FWO DMP Template - Flemish Standard Data Management Plan

Version KU Leuven

Project supervisors (from application round 2018 onwards) and fellows (from application round 2020 onwards) will, upon being awarded their project or fellowship, be invited to develop their answers to the data management related questions into a DMP. The FWO expects a **completed DMP no later than 6 months after the official start date** of the project or fellowship. The DMP should not be submitted to FWO but to the research co-ordination office of the host institute; FWO may request the DMP in a random check.

At the end of the project, the **final version of the DMP** has to be added to the final report of the project; this should be submitted to FWO by the supervisor-spokesperson through FWO's e-portal. This DMP may of course have been updated since its first version. The DMP is an element in the final evaluation of the project by the relevant expert panel. Both the DMP submitted within the first 6 months after the start date and the final DMP may use this template.

The DMP template used by the Research Foundation Flanders (FWO) corresponds with the Flemish Standard Data Management Plan. This Flemish Standard DMP was developed by the Flemish Research Data Network (FRDN) Task Force DMP which comprises representatives of all Flemish funders and research institutions. This is a standardized DMP template based on the previous FWO template that contains the core requirements for data management planning. To increase understanding and facilitate completion of the DMP, a standardized **glossary** of definitions and abbreviations is available via the following link.

1. General Project Information	
Name Grant Holder & ORCID	Fatemeh Behrad (0000-0003-2629-0854)
Contributor name(s) (+ ORCID) & roles	Supervisor: Johan Wagemans (0000-0002-7970-1541) Co-supervisor: Tinne Tuytelaars (0000-0003-3307-9723)
Project number [1] & title	3H230183 & The Art of Perception: Empowering Neural Networks to See Beauty Like Humans
Funder(s) GrantID [2]	FWO & 1159925N
Affiliation(s)	☑ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: ROR identifier KU Leuven: 05f950310

[1] "Project number" refers to the institutional project number. This question is optional. Applicants can only provide one project number.

[2] Funder(s) GrantID refers to the number of the DMP at the funder(s), here one can specify multiple GrantIDs if multiple funding sources were used.

Please provide a short project description	This project bridges the gap between computer science and experimental psychology by integrating principles from human perceptual organization into neural networks to enhance their performance in image aesthetic assessment. This effort will advance computational aesthetics while providing psychology with a validated computational model leading to a better understanding of human image aesthetics. I will curate two datasets: one for quantifying the aesthetic scores of images and the other for revealing the reasons behind the aesthetic quality of images, focusing on addressing limitations in existing datasets. Three deep neural networks will be developed. The initial model, tailored for estimating the aesthetic score of images, integrates insights from human visual perception and strives to preserve high-resolution information. The second model will aim for a more descriptive output beyond a single aesthetic score by integrating the first model into an image captioning model. This will help us generate verbal descriptions that capture subtle nuances of aesthetic information. This new model will be adapted for image aesthetic assessment, incorporating visual, aesthetic, and contextual information provided by comments. All models will be analyzed to gain insights leading to a deeper understanding of factors shaping aesthetics.

2. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data [3].

				ONLY FOR DIGITAL DATA	*ONLY FOR DIGITAL DATA*	*ONLY FOR DIGITAL DATA*	*ONLY FOR PHYSICAL DATA*
Dataset Name	Description	New or Reused	Digital or Physical	Digital Data Type	Digital Data Format	Digital Data Volume (MB, GB, TB)	Physical Volume
AVA	Image aesthetic assessment (IAA) dataset	Reuse existing data	Digital	Images & Numerical	JPEG & CSV	32.2 GB	
AADB	IAA dataset	Reuse existing data	Digital	Images & Numerical	JPEG & TXT	2.25 GB	
TAD66k	IAA dataset	Reuse existing data	Digital	Images & Numerical	JPEG & CSV	2.22 GB	
BAID	IAA dataset	Reuse existing data	Digital	Images & Numerical	JPEG & CSV	28.5 GB	
PARA	IAA dataset	Reuse existing data	Digital	Images & Numerical	JPEG & CSV	11.1 GB	
KonIQ10k	Image quality assessment (IQA) dataset	Reuse existing data	Digital	Images & Numerical	JPEG & CSV	5.09 GB	
SPAQ	IQA dataset	Reuse existing data	Digital	Images & Numerical	JPEG & XLSX	33 GB	
WP1 dataset	IAA dataset	Generate new data	Digital	Images & Numerical	JPEG	< 100 GB	
WP3 dataset	IAA dataset	Generate new data	Digital	Images & Textual	JPEG & CSV	< 100 GB	
Baseline AI Models	Pretrained Vision	Reuse existing data	Digital	Model & Software	PTH & PY	< 100 GB	

[3] Add rows for each dataset you want to describe.

	transformers and CLIP model						
AI Models	Image aesthetic assessment models and image aesthetic-related captioning model	Generate new data	Digital	Model & Software	PTH & PY	< 100 GB	

GUIDANCE:
The data description forms the basis of your entire DMP, so make sure it is detailed and complete. It includes digital and physical data and encompasses the whole spectrum ranging from raw data to processed and analysed data including analysis scripts and code. Physical data are all materials that need proper management because they are valuable, difficult to replace and/or ethical issues are associated. Materials that are not considered data in an RDM context include your own manuscripts, theses and presentations; documentation is an integral part of your datasets and should described under documentation/metadata.
RDM Guidance on data

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type.	AVA: https://doi.org/10.1109/CVPR.2012.6247954 AADB: https://doi.org/10.48550/arXiv.1606.01621 TAD66k: https://doi.org/10.24963/ijcai.2022/132 PARA: https://doi.org/10.48550/arXiv.2203.16754 BAID: https://doi.org/10.48550/arXiv.2303.15166 SPAQ: https://doi.org/10.1109/CVPR42600.2020.00373 KonIQ10k: https://doi.org/10.48550/arXiv.1910.06180 Vision transformer: https://doi.org/10.48550/arXiv.2010.11929 CLIP model: https://doi.org/10.48550/arXiv.2103.00020

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, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.	☒ Yes, human subject data; provide SMEC or EC approval number: G-2024-8885 ☐ Yes, animal data; provide ECD reference number: ☐ Yes, dual use; provide approval number: ☐ No Additional information:
Will you process personal data[4]? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).	☐ Yes (provide PRET G-number or EC S-number below) ☒ No Additional information:
Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.	☐ Yes ☒ No If yes, please comment:
Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements, research collaboration agreements)? If so, please explain to what data they relate and what restrictions are in place.	☐ Yes ☒ No If yes, please explain:
Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain to what data they relate and which restrictions will be asserted.	☐ Yes ☒ No If yes, please explain:

[4] See Glossary Flemish Standard Data Management Plan

3. Documentation and Metadata	
Clearly describe what approach will be followed to capture the accompanying information necessary to keep **data understandable and usable**, for yourself and others, now and in the future (e.g. in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded). *RDM guidance on documentation and metadata*.	All datasets reused in our project are publicly available and described in detail in published conference papers. For the datasets we create, we adopt a similar approach by documenting all relevant information, including data creation processes, content, and usage guidelines, in a publication. Additionally, each dataset will include a README file to provide clear and concise instructions for accessing and understanding the data.

Will a metadata standard be used to make it easier to **find and reuse the data**?
 	
If so, please specify which metadata standard will be used. If not, please specify which metadata will be created to make the data easier to find and reuse.
 	
REPOSITORIES COULD ASK TO DELIVER METADATA IN A CERTAIN FORMAT, WITH SPECIFIED ONTOLOGIES AND VOCABULARIES, I.E. STANDARD LISTS WITH UNIQUE IDENTIFIERS.
	☐ Yes
	☒ No
	If yes, please specify (where appropriate per dataset or data type) which metadata standard will be used:
	
	If no, please specify (where appropriate per dataset or data type) which metadata will be created:
	
	For each dataset created in this project, metadata will be documented in a structured CSV file accompanying the dataset.
	• WP1 Dataset: The metadata file will include key details such as:
	  ○ Image sources.
	  ○ Number of ratings per image.
	  ○ Personal characteristics of participants (e.g., gender, age, education level, and art experience).
	  ○ Aesthetic scores for each image.
	• WP3 Dataset: The metadata file will follow a similar structure to WP1 but will include aesthetic-related captions instead of aesthetic scores.

4. Data Storage & Back-up during the Research Project	
Where will the data be stored? *Consult the interactive KU Leuven storage guide to find the most suitable storage solution for your data.*	☐ Shared network drive (J-drive) ☐ Personal network drive (I-drive) ☒ OneDrive (KU Leuven) ☒ SharePoint online ☐ SharePoint on-premis ☐ Large Volume Storage ☐ Digital Vault ☒ Other: GitLab and GitHub

How will the data be backed up? *WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS?*	☐ Standard back-up provided by KU Leuven ICTS for my storage solution ☒ Personal back-ups I make (specify) ☒ Other (specify) Datasets and AI model weights will be stored within our lab's supercomputers. The code will be stored on GitLab provided by KU Leuven.
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 ☐ No If no, please specify: Our supercomputers, OneDrive, SharePoint, GitHub, and GitLab have enough space for the project's needs.
How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons? *CLEARLY DESCRIBE THE MEASURES (IN TERMS OF PHYSICAL SECURITY, NETWORK SECURITY, AND SECURITY OF COMPUTER SYSTEMS AND FILES) THAT WILL BE TAKEN TO ENSURE THAT STORED AND TRANSFERRED DATA ARE SAFE.* *Guidance on security for research data*	All systems in our lab are secured with strong, unique passwords, limiting access to authorized users only. Data stored on platforms like SharePoint, OneDrive, GitHub, and GitLab is further safeguarded through Two-Factor Authentication adding an additional layer of security beyond passwords.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	We have access to free data storage (SharePoint, OneDrive, GitHub, and GitLab by KU Leuven and Our supercomputers).

5. Data Preservation after the end of the Research Project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...). *Guidance on data preservation*	☒ All data will be preserved for 10 years according to KU Leuven RDM policy ☐ All data will be preserved for 25 years according to CTC recommendations for clinical trials with medicinal products for human use and for clinical experiments on humans ☐ Certain data cannot be kept for 10 years (explain)
Where will these data be archived (stored and curated for the long-term)? *Dedicated data repositories are often the best place to preserve your data. Data not suitable for preservation in a repository can be stored using a KU Leuven storage solution, consult the interactive KU Leuven storage guide.*	☒ KU Leuven RDR ☐ Large Volume Storage (longterm for large volumes) ☐ Shared network drive (J-drive) ☒ Other (specifiy): The code and model weights will be stored on GitHub after publishing the results in a journal or conference.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	We do not expect any costs. Our datasets will be uploaded to KU Leuven RDR, providing storage that is preserved for at least 10 years post-project completion for free. Additionally, datasets, model weights, and codes will be stored within our lab's supercomputers for free. GitHub and KU Leuven's GitLab will provide free storage for more than 10 years.

6. Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project? Please explain per dataset or data type which data will be made available. *Note that 'available' does not necessarily mean that the data set becomes openly available, conditions for access and use may apply. Availability in this question thus entails both open & restricted access. For more information: https://wiki.surfnet.nl/display/standards/info-eu-repo/#infoeurepo-AccessRights*	☒ Yes, as open data ☐ Yes, as embargoed data (temporary restriction) ☐ Yes, as restricted data (upon approval, or institutional access only) ☐ No (closed access) ☐ Other, please specify: Because our datasets do not contain any sensitive information that could compromise the anonymity of the annotators, they will be made publicly available online. Also, our code and model weights will be made publicly available for reproducibility of our results.
If access is restricted, please specify who will be able to access the data and under what conditions.	
Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain per dataset or data type where appropriate.	☐ Yes, privacy aspects ☐ Yes, intellectual property rights ☐ Yes, ethical aspects ☐ Yes, aspects of dual use ☐ Yes, other ☒ No If yes, please specify:
Where will the data be made available? If already known, please provide a repository per dataset or data type.	☒ KU Leuven RDR ☒ Other data repository (specify) ☐ Other (specify) We also share our code and models on GitHub to ensure accessibility and enable a broader audience to benefit from them.

When will the data be made available?	☒ Upon publication of research results ☐ Specific date (specify) ☐ Other (specify)
Which data usage licenses are you going to provide? If none, please explain why. *A DATA USAGE LICENSE INDICATES WHETHER THE DATA CAN BE REUSED OR NOT AND UNDER WHAT CONDITIONS. IF NO LICENCE IS GRANTED, THE DATA ARE IN A GREY ZONE AND CANNOT BE LEGALLY REUSED. DO NOTE THAT YOU MAY ONLY RELEASE DATA UNDER A LICENCE CHOSEN BY YOURSELF IF IT DOES NOT ALREADY FALL UNDER ANOTHER LICENCE THAT MIGHT PROHIBIT THAT.* *Check the RDR guidance on licences for data and software sources code or consult the License selector tool to help you choose.*	☒ CC-BY 4.0 (data) ☐ Data Transfer Agreement (restricted data) ☒ MIT licence (code) ☐ GNU GPL-3.0 (code) ☐ Other (specify)
Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here. *INDICATE WHETHER YOU INTEND TO ADD A PERSISTENT AND UNIQUE IDENTIFIER IN ORDER TO IDENTIFY AND RETRIEVE THE DATA.*	☒ Yes, a DOI will be added upon deposit in a data repository ☐ My dataset already has a PID ☐ No
What are the expected costs for data sharing? How will these costs be covered?	We do not expect any costs and GitHub allows us to share the codes and pre-trained models for free. Also, RDR provides enough free storage for our project.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Fatemeh Behrad & Johan Wagemans

Who will manage data storage and backup during the research project?	Fatemeh Behrad
Who will manage data preservation and sharing?	Fatemeh Behrad & Johan Wagemans
Who will update and implement this DMP?	Fatemeh Behrad