Jessamyn Gibert
Professor Matthew Mayernik
LIS 545
March 11, 2024

LIS 545 Final Report

Data

This dataset titled “MoMA Artworks on View” shows every work of art that was on view at the Museum of Modern Art (MoMA) on February 1st 2022. The information was taken directly from the MoMA website, where they list all of the works currently on view. The data includes one file in the form of an excel spreadsheet (.xlsx). There is no specific software required to open or analyze the file – it could be opened in Excel, Google Sheets, or simply viewed off the Kaggle website. Under “license” it is listed as public domain, so there does not seem to be any usage restrictions. The spreadsheet contains information on 1,302 works of art, with 14 columns that provide various details. The columns are labeled: artist, title, description, year, medium, dimensions, copyright, credit, department, floor, room, image hyperlink, art ID, and view status. The location categories refer to where in the MoMA the works were on view, and the image hyperlinks direct to the MoMA website. “Credit” refers to which collection the work is loaned from, or where it was acquired if a part of the MoMA permanent collection.

Metadata

In terms of the metadata of the excel spreadsheet file itself, there is limited information on the Kaggle website. There is a small “about” section indicating that the data was collected from the MoMA website, and that it lists all works on view during February 1, 2022. The Kaggle owner of the file is listed as well, which is assumed to also be the creator of the spreadsheet. Not included is a DOI citation, detailed collection methodology, or date of creation. This information is structured according to the Kaggle standard (each dataset page on this site has the same

format), but not all of the information is complete. Considering the metadata of the individual works of art, the spreadsheet provides thorough information. Each of the 1,302 pieces of art are listed not only by their title and author, but also by a description, year of creation, medium, dimensions, copyright, credit, department, location, image and ID. All of this is included in the spreadsheet, and is clearly labeled and organized.

Discussion

This data could be enriched in several ways. To improve discoverability on the Kaggle site, more tags could be added to help the dataset appear in more searches. I was looking for a dataset related to art that fit the assignment parameters, but it took me a while to find this one. It currently has the tags, “Arts and Entertainment,” “Text,” and “Art,” but I would also include “Museums,” “Artists,” “Photography,” “Art History,” and “Exhibit.” This would allow more users to discover the dataset. Additionally, it would be helpful to include more information in the “About” section. Right now it reads, “This dataset contains data for all the works currently on view at the Museum of Modern Art as of February 1st 2022. The data includes information such as the medium of the art piece, the artist, a description of the art piece, dimensions of the art, etc.” Additional information about the dataset’s purpose and data collection process would be helpful in understanding reasons behind the spreadsheet’s organization. There is also no clear system for how the data is listed – none of the sections are in chronological or alphabetical order. I think the best way to list these would be alphabetically by artist, followed by year of creation. Something I would like to see in this dataset is metadata related to the artists’ demographics. Levels of diversity in museums has become a major area of contention in the art world. At this point in time, museums are being asked to evaluate the diversity of their collections and reckon with the deeply colonial history of museums in the United States. The inclusion of information

on gender, race, and nationality of each artist would be greatly beneficial to individuals researching levels of diversity in art museums. I would also like to see data on which exhibitions these works were a part of – this information can be extremely valuable to Art History research. The section titled "View Status" seems unnecessary: being on view as of February 1, 2022 is the common denominator for every entry in the dataset. The only options for this column are "on view" and "new on view," but there is no information describing what "new on view" specifically means.

Publications

When doing a Google search for the title of this dataset, the only results are from the Kaggle site where the spreadsheet is stored. There are no publications listed or provided with the dataset. The only source cited is the collection page on the MoMA website, where the works currently on view can be seen.

A**rtstor Overview**

After looking through several data repositories via re3data.org, Artstor seems to be the most appropriate for my data set. Artstor is a partner of JSTOR and under the organization ITHAKA. It houses over 2 million images across about 300 art and architecture collections. These collections are divided into two categories: core collections and public collections. The public collections can be viewed by anyone, and the core collections can be viewed by those with an account or who are associated with a university or other institution that subscribes to JSTOR. Core collections contain collections of larger name museums and institutions, such as the Ivy League schools, the MoMA, and the Metropolitan Museum of Art. Public collections have a huge range of smaller collections across the United States. Each collection is listed by the name of the institution followed by the name of the collection. Upon selecting a collection, the site

directs you to a list of all the images in that collection. When clicking on an image, the user is brought to a page with the image along with pieces of metadata – title, medium, date, location, collection, ID number, source, rights, license, and file properties. This format would complement my dataset, as each data entry includes similar metadata elements as is listed in the Artstor entries. Artstor would also be an appropriate match for the data set because it focuses on the visual arts. The fact that there are MoMA collections already listed is further indication that this would be a good fit. It could be easily titled to match the existing Artstor format: "Museum of Modern Art: Works on view on February 1, 2022."

Contributing to Artstor

Artstor has a webpage dedicated to outlining the guidelines for contributing to the repository. Anyone can submit data for consideration, but there are specific parameters for what will be accepted. On their website it reads, "The collections in the Artstor Digital Library are community-built and community-contributed. We welcome collaboration from the international community to help us broaden access to visual materials for learning and scholarship." First, the individual pieces of data must be some work of art – this can be a painting, audio recording, video, photograph, etc. Second, the ownership rights of the works needs to be identified, and Artstor performs a legal review of the work to determine if they can actually store and display it on the site. Third, the data must already be in a digital format (i.e. not something that needs to be photographed or scanned). Images of works must be at least 72 ppi with 1032 pixels in length or width, but a higher resolution 300ppi and greater is strongly preferred. Materials can only be submitted after contacting Artstor first for more information. If they decide to move on with the process, Artstor will review all of the dataset before accepting it into the repository. Because of this format, there are no instructions for what should be in the submission information package.

Metadata

In terms of metadata, Artor has their own controlled list of classification terms, and uses country terms from the Getty Research Institute's Thesaurus of Geographic Names. They also use the Getty Research Institute's Union List of Artist Names. Artstor may add additional metadata to the dataset in order for it to best align with the other works in the repository. There does not seem to be requirements for how the metadata is submitted, but it is important to include artist, subjects covered, date range, medium, and location.

Accessibility

Logging into Artstor is required to view and download media from the "core collections," but there is no login necessary for works in the "public collections." Just like JSTOR, those who are associated with institutions that subscribe to Artstor are able to login and gain access. There does not seem to be a way for individuals not associated with such institutions to view core collection materials. Each work is displayed with the following metadata: title, medium, date, location, collection, ID number, source, rights, license, and file properties. However, the only option is to download the image of the work which is downloaded as a jpeg. There is no way to download the accompanying metadata.

Data Citation

For this specific dataset, I would assign the dataset a DOI and follow A Chicago Manual of Style format. This recommendation is because Chicago style is commonly used in the Art History field, and the Museum of Modern Art also uses this style. In their document titled "FORMAT FOR CITATIONS TO MATERIALS FROM THE MUSEUM OF MODERN ART ARCHIVES," the formats listed most closely align with the Chicago Manual of Style.

Preservation

Because this dataset is in the form of an excel spreadsheet and doesn't require any specific software to view, I would argue that its format is not in danger of becoming obsolete. Even if there is some future where excel or other spreadsheet programs become obsolete, the data itself is easily understood in a plain text or written format. The category in the dataset that is not readable text, is the links to the MoMA website where a user can view the images of the art pieces. I do not see this format becoming obsolete either, but for preservation reasons, I am positive that the MoMA keeps both physical and digital documentation of these collections. Furthermore, these artworks can be identified based on all of the additional information provided throughout the dataset, which will be enough to find either physical versions or digital photographs of them. If this dataset is the only place where documentation of these works are kept, I would suggest adding the jpeg images into the dataset, or addresses of the museums that hold these works in their permanent collections.

Copyright License

The question around a copyright license for this dataset is interesting, because all of the data points are works of art that have their own copyright license attached to them. I would argue that this dataset has to be open access and is not eligible for copyright, because it contains work that is copyrighted, and is information that is publicly available through the Museum of Modern Art website. If anyone had the authority to develop a copyright license, it would be the MoMA.

Privacy Statement

None of the data in this data set are anonymous, because they are derived from artists who have their work publicly on view at the Museum of Modern Art. Likewise, none of the information in it is sensitive for the same reason.

Works Cited:

Dataset: https://www.kaggle.com/datasets/jackogozaly/moma-artworks-on-view/data

MoMA Collection Page: https://www.moma.org/collection/works/

Artstor: https://library.artstor.org/#/home

Artstor Guidelines For Contributing:
https://www.artstor.org/contribute/guidelines-for-contributing

“How to Cite Datasets and Link to Publications”:
https://www.dcc.ac.uk/guidance/how-guides/cite-datasets#sec:building-infrastructure

MoMA Document, “Format For Citations To Materials From The Museum Of Modern Art Archives”: https://www.moma.org/momaorg/shared/pdfs/docs/learn/PermissionForm.pdf