

Team Name: **DataOmini**

Team Supervisor: **Bushra Kidwai**

Team Member Information

Member 1 Name: Xianjie (Iris) Ma UPenn Email: irisma@seas.upenn.edu Github Username: IrisMaxj	Member 3 Name: Xue(Sue) Liu UPenn Email: xliu28@seas.upenn.edu Github Username: xueliu28
Member 2 Name: Yinjie Zhang UPenn Email: yinjiez@seas.upenn.edu Github Username: yinjiez	Member 4 Name: Lingzi Xiao UPenn Email: lzxiao@seas.upenn.edu Github Username: lingzix

Description of our Web-Application design motivations and ideas:

We want to build this virtual art gallery to let people have a glimpse of the amazing artworks exhibited/hosted by National Gallery of Art (NGA). Especially in the pandemic era, with many national/international travel restrictions and inconvenience, we want to provide an alternative virtual gallery browsing experience to people who currently cannot visit the NGA onsite.

Meanwhile, this web-application also serves an educational purpose, as it provides an introduction to some of the finest artworks by a group of prominent artists around the world. For instance, our searching functionality will allow the application of “filters” to let users browse artworks by artists (possibly the donors of these artworks), by their geological origins, by the chronolog, etc.. We will also provide some analytical results of these artworks based on their historical/geographical/etc. associations. As these can provide more insights about the rationales and motivations that fostered the birth of these artworks.

We also wish our site to be a source of aesthetic inspiration for creators. We want to achieve this by tailoring the aesthetic designs of our webpages with the visual/color components of the artworks being displayed. (i.e. with the help of some amazing REACT-library tools such as “react-color-extractor” package). There will also be some interesting functionalities such as “to give the user an idea about the actual dimensions of artworks by asking the height of the user, and display some artworks with regarding to this height”. More interesting ideas are being brainstormed and discussed in process.

Lastly, we hope our virtual site experience can encourage people to make physical tours to NGA, which will complement them with a more holistic perceptual experience of these artworks.

Datasets Description Information:

Data Source: **National Gallery of Art (NGA)** Open Data

- <https://www.nga.gov/open-access-images/open-data.html>
- <https://github.com/NationalGalleryOfArt/opendata>

We are interested in using most but not all (i.e. **14** out of 17 relations) of the provided datasets* by NGA:

- Major Relations: “**objects.csv**”, “**constituents.csv**”
 - linked by “objects_constituents.csv”
- Supplementary source: “media_items.csv”

* Note:

1. Detailed explanation of the entire datasets are provided by NGA in their “Data Dictionary.docx” published here: <https://github.com/NationalGalleryOfArt/opendata/tree/main/documentation>
2. basic informations about the datasets of our interests are as updated by NGA on 2022-02-14
3. *due to the nature of artworks and its associated data features (most of them are semantic data), we cannot really generate a summary statistics for our relations’ attributes.*

Category #1: Artworks/Objects*

#1-“**objects.csv**” (55MB) = 136929 rows x 29 cols

Description: Art objects are a core entity of the collection extract. Objects are physical or logical constructs.

#2-“objects_constituents.csv” (17.7MB) = 467716 rows x 12 cols

Description: One of the most important elements of the extract, the object / constituent relationships describe the association between a constituent and an object.

#3-“published_images.csv” (28.7MB) = 103047 rows x 12 cols

Description: Images that have been published to NGA web properties. Not all the artworks hosted by NGA have a corresponding digital image.

#4-“objects_terms.csv” (14.3MB) = 386242 rows x 6 cols

Description: relationships between terms and objects

#5-“object_associations.csv” (599KB) = 24655 rows x 4 cols

Description: relationships between art objects.

#6-“objects_dimensions.csv” (9.5MB) = 206598 rows x 5 cols

Description: records various dimensions for objects

#7-“objects_text_entries” (33.9MB) = 208280 rows x 4 cols

Description: stores long texts associated with objects

#8-“objects_altnums.csv” (1.3MB) = 65075 cols x 3 cols

Description: relationships between art objects - see accompanying documentation about types of art objects and relationships between them

#9-“Objects_historical_data.csv” (447KB) = 5131 cols x 7 cols

Description: records a very light audit trail for certain important object properties

***Datasets Cleaning Intention:** The final collection of artworks/objects included in our database will be strictly less than the number of 136929 (the size of “objects.csv”) and 103047 (the size of “published_images.csv”). For the purpose of visual presentation, we will only include the artworks that have a publicly accessible digital image to be displayed on our website. Also, NGA has a variety of artworks collections, including sculptures, hand-crafted artworks like coins, woodworks etc., which are not in our major-scope of focus. We are most interested in paintings and drawings (i.e. 2D-artworks). Hence, we may reduce the size of our collections to only contain these 2D-artworks.

Category #2: Constituents of the art objects

#10-“**constituents.csv**” (2.3MB)= 18282 rows x 13 cols

Description: A constituent is a single person or a group of persons (see constituent type). Constituents in the extract typically have a relationship with one or more objects in the public objects extract. Relationships are defined in another table, but a constituent may be an artist, an owner, or a donor, or all of these.

#11-“constituents_text_entries.csv” (984KB) = 6891 rows x 4 cols

Description: stores long texts associated with constituents

#12-“constituents_altnames.csv” (1.3MB) = 23027 rows x 6 cols

Description: Alternative names used by or assigned to constituents for various purposes as indicated by the nametype column. Curators might prefer to use one name whereas the constituent might be referred to by in other publications or sources by a different name.

Category #3: Media* sources used by NGA to enhance the overall visiting/browsing experience

#13-“media_items.csv” (3.8MB)= 2621 rows x 3 cols

Description: a table containing audio and video items published to www.nga.gov

#14-“media_relationships.csv” (118KB) = 2969 rows x 3 cols

Description: records relationships between media items and core NGA art entities

***Note:** We are still in the stage of brainstorming how we can utilize these media sources to better enhance the user experience of our website (i.e. when a user is browsing our website, we can play musical pieces that are suiting the corresponding artworks or website-functionalities). So these “media”-related data may or may not be included in our final project actualization.

Anticipated Query Functionality:

Query #1:

- Search & Browse artworks by classification/subclassification and display relevant informations:

- Ex. artists/titles/medium/shape/theme/color/season/etc.
- Relevant query relations: “objects.csv”, “public_images.csv”, “objects_dimensions.csv”, “object_text_entries.csv”

Query #2:

- “Guess what you like”: art recommendation by inquiring users about their preferred color patterns/historical period/country
- Relevant query relations: “objects.csv”, “public_images.csv”

Query #3:

- View art collections: query and display artworks from the same collection/series/exhibits
- Relevant query relations: “objects.csv”, “public_images.csv”

Query #4:

- Prompt User for entering his/her height, and query for some artworks with real size in the perspective of user’s height
- Relevant query relations: “objects.csv”, “public_images.csv”, “objects_dimensions.csv”

Query #5:

- Display a collection of artworks (can be further filtered by user’s specification) in smallest->largest size order;
- Relevant query relations: “objects.csv”, “public_images.csv”, “objects_dimensions.csv”

Query #6:

- Artworks collections in different time span or originated regions
 - Similar function: List art pieces for the same artist and arrange by time course
- Relevant query relations: “objects.csv”, “public_images.csv”

Query #7:

- Analysis: How portraits have changed across time (Optional: user can specify a time range to look into)
- Relevant query relations: “objects.csv”, “public_images.csv”

Query #8:

- Analysis: Artworks by constituents (i.e. donor, artist, owner). We will perform some statistical calculation, and display the analysis result with some data visualization tools
- Relevant query relations: “objects.csv”, “object_constituents.csv”, “constituents.csv”