.org/10.1080/1369118X.2019.1657162>
- Villaespesa, E., & Murphy, O. (2021). This is not an apple! Benefits and challenges of applying computer vision to museum collections. *Museum Management and Curatorship*, 36(4), 362–383.
<https://doi.org/10.1080/09647775.2021.1873827>

About the Authors:

Renata Perim is a researcher and art director. Postdoctoral fellow at FAUUSP, supervised by Giselle Beiguelman. She is a member of the Fapesp Thematic Project Digital Archives and Research. She has a PhD in Design and a Master's in Visual Arts.

Giselle Beiguelman is an artist and professor at FAU-USP. She is the author of *Image Politics: Surveillance and Resistance in the Datasphere* and *Memory of Amnesia: Politics of Oblivion*, among others. His artistic works are part of museum collections in Brazil and abroad.

Acknowledgements: The main references, thought and proposals presented in this article came from the meetings that took place within the scope of the Thematic Project Digital Collections and Research and from the readings and mediation developed by its participants mentioned at the beginning of this paper.