DMP title

Project Name ASM FWO DMP - DMP title

Project Identifier u0124860

Grant Title 11C9522N

Principal Investigator / Researcher Anne-Sofie Maerten

Project Data Contact annesofie.maerten@kuleuven.be

Description In this PhD project, aesthetic ratings are collected for a large set of paintings and artistic photographs. In addition, demographic information of the raters is collected. This will result in a large (anonymized) dataset that will be used to train deep neural networks to predict aesthetic preferences for certain groups.

Institution KU Leuven

1. General Information
Name applicant

Anne-Sofie Maerten

FWO Project Number & Title

11C9522N

Beauty is in the eye of the algorithm: understanding what drives individual and group differences in aesthetic preferences.

Affiliation

- KU Leuven

2. Data description
Will you generate/collect new data and/or make use of existing data?

- Generate new data

Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).

type of data: images of artworks (from WikiArt) and artistic photographs (from FLICKR), annotated by multiple raters on their aesthetic value and for a smaller set their perceived complexity and order.

format: .jpg, .csv

volume: approx 1TB

how created: images are scraped from WikiArt and FLICKR (freely available) and annotated in an online study on the platform Prolific.

3. Legal and ethical issues
Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

- Yes

Privacy Registry Reference: G-2021-4569

Short description of the kind of personal data that will be used: The personal data that are collected include age, gender, nationality, education level and some art interest and knowledge questionnaires. All data is gathered anonymously through the platform Prolific, meaning that we only have the answers to the demographic questions and prolific only has the contact information of the participants to be able to reimburse them. Therefore, the data can not be linked to the individual that provided the data.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, add the reference to the formal

approval by the relevant ethical review committee(s)

- Yes

[G-2021-4569](G-2021-4569)

Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?

- No

Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

- Yes

Some of the artworks are protected by copyrights, and can therefore not be used as illustration in published papers. A copyright claim will be shown before granting access to the dataset.

4. Documentation and metadata
What documentation will be provided to enable reuse of the data collected/generated in this project?

For the large-scale artwork and artistic photograph dataset the following metadata will be stored: url to the image, name of the work, artist, different style categorizations, content, size in pixels, the aesthetic rating by multiple annotators, for a small set the perceived complexity and order ratings, demographics of the annotators.

The methodology and protocol of the online study will be described in a paper. A ReadMe file of the image collection will be written.

Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

- No

For the large-scale artwork and artistic photograph dataset the following metadata will be stored: url to the image, name of the work, artist, different style categorizations, content, size in pixels, the aesthetic rating by multiple annotators, for a small set the perceived complexity and order ratings, demographics of the annotators. These will be storen in a .csv file and will be saved on an external hard drive as well as on the online server of the lab.

5. Data storage and backup during the FWO project
Where will the data be stored?

The dataset will be stored on an external hard drive that will be saved in the lab, as well as on the online server of the lab.

How is backup of the data provided?

The data is partially stored on OneDrive and partially on the online server of the lab.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available then explain how this will be taken care of.

- Yes

I bought an external hard drive to ensure enough storage capacity for the data.

What are the expected costs for data storage and back up during the project? How will these costs be covered?

The external hard drive was bought with part of my bench fee. If additional unexpected costs are required, these will also be covered by my bench fee.

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The data is fully anonymised and will be shared upon request with other researchers (after they sign the copyright claim and agree to the use solely for research purposes). The data will be stored on the online server of the lab which is protected according to the KU Leuven standards.

6. Data preservation after the FWO project
Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).

The large-scale bench mark dataset will be retained for the expected 5 year period and many more years. The dataset will be publicly available (upon request) to other researchers. All aspects of the dataset will be retained and available since we receive anonymised data through the platform Prolific.

Where will the data be archived (= stored for the longer term)?

The dataset will be stored on an external hard-drive as well as on the online sever of the lab, conform with the KU Leuven RDM policy for protection of data.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

The external hard drive to save the dataset on will be covered by my FWO bench fee. The cost for storage on the online server are covered by the lab and the PI's funds and resources.

7. Data sharing and reuse
Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

- Yes. Specify:

The dataset will only be shared with third parties for research purposes, after they sign a copyright claim for the images in the dataset. Apart from that, the dataset is fully anonymised and does not contain sensitive personal information, therefore we do not see concerns in sharing the data with other researchers.

Which data will be made available after the end of the project?

The full dataset will be available online upon request.

Where/how will the data be made available for reuse?

- In a restricted access repository

The data will be available on request after signing a copyright claim on a website for the project.

When will the data be made available?

- Upon publication of the research results

Who will be able to access the data and under what conditions?

Third parties can access the data for research purposes upon request and after signing a copyright claim.

What are the expected costs for data sharing? How will the costs be covered?

There aren't any specific costs linked to the sharing of the data.

8. Responsibilities
Who will be responsible for data documentation & metadata?

Anne-Sofie Maerten (PhD)

Who will be responsible for data storage & back up during the project?

Anne-Sofie Maerten (PhD)

Who will be responsible for ensuring data preservation and reuse ?

Anne-Sofie Maerten (PhD) and Prof. Johan Wagemans (PI).

Who bears the end responsibility for updating & implementing this DMP?

The PI bears the end responsibility of updating & implementing this DMP.