

FWO DMP Template - Flemish Standard Data Management Plan

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	Bingnan Liu 0000-0003-2542-884X
Contributor name(s) (+ ORCID) & roles	Xavier Casadevall i Solvas
Project number ¹ & title	GPUE/22/035 New machine-vision and automation methods for high throughput combinatorial drug screening
Funder(s) GrantID ²	GPUE/22/035
Affiliation(s)	<input checked="" type="checkbox"/> KU Leuven <input type="checkbox"/> Universiteit Antwerpen <input type="checkbox"/> Universiteit Gent <input type="checkbox"/> Universiteit Hasselt <input type="checkbox"/> Vrije Universiteit Brussel <input type="checkbox"/> Other: Provide ROR ³ identifier when possible:
Please provide a short project description	Lens-free imaging (LFI) is a novel imaging technology that uses holographic reconstruction algorithms to replace optical devices such as lenses. This project aims to explore advanced holographic reconstruction algorithms for improved LFI quality. The methods to be investigated include image super-resolution, 2D and 3D image reconstruction and neural network generalization combining computational imaging. The device and approaches will be applied to on-chip microscopy for high-throughput drug screening.

2. Research Data Summary

¹ "Project number" refers to the institutional project number. This question is optional since not every institution has an internal project number different from the GrantID. 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.

³ Research Organization Registry Community. <https://ror.org/>

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 data⁴.

Dataset Name	Description	New or Reused	Digital or Physical	ONLY FOR DIGITAL DATA	ONLY FOR DIGITAL DATA	ONLY FOR DIGITAL DATA	ONLY FOR PHYSICAL DATA
				Digital Data Type	Digital Data Format	Digital Data Volume (MB, GB, TB)	Physical Volume
Microscopic Image Sequence	From the camera software provided by Basler (Basler - mDisplay (baslerweb.com))	<input checked="" type="checkbox"/> Reuse existing data	<input checked="" type="checkbox"/> Digital	<input checked="" type="checkbox"/> Experimental	<input checked="" type="checkbox"/> other: .bmp	<input checked="" type="checkbox"/> < 1 TB	
Microscopic Image	From collaborative Lab - Laboratory for Molecular Virology and Gene Therapy (kuleuven.be) , via Zeiss ZEN Microscopy Software	<input checked="" type="checkbox"/> Reuse existing data	<input checked="" type="checkbox"/> Digital	<input checked="" type="checkbox"/> Experimental	<input checked="" type="checkbox"/> other: .tif	<input checked="" type="checkbox"/> < 1 GB	
Opensource Codes and	Reused from Github.com	<input checked="" type="checkbox"/> Reuse existing data	<input checked="" type="checkbox"/> Digital	<input checked="" type="checkbox"/> Software	<input checked="" type="checkbox"/> other: .py .c .cpp .cu .h .j	<input checked="" type="checkbox"/> NA	

⁴ Add rows for each dataset you want to describe.

Corresponding Documents	(One program project contains multiple formats)				son .va .hap .txt .md		
New Codes and Corresponding Documents	(One program project contains multiple formats of codes)	<input checked="" type="checkbox"/> Generate new data	<input checked="" type="checkbox"/> Digital	<input checked="" type="checkbox"/> Software	<input checked="" type="checkbox"/> other: .py .c .cpp .cu .h .json .txt .md	<input checked="" type="checkbox"/> NA	
Standard Computer Vision Public Datasets	Open-source Public Datasets (Image)	<input checked="" type="checkbox"/> Reuse existing data	<input checked="" type="checkbox"/> Digital	<input checked="" type="checkbox"/> Software	<input checked="" type="checkbox"/> other: .jpg	<input checked="" type="checkbox"/> < 10 TB	

GUIDANCE:

DATA CAN BE DIGITAL OR PHYSICAL (FOR EXAMPLE BIOBANK, BIOLOGICAL SAMPLES, ...). DATA TYPE: DATA ARE OFTEN GROUPED BY TYPE (OBSERVATIONAL, EXPERIMENTAL ETC.), FORMAT AND/OR COLLECTION/GENERATION METHOD.

EXAMPLES OF DATA TYPES: OBSERVATIONAL (E.G. SURVEY RESULTS, SENSOR READINGS, SENSORY OBSERVATIONS); EXPERIMENTAL (E.G. MICROSCOPY, SPECTROSCOPY, CHROMATOGRAMS, GENE SEQUENCES); COMPILED/AGGREGATED DATA⁵ (E.G. TEXT & DATA MINING, DERIVED VARIABLES, 3D MODELLING); SIMULATION DATA (E.G. CLIMATE MODELS); SOFTWARE, ETC.

EXAMPLES OF DATA FORMATS: TABULAR DATA (.POR, .SPSS, STRUCTURED TEXT OR MARK-UP FILE XML, .TAB, .CSV), TEXTUAL DATA (.RTF, .XML, .TXT), GEOSPATIAL DATA (.DWG, .GML, ..), IMAGE DATA, AUDIO DATA, VIDEO DATA, DOCUMENTATION & COMPUTATIONAL SCRIPT.

DIGITAL DATA VOLUME: PLEASE ESTIMATE THE UPPER LIMIT OF THE VOLUME OF THE DATA PER DATASET OR DATA TYPE.

PHYSICAL VOLUME: PLEASE ESTIMATE THE PHYSICAL VOLUME OF THE RESEARCH MATERIALS (FOR EXAMPLE THE NUMBER OF RELEVANT BIOLOGICAL SAMPLES THAT NEED TO BE STORED AND PRESERVED DURING THE PROJECT AND/OR AFTER).

⁵ These data are generated by combining multiple existing datasets.

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.

Public datasets:

1. CIFAR-10
 - 60k labelled images with 10 classes
 - Use case: image classification
 - Link to Dataset -> <https://www.cs.toronto.edu/~kriz/cifar.html>
2. CIFAR-100
 - 600k labelled images with 100 classes (each class has 500 train and 100 test images)
 - Use case: image classification
 - Link to Dataset -> <https://www.cs.toronto.edu/~kriz/cifar.html>
3. CALTECH-101
 - 300x200 pixel images with 101 classes (each class has 40-800 images)
 - Use case: image classification
 - <https://data.caltech.edu/>
4. CALTECH-256
 - 30k labelled images with 256 classes
 - Use case: image classification
 - <https://data.caltech.edu/>
5. MNIST
 - 70k labelled images with 10 classes (handwritten digits 0-9)
 - Use case: image classification
 - Link to Dataset -> <http://yann.lecun.com/exdb/mnist/>
6. ImageNet
 - 15 million labelled images
 - Use case: image classification , object detection
 - Link to Dataset -> <http://image-net.org/index>
7. Pascal VOC
 - 11.5k labelled images with 20 classes
 - Use case: object detection, image segmentation
 - Link to Dataset -> <http://host.robots.ox.ac.uk/pascal/VOC/>
8. MS COCO

	<ul style="list-style-type: none"> ○ 200k labelled images with 80 classes ○ Use case: object detection, image segmentation, image captioning ○ Link to Dataset -> https://cocodataset.org/#home <p>9. Sports-1M</p> <ul style="list-style-type: none"> ○ 1M labelled videos (average length - 5.5 mins) with 487 classes ○ Use case: video classification, action recognition ○ Link to Dataset -> https://cs.stanford.edu/people/karpathy/deepvideo/index.html <p>10. YouTube-8M</p> <ul style="list-style-type: none"> ○ 8M labelled videos (average length - 4 mins) with 4800 classes ○ Use case: video classification, action recognition ○ Link to Dataset -> https://research.google.com/youtube8m/
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, please describe these issues further and refer to specific datasets or data types when appropriate.	<input type="checkbox"/> Yes, human subject data <input type="checkbox"/> Yes, animal data <input type="checkbox"/> Yes, dual use <input checked="" type="checkbox"/> No If yes, please describe:
Will you process personal data ⁶ ? If so, briefly describe the kind of personal data you will use. Please refer to specific datasets or data types when appropriate. If available, add the reference to your file in your host institution's privacy register.	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No If yes: <ul style="list-style-type: none"> - Short description of the kind of personal data that will be used: - Privacy Registry Reference:

⁶ See Glossary Flemish Standard Data Management Plan

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.	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> 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.	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> 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.	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No If yes, please explain:

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).	<p>The codes are following the code management tool Git. The versions are recorded.</p> <p>With clear documentation README.md</p> <p>For the Images, are managed by the contributors – colleagues from Biomimetics Lab and Laboratory for Molecular Virology and Gene Therapy. When I reuse them, I do not change the original management formats. More information: Protocols (containing info about both materials(setting, parameters, set-up, ...)and methods), the research progress and obtained data, what they represent and how they were generated, will be collected in an electronic notebook (eLABJournal, Bio-ITech).</p>
---	---

<p>Will a metadata standard be used to make it easier to find and reuse the data?</p> <p>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.</p> <p><i>REPOSITORIES COULD ASK TO DELIVER METADATA IN A CERTAIN FORMAT, WITH SPECIFIED ONTOLOGIES AND VOCABULARIES, I.E. STANDARD LISTS WITH UNIQUE IDENTIFIERS.</i></p>	<p><input checked="" type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>If yes, please specify (where appropriate per dataset or data type) which metadata standard will be used: Git Git (git-scm.com) for the codes.</p> <p>If no, please specify (where appropriate per dataset or data type) which metadata will be created:</p>
---	--

4. Data Storage & Back-up during the Research Project	
<p>Where will the data be stored?</p>	<p>Online storage:</p> <ul style="list-style-type: none"> - Microsoft OneDrive and Sharepoint with KU Leuven personel account. - GitHub (*Only for open-source) - KUL ICTS and SET-IT managed cloud drives <p>Local storage: Disks</p>

<p>How will the data be backed up?</p> <p><i>WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS? DESCRIBE THE LOCATIONS, STORAGE MEDIA AND PROCEDURES THAT WILL BE USED FOR STORING AND BACKING UP DIGITAL AND NON-DIGITAL DATA DURING RESEARCH.⁷</i></p> <p><i>REFER TO INSTITUTION-SPECIFIC POLICIES REGARDING BACKUP PROCEDURES WHEN APPROPRIATE.</i></p>	<p>My original data are double backed up on cloud: KUL cloud dirver (J, K drive) and OneDrive. Using ICTS permitted Onedrive client on Linux, all data are downloaded locally on hard disk.</p>
<p>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.</p>	<p><input checked="" type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>If yes, please specify concisely: 2TB for personal onedrive and 5TB for Teams sharepoint. 5TB for KUL SET-IT managed cloud service. If no, please specify:</p>
<p>How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?</p> <p><i>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.⁷</i></p>	<p>All online services (except open-source) are provided by the University (affiliated Microsoft account).</p>

⁷ Source: Ghent University Generic DMP Evaluation Rubric: <https://osf.io/2z5g3/>

What are the expected costs for data storage and backup during the research project? How will these costs be covered?	KU Leuven provides multiple options for (long term) data storage. Type 1 server backend storage with mirror backup for the FWO-SB project folder will cost € 270 per TB per year. The estimated maximal cost for the 4-year project would therefore be € 1080. Large datasets that do not require frequent access can be stored on a separate server for large volume storage, costing € 113,84 per TB per year. The estimated maximal cost for the 4-year project would therefore be € 455,36 if this type of data storage is required. All costs will be covered by the project budget.
---	---

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...).	All data obtained during this project will be retained for the expected 5 year period.
Where will these data be archived (stored and curated for the long-term)?	The digital data will be stored on the university's central servers (with automatic backup procedures) for at least 5 years, conform the KU Leuven RDM policy.

<p>What are the expected costs for data preservation during the expected retention period? How will these costs be covered?</p>	<p>Cost of the large volume storage will be € 128,39 per TB and year. We anticipate that we will need 1 TB for 5 years to keep the essential data available. This will amount to € 614,95 and will be covered by the project's budget.</p>
---	--

6. Data Sharing and Reuse

<p>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.</p> <p><i>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/#INFOEU-REPO-ACCESSRIGHTS</i></p>	<p><input type="checkbox"/> Yes, in an Open Access repository</p> <p><input checked="" type="checkbox"/> Yes, in a restricted access repository (after approval, institutional access only, ...)</p> <p><input type="checkbox"/> No (closed access)</p> <p><input type="checkbox"/> Other, please specify:</p>
<p>If access is restricted, please specify who will be able to access the data and under what conditions.</p>	<p>Before the publication, all data can only be accessed by relevant colleagues of KU Leuven.</p>
<p>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.</p>	<p><input type="checkbox"/> Yes, privacy aspects</p> <p><input type="checkbox"/> Yes, intellectual property rights</p> <p><input type="checkbox"/> Yes, ethical aspects</p> <p><input type="checkbox"/> Yes, aspects of dual use</p> <p><input type="checkbox"/> Yes, other</p> <p><input checked="" type="checkbox"/> No</p> <p>If yes, please specify:</p>
<p>Where will the data be made available? If already known, please provide a repository per dataset or data type.</p>	<p>For the work published, in principal codes will be available at Github.</p>

<p>When will the data be made available?</p> <p><i>THIS COULD BE A SPECIFIC DATE (DD/MM/YYYY) OR AN INDICATION SUCH AS 'UPON PUBLICATION OF RESEARCH RESULTS'.</i></p>	<p>After publications.</p>
<p>Which data usage licenses are you going to provide? If none, please explain why.</p> <p><i>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.</i></p> <p><i>EXAMPLE ANSWER: E.G. "DATA FROM THE PROJECT THAT CAN BE SHARED WILL BE MADE AVAILABLE UNDER A CREATIVE COMMONS ATTRIBUTION LICENSE (CC-BY 4.0), SO THAT USERS HAVE TO GIVE CREDIT TO THE ORIGINAL DATA CREATORS." ⁸</i></p>	<p>The codes are with MIT license.</p>
<p>Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here.</p> <p><i>INDICATE WHETHER YOU INTEND TO ADD A PERSISTENT AND UNIQUE IDENTIFIER IN ORDER TO IDENTIFY AND RETRIEVE THE DATA.</i></p>	<p><input type="checkbox"/> Yes <input checked="" type="checkbox"/> No If yes:</p>
<p>What are the expected costs for data sharing? How will these costs be covered?</p>	<p>0</p>

⁸ Source: Ghent University Generic DMP Evaluation Rubric: <https://osf.io/2z5g3/>

7. Responsibilities

Who will manage data documentation and metadata during the research project?	Me and KU Leuven IT services.
Who will manage data storage and backup during the research project?	Me
Who will manage data preservation and sharing?	Me
Who will update and implement this DMP?	Me