Portfolio 3 - assignment

By Mushtaba Osmani (wgq323), Christan Lauersen (bfd962) & Mickey Johansson (mvx394)

1. SMK's approach and reasoning for using an API

SMK have given an API approach to their data. This is for free use "to create apps and services that make use of SMK art (REFERENCE). Furthermore, SMK has started a project called "SMK Open" (https://www.smk.dk/article/smk-open/) which aims at making their art available to everybody. As is stands now, not everybody has easy access to a physical museum and even if one were to go through and see every single work in SMK, it would still be less than 1% of the entire collection. By digitizing the art, it would be possible to solve this problem and further improve other aspects such as making it possible to study the works in detail and continue development. It would also be possible to use the art in everything from research-and school projects to printing on posters or pillowcases.

2. Description of our URL and its content

The SMK api (https://api.smk.dk/api/v1/docs#/artworks/searchArt) was used to search for the key word "sculpture". This gave a result of 4006 times, which was cut down to 200 items in our JSON-file. The accompanying url was then copied into a JSON beautifier (https://codebeautify.org/jsonviewer) to easier research the content of the file.

a) Different types of metadata in each element:

Every element inherits a large amount of metadata, but this data does not match across every element. For example, some of the elements has "distinguishing_features", "labels" and images of the art.

The metadata that recurs on most of the elements include:

'Created' - A date resembling when the element was created in the API.

'Credit_line' - The name of the credited people, most often the artist.

‘Id’ - The ID of the artwork.

‘Object_names’ - The name of the artwork.

‘Production’ - Information about the production, including ‘creator’, ‘date’ and information about the artist.

Some of the elements do not have a “credit_line”, “production” or “production_date”, which could be because no such information exists about the artist behind the sculpture.

When flattening the data into a dataframe, we get 249 different columns, which each represents a different set of metadata. Screenshots of the entire list is are in appendix xx.
Many of the columns have missing values, since many elements have different metadata compared to the other elements. We made a rough clean-up of the data, by removing every column that had more than 25 rows with missing data. This gave us a dataframe with 23 columns, from which we chose specific columns that we found useful in our analysis. This picture shows the column names and types of the data we ended up using after cleaning it up.

```
nydf.dtypes
```

Out[574]:

creator	object
has_image	bool
name	object
birthyear	float64
nationality	object
py_start	float64
py_end	float64
public_domain	bool
language	object
title	object

dtype: object

b) Metadata comparison with Dublin Core

We will now attempt to expand on whether our metadata is in compliance with the Dublin Core Metadata Element Set. This element consists of fifteen properties that are used in resource description. The practice of the Dublin Core has been in play since 1998 and aims to standardize the process of creating metadata. The fifteen elements consist of the

following: contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject, title & type (DCIM Usage Board, 2012).

As we examine our JSON file we can argue that the metadata we work with is mostly in compliance with the fifteen elements of the Dublin Core. It has followed the format of including aspects such as creator, date, description, format, rights, title, and type. According to Dublin Core, the creator is the entity responsible for making the resource and in our case that is evidently clear that SMK has followed that process. In terms of date, there is also compliance with Dublin Core as there are start date, end date and period time frame of each sculpture. Our metadata has 'notes' or content description, which in our understanding is in compliance with the Dublin Core element of description. Most of the sculptures have NaN in place, however, a few have few notes of the sculpture. Format element is in our metadata named dimensions. Language is in our metadata named title_0_language and will is element 'language' include the language of the resource. Rights element in our metadata is following the Dublin Core as it is named 'rights' and includes rights information about various properties that are associated with the resource, which is exactly what the Dublin Core describes 'rights' element as being. In terms of title SMK's metadata is following Dublin Core's description of a 'title' element as well. Lastly, there is the element of 'type' which SMK again includes, however, in an alternate name that is 'materials' under the element of 'techniques'. The 'materials' row in our metadata consists of the materials that are used in the production of each sculpture.

In essence we can conclude that SMK's metadata is in compliance with Dublin Core's fifteen elements, however, it is not a full compliance, as there are aspects that are not followed in identical fashion whether it is excluding some elements or changing their names such as 'title_o_language.

c) Useful numbers for statistic calculations

There are a few relevant numbers one could choose to look at within the dataframe. For example, one could choose to look at the "production_date" column. This reveals from when the art is created and further, from what year/decade most sculptures are created. Together with "creator_date_of_birth" found in "production", it would be possible to find out how old most artists were when they created their art. By examining "production_date" and their "start" and "end" further, it would be possible to compare the "credit_line" (the artist) to find out how long it takes each artist to finish their work.

Additionally, with the “dimensions” column, one could inspect the average height and width of each sculpture. Perhaps compare it to the artist to see if there is a size-pattern, if the artist usually creates big or small sculptures.

d) Questions that this dataset could answer

The “credit_line” shows the name of the creator, meaning it is possible, with the value_counts() function, to see what artist has the most sculptures in SMK’s database. Furthermore, the “creator_nationality”, within the “production” column, enables one to count what nationality is represented the most. Additionally, in the “techniques”-column, one could look at what techniques and/or materials are used most often, when and by whom?

Bibliography

DCIM Usage Board (2012. Dublin Core Metadata Element Set, Version 1.1: Reference Description. Retrieved the 29th of November 2019 from: https://business.twitter.com/en/help/campaign-setup/create-a-tweet-engagement-campaign.html

Appendix

acquisition_date	object
acquisition_date_precision	object
alternative_images_0_height	float64
alternative_images_0_iiif_id	object
alternative_images_0_iiif_info	object
alternative_images_0_mime_type	object
alternative_images_0_native	object
alternative_images_0_orientation	object
alternative_images_0_size	float64
alternative_images_0_thumbnail	object
alternative_images_0_width	float64
alternative_images_1_height	float64
alternative_images_1_iiif_id	object
alternative_images_1_iiif_info	object
alternative_images_1_mime_type	object
alternative_images_1_native	object
alternative_images_1_orientation	object
alternative_images_1_size	float64
alternative_images_1_thumbnail	object
alternative_images_1_width	float64
alternative_images_2_height	float64
alternative_images_2_iiif_id	object
alternative_images_2_iiif_info	object
alternative_images_2_mime_type	object
alternative_images_2_native	object
alternative_images_2_orientation	object
alternative_images_2_size	float64
alternative_images_2_thumbnail	object
alternative_images_2_width	float64
alternative_images_3_height	float64
alternative_images_3_iiif_id	object
alternative_images_3_iiif_info	object
alternative_images_3_mime_type	object
alternative_images_3_native	object
alternative_images_3_orientation	object
alternative_images_3_size	float64
alternative_images_3_thumbnail	object
alternative_images_3_width	float64
alternative_images_4_height	float64
alternative_images_4_iiif_id	object
alternative_images_4_iiif_info	object
alternative_images_4_mime_type	object
alternative_images_4_native	object
alternative_images_4_orientation	object
alternative_images_4_size	float64
alternative_images_4_thumbnail	object
alternative_images_4_width	float64
alternative_images_5_height	float64
alternative_images_5_iiif_id	object
alternative_images_5_iiif_info	object
alternative_images_5_mime_type	object
alternative_images_5_native	object
alternative_images_5_orientation	object
alternative_images_5_size	float64
alternative_images_5_thumbnail	object
alternative_images_5_width	float64
alternative_images_6_height	float64
alternative_images_6_iiif_id	object
alternative_images_6_iiif_info	object
alternative_images_6_mime_type	object
alternative_images_6_native	object
alternative_images_6_orientation	object
alternative_images_6_size	float64
alternative_images_6_thumbnail	object
alternative_images_6_width	float64
alternative_images_7_height	float64
alternative_images_7_iiif_id	object
alternative_images_7_iiif_info	object
alternative_images_7_mime_type	object
alternative_images_7_native	object
alternative_images_7_orientation	object
alternative_images_7_size	float64
alternative_images_7_thumbnail	object
alternative_images_7_width	float64
alternative_images_8_height	float64
alternative_images_8_iiif_id	object
alternative_images_8_iiif_info	object
alternative_images_8_mime_type	object
alternative_images_8_native	object
alternative_images_8_orientation	object
alternative_images_8_size	float64
alternative_images_8_thumbnail	object
alternative_images_8_width	float64
alternative_images_9_height	float64
alternative_images_9_iiif_id	object
alternative_images_9_iiif_info	object
alternative_images_9_mime_type	object
alternative_images_9_native	object
alternative_images_9_orientation	object
alternative_images_9_size	float64
alternative_images_9_thumbnail	object
alternative_images_9_width	float64
colors_0	object
colors_1	object
colors_2	object
colors_3	object
colors_4	object
content_description_0	object
created	object
credit_line_0	object
current_location_name	object
dimensions_0_notes	object
dimensions_0_part	object
dimensions_0_type	object
dimensions_0_unit	object
dimensions_0_value	object
dimensions_1_notes	object
dimensions_1_part	object
dimensions_1_type	object
dimensions_1_unit	object
dimensions_1_value	object
dimensions_2_notes	object
dimensions_2_part	object
dimensions_2_type	object
dimensions_2_unit	object
dimensions_2_value	object
dimensions_3_notes	object
dimensions_3_part	object
dimensions_3_type	object
dimensions_3_unit	object

```
dimensions_3_value                  object
dimensions_4_notes                  object
dimensions_4_part                   object
dimensions_4_type                   object
dimensions_4_unit                   object
dimensions_4_value                  object
dimensions_5_notes                  object
dimensions_5_part                   object
dimensions_5_type                   object
dimensions_5_unit                   object
dimensions_5_value                  object
distinguishing_features_0           object
distinguishing_features_1           object
exhibitions_0_date_end              object
exhibitions_0_date_start            object
exhibitions_0_exhibition            object
frame_notes_0                       object
frame_notes_1                       object
has_image                             bool
id                                  object
iiif_manifest                       object
image_cropped                       object
image_height                       float64
image_iiif_id                       object
image_iiif_info                     object
image_mime_type                     object
image_native                        object
image_orientation                   object
image_size                         float64
image_thumbnail                     object
```

```
image_width                        float64
labels_0_date                       object
labels_0_source                     object
labels_0_text                       object
labels_0_type                       object
labels_1_date                       object
labels_1_source                     object
labels_1_text                       object
labels_1_type                       object
materials_0_material                object
materials_1_material                object
materials_2_material                object
materials_3_material                object
materials_4_material                object
materials_5_material                object
materials_6_material                object
materials_7_material                object
modified                            object
notes_0                             object
notes_1                             object
notes_2                             object
notes_3                             object
notes_4                             object
number_of_parts                    float64
object_history_note_0               object
object_names_0_name                 object
object_number                       object
on_display                            bool
part_of_0                           object
part_of_1                           object
```

```
part_of_2                           object
part_of_3                           object
part_of_4                           object
part_of_5                           object
part_of_6                           object
parts_0                             object
parts_1                             object
parts_10                            object
parts_11                            object
parts_12                            object
parts_13                            object
parts_14                            object
parts_15                            object
parts_16                            object
parts_17                            object
parts_18                            object
parts_19                            object
parts_2                             object
parts_20                            object
parts_21                            object
parts_22                            object
parts_3                             object
parts_4                             object
parts_5                             object
parts_6                             object
parts_7                             object
parts_8                             object
parts_9                             object
production_0_creator                object
production_0_creator_date_of_birth  object
```

```
production_0_creator_date_of_death  object
production_0_creator_history        object
production_0_creator_lref           object
production_0_creator_nationality    object
production_0_place                  object
production_1_creator                object
production_1_creator_date_of_birth  object
production_1_creator_date_of_death  object
production_1_creator_history        object
production_1_creator_lref           object
production_1_creator_nationality    object
production_date_0_end               object
production_date_0_period            object
production_date_0_start             object
production_date_1_end               object
production_date_1_period            object
production_date_1_start             object
production_dates_notes_0            object
production_dates_notes_1            object
public_domain                         bool
related_objects_0_notes             object
responsible_department              object
rights                              object
techniques_0_technique              object
titles_0_language                   object
titles_0_notes                      object
titles_0_title                      object
titles_0_translation                object
titles_0_type                       object
titles_1_language                   object
```

```
titles_1_notes                      object
titles_1_title                      object
titles_1_translation                object
titles_1_type                       object
titles_2_language                   object
titles_2_notes                      object
titles_2_title                      object
titles_2_translation                object
titles_2_type                       object
dtype: object
```