

GENERAL

Various data collection methods are employed depending on the specific experiments planned within the ERC project:

Data(set)	Description	New or reused	File format	Estimated max volume
Aesthetic preferences for images	Online studies; Preference ratings of images collected in experiments programmed in the jsPsych framework	New	Raw image data is exported in .json and preference ratings are saved in .csv for further analyses.	<1GB
Drawings/Notes from focus groups with artists and teachers in art and design	Pen and paper data collected in focus groups will be scanned and preserved as pdf files	New	.pdf	<1TB
Recordings from focus groups with artists and teachers in art and design	Audio- and video recordings of focus groups	New	Stored as .mp3 or .mp4-files	<1TB
Psychophysical data	Psychophysical data collected in in-lab experiments developed in the PsychoPy framework	New	Raw data exported as .csv files	<1GB
Questionnaire data	Collected on a computer or tablet, or pen-and-paper questionnaires	New	Data collected on computer or tablet are stored as .csv or .txt-files; paper questionnaires will be scanned and stored as .pdf-files.	
Eye tracking data	Collected during in-lab experiments using the SR research eye tracking	New	Raw data stored as .edf. Since the raw data are compressed and encoded to be only	< 1 TB

	equipment.		<p>readable with the manufacturer's analysis software (dataviewer), we also provide datafiles in a text format.</p> <p>If applicable, parameters used for preprocessing are described in a readme file.</p>	
	Collected using the Tobii Glasses mobile eye tracking devices during museum exhibitions	New	<p>Raw data stored in folders with structure for Tobii eyetrackers (including compressed data in .json-files and videos in .mkv format from the scene camera).</p> <p>Since the raw data are compressed and encoded to be only readable with the manufacturer's analysis software (Tobii Pro Lab), we also provide datafiles in a text format.</p> <p>If applicable, parameters used for possible preprocessing are described in a readme file.</p>	< 1 TB
Model source code	In addition to training and evaluation, the model source code is used to predict aesthetic scores.	New and reused: specifies whether the source code is newly written or reused, including pretrained models and existing codes.	Typically provided in Python (.py) or Jupyter notebook (.ipynb) files. Additionally, a ReadMe file (.me) explaining how to use the code.	< 1 TB
Trained model	Metadata and training details of	New	Stored in Python frameworks such as	< 1 TB

	the saved model after training, including all learned parameters and weights of the model.		PyTorch, Keras, or Tensorflow.	
Generated datasets	Datasets that have been created during the project, including any synthetic data, data transformations applied to original datasets, or artificial images generated by AI tools.	New	Images are described in .csv-files, or stored as .jpeg, .png or .pdf-files.	< 1 TB
(Annotated) image datasets	Datasets with behavioral ratings used for model training.	New (e.g. LAPIS)	Images are described in .csv-files, or stored as .jpeg, .png or .pdf-files.	< 1 TB
Benchmark datasets	Standardized datasets, publicly available, used for evaluating and comparing the performance of models.	<p>Reused</p> <p>Benchmark datasets for aesthetics: JenAesthetics, PARA, TAD66k, ...</p> <p>Benchmark datasets for natural images: MNIST, ImageNet, ...</p> <p>More may be added during the course of the project.</p>	Images are described in .csv-files, or stored as .jpeg, .png or .pdf-files.	< 1 TB

1. MAKING DATA FINDABLE (*dataset description: metadata, persistent and unique identifiers e.g., DOI*)

During the research project, the research data are stored on the researchers' computer and software scripts are uploaded on the Gitlab repository hosted by KU Leuven. The computers and storage solutions are managed by KU Leuven which means that all data is backed up automatically.

Upon completion of a study, data will be stored on a KU Leuven long-term data storage solution (e.g., Sharepoint) using a similar folder structure within each study. Additionally, all data, code and material will be handed to the PI upon completion of a study.

Furthermore, the research data will be made available via de KU Leuven data repository RDR or OSF where they will be assigned a unique DOI for further accessibility. Both RDR and OSF use metadata to facilitate findability of data.

2. MAKING DATA OPENLY ACCESSIBLE *(which data will be made openly available and if some datasets remain closed, the reasons for not giving access; where the data and associated metadata, documentation and code are deposited (repository?); how the data can be accessed (are relevant software tools/methods provided?)*

Data from online experiments, computer-based experiments, questionnaires and eye-tracking-experiments will be anonymized and shared via the RDR managed by the KU Leuven or the osf.io platform (field standard). Next to the research data file, we will share –as far as possible- stimuli, experiment programs, analysis code and extensive documentation of the data files. If full anonymization is not possible, the datasets will only be shared upon request.

Code used for machine learning studies is stored on GitLab. The trained models, datasets, and training details are shared via the RDR managed by the KU Leuven or the osf.io platform (field standard).

For scientific publications, open access will be guaranteed either through the journal's open access system or through publishing the preproof publications on the open access system of the KU Leuven (i.e., Lirias).

3. MAKING DATA INTEROPERABLE *(which standard or field-specific data and metadata vocabularies and methods will be used)*

The anonymized data and/or the data formatted as text or .csv files are shared (depending on the size of the raw datafiles). Data will be deposited in a format that can be accessible for everyone whenever possible. Most of our experiments yield data that can be represented in .csv file format. Datasets consist of images represented in common image file formats (e.g., .jpeg, .png, ...).

We will use relevant metadata descriptor standards that apply to each specific type of data set (e.g., data dictionaries/code books for behavioral and eye-tracking studies, code is documented and made publicly available in Gitlab).

For the studies based on machine learning, datasets and code are released in .readme files that explain in detail the content of code, annotation and data, and how to use them.

4. INCREASE DATA RE-USE *(what data will remain re-usable and for how long, is embargo foreseen; how the data is licensed; data quality assurance procedures)*

Rich metadata and documentation will be provided for all data files. This can include:

- A codebook that will accompany the output of each experiment, which properly describes the content of the data files.
- A file containing the methodology of the experiment stored as .txt, .pdf, or .readme-files

- Annotated scripts/code for running the experiment and analysis if applicable. Data are analyzed using R, Matlab and/or Python, saved in their native formats (respectively, .R, .m and .py). Scripts formatting the text data files outputted by the manufacturer's software (EDF2ASC, dataviewer or Tobii Pro Lab) prior to formal analyses, are provided.

Research output will be made available via RDR or OSF, at the moment of publication. Creative commons attribution licenses will be selected in RDR or OSF.

To ensure qualitative data storage, XXXXXXXXXXXXXXXXXXXX will check the completeness of the documentation.

5. ALLOCATION OF RESOURCES and DATA SECURITY *(estimated costs for making the project data open access and potential value of long-term data preservation; procedures for data backup and recovery; transfer of sensitive data and secure storage in repositories for long term preservation and curation)*

There are no costs attached to apply the FAIR-principles to the data and research outputs because repositories and other resources are provided by KU Leuven free of charge, including GitLab and RDR. Additionally, also the usage of OSF is free of charge. Individual researchers are responsible for data management during the course of their projects. The grant-holder will ensure the long-term storage of the data at KU Leuven.