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	Erkki Tobias Bartczak, 0000-0002-1414-596X
Contributor name(s) (+ ORCID) & roles	Maarten Vergauwen, 0000-0003-3465-9033, first promoter Maarten Bassier, 0000-0001-8526-8847, daily supervisor
Project number [1] & title	UAV-assisted bridge inspection
Funder(s) GrantID [2]	1SH4O24N
Affiliation(s)	☒ KU Leuven
Please provide a short project description	This project aims to develop a novel routine bridge inspection methodology using unmanned aerial vehicles (UAVs). We propose to automate the data acquisition, damage detection and data interpretation steps using highly innovative machine learning and computer vision concepts.

2. Research Data Summary

[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.

WP1 – UAV data acquisition and photogrammetric processing

					ONLY FOR DIGITAL DATA	*ONLY FOR DIGITAL DATA*	*ONLY FOR DIGITAL DATA*
Dataset Name	**Description**	**New or Reused**	**Origin of Data**	**Digital or Physical**	**Digital Data Type**	**Digital Data Format**	**Digital Data Volume (MB, GB, TB)**
1 Photogrammetric bridge surveys	Creating 3D models of bridges, estimate camera positions and detect damages in images	☒ Generate new data	observational	☒ Digital ☐ Physical	☒ Images ☒ DensePointClouds ☒ 3D Model	.png .e57 .obj	☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☒ < 5 TB

Input images from survey will be processed in Agisoft Metashape or RealityCapture.

WP2 – Data processing for damage detection and characterisation

					ONLY FOR DIGITAL DATA	*ONLY FOR DIGITAL DATA*	*ONLY FOR DIGITAL DATA*
Dataset Name	**Description**	**New or Reused**	**Origin of Data**	**Digital or Physical**	**Digital Data Type**	**Digital Data Format**	**Digital Data Volume (MB, GB, TB)**
2 Damage detection models	Using third party images and existing images from 1 to train deep learning models.	☐ Generate new data ☒ Reuse existing data	Departement Mobiliteit en Openbare Werken (MOW)	☒ Digital ☐ Physical	☒ deep learning models ☒ Images	.pt .png	☒ < 5 TB

Data will be processed using deep learning models such as YOLOv8.

WP3 – Development of damage interpretation and tracking methodology

					ONLY FOR DIGITAL DATA	*ONLY FOR DIGITAL DATA*	*ONLY FOR DIGITAL DATA*
Dataset Name	**Description**	**New or Reused**	**Origin of Data**	**Digital or Physical**	**Digital Data Type**	**Digital Data Format**	**Digital Data Volume (MB, GB, TB)**
Linked Data Models	Connecting the identified damages and the 3D model	☒ Generate new data ☐ Reuse existing data	Output from WP1 and WP2	☒ Digital ☐ Physical	☒ Numerical	.csv	☒ < 1 GB

Input data will be processed using RDFs in Visual Studio Code.

WP4 – Validation and representation of automated bridge inspections

					ONLY FOR DIGITAL DATA	*ONLY FOR DIGITAL DATA*	*ONLY FOR DIGITAL DATA*
Dataset Name	**Description**	**New or Reused**	**Origin of Data**	**Digital or Physical**	**Digital Data Type**	**Digital Data Format**	**Digital Data Volume (MB, GB, TB)**
Valorisation	Analysing the outputs of the project on a greater scale.	☐ Generate new data ☒ Reuse existing data	Project WP3	☒ Digital ☐ Physical	☒ Numerical ☒ Textual	.csv .pdf	☒ < 1 GB

Documentation

					ONLY FOR DIGITAL DATA	*ONLY FOR DIGITAL DATA*	*ONLY FOR DIGITAL DATA*
Dataset Name	**Description**	**New or Reused**	**Origin of Data**	**Digital or Physical**	**Digital Data Type**	**Digital Data Format**	**Digital Data Volume (MB, GB, TB)**
manuscripts, theses and presentations	manuscripts, theses and presentations	☒ Generate new data ☐ Reuse existing data	project	☒ Digital ☐ Physical	☒ Presentations ☒ Numerical ☒ Textual ☒ Videos	.csv .pdf .txt .pptx .mp4	☒ < 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.	Damage detection model input images may come from: Departement Mobiliteit en Openbare Werken (MOW), bart.verbeke@mow.vlaanderen.be
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.	☒ No

Will you process personal data[3]? 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 ☒ No
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: Potential for spin-off
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
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: The deep learning model training dataset may be used within our research though MOW remains the provider and owner of the dataset.

3. Documentation and Metadata

[3] See Glossary Flemish Standard Data Management Plan

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.*	Per work package, all input, output data, and scripts will be collected. Specific folders will be created to contain (1) input data, (2) processing files, and (3) output data. Included in the Input Data folder will be a text file with a clear description of what the data within the folder represent, including the type, format, source of each dataset, and dates acquired. Any scripts or intermediate data used to generate output data will be kept in the Processing folder with a corresponding text file describing applied tools and methodology used to process the data, as well as explanations of the file names. The Output data folder will house all final processing outputs to be used in further work packages and will also have a text file describing each output data as well as how they were generated.
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 For each bridge survey, the following folder structure will be used: 0_ProjectData, 1_RawData, 2_PreProcessed, 3_Processing, 4_Outputs, 5_Analysis, 6_MachineLearning, 7_LocalCode, 8_DataFusion, 9_FinalResults The process of each step will be documented in a word file saved in: 0_ProjectData

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.*	☒ Personal network drive ☒ OneDrive (KU Leuven) ☒ Large Volume Storage

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)
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 The large amount of training image data for 2 will be separately storage on a portable 4 TB hard drive, which holds currently 0.9 TB image data.
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*	The data is password protected.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	The data storage volumes for cloud storage provided by the Department will suffice. Additional offline backups will be done on external hard drives (estimated cost €200 for 5TB).

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) The data provided by MOW will not be retained after the end of the project as we only have the right to use it and not share it. All other data created during this project will be retained for the expected 10 year period.
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):
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	The data will be stored on the university's central servers. The expected cost for preserving this data is €13 per year. This cost will be covered by the working budget of Maarten Vergauwen, the main promotor.

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: Some output data will be publicly available if it is necessary to reproduce the results of the published papers. The Python scripts are generally considered intellectual property and may not necessarily made public. Images that may be used for deep learning model training may or may not be made open access.
If access is restricted, please specify who will be able to access the data and under what conditions.	Any contributor to the project, including third parties who contributed by delivering training images, may get access upon granted request.
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 The outputs of this research project may have direct commercial deliverables. Especially the dataset from WP 1 and 2 may potentially be used for commercial benefits. Therefore, these datasets will only be shared if it is in the best interest for the (continuation) of the project's intent, e.g. to improve Deep Learning models by fusing our and a third-party training data.
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)

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 PID 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?	There are no expected costs related to data sharing.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	The PhD researcher will be responsible for data documentation and metadata.

Who will manage data storage and backup during the research project?	The PhD researcher will be responsible for data storage and backup during the project.
Who will manage data preservation and sharing?	The PhD researcher will be responsible for compiling a folder with all data and corresponding metadata that needs to be preserved. Our division's data storage team will be responsible for storing the data thereafter, with the supervision of the promoters.
Who will update and implement this DMP?	The promoters bear the end responsibility of updating and implementing this DMP.