# 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 | **Quinten Goovaerts (0000-0002-2168-1103)** |
| Contributor name(s) (+ ORCID) & roles | **Kalyan Das (0000-0002-8897-324X) – Supervisor**<br>**Smita Patel (0000-0002-2523-4933) – Professor of collaboration group (for biochemical validation)**<br>**Jiayu Shen (0000-0001-8447-9378) – PhD student of collaboration group (for biochemical validation)** |
| Project number[1] & title | **3M200511**<br>**High-resolution single-particle cryo-EM structures to reveal the human mitochondrial DNA transcription machinery at work: a step towards understanding mitochondrial related disorders.** |
| Funder(s) GrantID[2] | **1162823N** |
| Affiliation(s) | **X  KU Leuven**<br>☐ Universiteit Antwerpen<br>☐ Universiteit Gent<br>☐ Universiteit Hasselt<br>☐ Vrije Universiteit Brussel<br>☐ Other:<br>Provide ROR[3] identifier when possible: **https://ror.org/03w5j8p12** |

---

[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/

| Please provide a short project description | Gene transcription is carried out by RNA polymerases, which are crucial biological machines that synthesize RNA molecules from protein-coding and non-protein coding genes. This enzymatic process is highly regulated and defects in genes or gene regulation are linked to a myriad of diseases. In eukaryotes, these reactions occur in the nucleus and mitochondria and they have been widely studied in the last decades. Despite significant advances in understanding of gene transcription and regulation fundamentals, our understanding of mitochondrial RNA polymerases remains scarce. In addition, a complete overview of the successive molecular events involving the human mitochondrial RNA polymerases is still missing, however, is required for exploring effective therapies for patients with certain mitochondrial disorders. The proposed fundamental research addresses this limitation by providing a structural and mechanistic framework of the human mitochondrial DNA transcription machinery at work and provides new avenues to better understand the details of normal and pathological biology of the human mitochondrial RNA polymerase at near atomic resolution. This goal will be achieved by trapping consecutive high-resolution structures of the human mitochondrial DNA transcription initiation complex at work using single-particle cryo-EM and biochemical assays to reveal its step-by-step progression in the transcription initiation pathway. |
|---|---|

## 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[4].

| | | | | 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 |
| Sequencing results | Sequencing data to check if plasmids have the correct insert and/or mutations were successful | ☒ Generate new data | ☒ Digital | ☒ Experimental | ☒ .pdf<br>☒ .txt<br>☒ .ab1 | ☒ < 100 MB | |
| Protein Purification chromatography results (Size exclusion & Affinity Chromatography) | Elution profiles of various purifications | ☒ Generate new data | ☒ Digital | ☒ Experimental | ☒ .csv | ☒ < 100 MB | |
| Biophysical characterization results | Proteins will be characterized using biophysical techniques such as DLS, MALS, TSA, | ☒ Generate new data | ☒ Digital | ☒ Experimental | ☒ .dexp<br>☒ .afe7<br>☒ .eds<br>☒ .txt<br>☒ .scn | ☒ < 100 MB | |

---

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

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | NanoDrop and SDS-PAGE | | | | | | |
| Cryo-EM datasets | After complex formation, samples for cryo-EM data collection will be prepared. Multiple intermediate states will be collected | ☒ Generate new data | ☒ Digital | ☒ Experimental | ☒ .tiff<br>☒ .mrc<br>☒ .xml<br>☒ .jpg<br>☒ .emi<br>☒ .emd<br>☒ .sxml<br>☒ .bmp | ☒ < 5 TB (per dataset) | |
| Physical samples | Cell pellets, Purified proteins, ordered DNA sequences and purified initiation complexes will be stored in -80 °C | ☒ Generate new data | ☒ Physical | ☒ Experimental | ☒ NA | ☒ NA | |
| Images and movies | Images* generated by ChimeraX, Chimera, PyMol. Movies* generated by ChimeraX.<br><br>*made with (un)published protein structures | ☒ Generate new data<br>☒ Reuse existing data | ☒ Digital | ☒ Experimental<br>☒ Compiled | ☒ .cxs<br>☒ .py<br>☒ .pyc<br>☒ .tiff<br>☒ .jpg<br>☒ . mp4 | ☒ < 100 GB | |
| Documentation | Experimental protocols (adjusted from literature), manuscripts, analysis, and | ☒ Generate new data<br>☒ Reuse existing data | ☒ Digital<br>☒ Physical | ☒ Experimental<br>☒ Observational | ☒ .docx<br>☒ .xlsx<br>☒ .pdf<br>☒ .ppt | ☒ < 100 GB | Also on paper in laboratory notebooks |

| | | | | | | |
|---|---|---|---|---|---|---|
| | presentations | | | | | |
| Protein density maps & Protein structures | Cryo-EM datasets will be used for structure solution of target protein complexes. Software includes: Relion, CryoSPARC, COOT and Phenix. | ☒ Generate new data | ☒ Digital | ☒ Compiled/ Aggregated data | ☒ .eff<br>☒ .cif<br>☒ .geo<br>☒ .pdb<br>☒ .txt<br>☒ .cs<br>☒ .mrc<br>☒ .star | ☒ < 100 GB | |

| | |
|---|---|
| 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. | Published protein structures are used for comparison and analysis purpose. All protein structures and electron density maps are freely available in the Protein Data Bank and the Electron Microscope Data Bank respectively; and are made available for sharing with the scientific community. |

---

[5] These data are generated by combining multiple existing datasets.

| | |
|---|---|
| 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. | ☐ Yes, human subject data<br>☐ Yes, animal data<br>☐ Yes, dual use<br>☒ No<br>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. | ☐ Yes<br>☒ No<br>If yes:<br><br>   -   Short description of the kind of personal data that will be used:<br>   -   Privacy Registry Reference: |
| Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)?<br>If so, please comment per dataset or data type where appropriate. | ☐ Yes<br>☒ No<br>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)?<br>If so, please explain to what data they relate and what restrictions are in place. | ☐ Yes<br>☒ No<br>If yes, please explain: |

---

[6] See Glossary Flemish Standard Data Management Plan

| 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<br>☒ No<br>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). | **Reused data:** All reused data/protocols are published online and hence is understandable and always accessible and usable<br><br>**Experimental data:** Experimental data will be divided in physical and digital data.<br><br>Physical samples: A list of all physical samples is maintained by me and updated when certain samples are used in further experiments. In addition, all samples are stored on 4°C, -20°C or -80°C and shared in an online inventory tool with the lab (Quartzy). All lab members have access to this tool and can digitally search for the content of the freezers and fridges. Self-explaining names are used to store physical samples *e.g.* name of protein/complex, buffer components, date of production, …<br><br>Digital data: Experimental data obtained from sequencing, purifications and biophysical characterization are saved on the shared drive\* in appropriate folders. These folders are categorized based on the used technique or experiment and a subdivision is made based on user or project. Data folders containing raw and processed data are hierarchically organized and labelled based on the date of data generation, the number of the experiment and the source of the data (DDMMYYYY - self-explaining names).<br><br>Images and movies created by software are also saved on the shared drive in appropriate folders using the system above.<br><br>Cryo-EM datasets will be created by default with metadata imprinted by the image acquisition software automatically. This includes information on project (user), date and time, equipment parameters etc. Data generated with the cryo-EM will be saved on an in-line server and will be backed-up on an external 600 TB capacity server available in our lab. As a large server capacity is available all kind of raw and (meta)data generated through the microscope will be stored for several years. Prof. Kalyan Das will have the end responsibility for ensuring data preservation.<br><br>**Compiled/aggregated data:** Protein density maps, protein coordinates are saved on the shared drive in appropriate folders using the system described above. In addition, density maps are also saved on the user's CryoSPARC account and shared drive on the HPC computer of KU Leuven.<br><br>**Documentation:** Experimental protocols and results will be recorded in physical lab books and stored into Word or Excel files. A summary and the analysis of the results is done in word or PowerPoint files. Data documentation, storage and back-up during the project is my responsibility.<br> \*Data will be stored on servers centrally managed by ICTS KU Leuven, also providing back-up |

| | capacity allowing disaster recovery and guaranteeing long-term data availability. Sharing possibilities are in place via OneDrive.<br><br>Overall, all kind of data is stored in the appropriate folder on the shared drive as following "DDMMYYYY-self-explaining-name". |
|---|---|
| Will a metadata standard be used to make it easier to **find and reuse the data**?<br><br>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.<br><br>*REPOSITORIES COULD ASK TO DELIVER METADATA IN A CERTAIN FORMAT, WITH SPECIFIED ONTOLOGIES AND VOCABULARIES, I.E. STANDARD LISTS WITH UNIQUE IDENTIFIERS.* | ☒ Yes<br>☐ No<br><br>Cryo-EM data-sets will stored on the 600 TB server and labelled as following: DDMMYYYY_name-project_details-parameters.<br>Hard drives are labelled and also stored in the lab. |

| 4.  Data Storage & Back-up during the Research Project |
|---|

| Where will the data be stored? | **Digital data** (including protocols, raw data, analysed data, produced images/movies, processed data, density maps, protein structures…) will be stored on a shared network (J: Drive), which only can be accessed by members of our lab. In addition, data is synched with OneDrive providing a copy of the original data if required. Data stored on these servers are centrally managed by ICTS KU Leuven, also providing back-up capacity allowing disaster recovery and guaranteeing long-term data availability.<br>**Cryo-EM data** will be saved on an in-line server and will be backed-up on an external 600 TB capacity server available in our lab.<br>**Hard copy notebooks** with raw data will be stored physically in our laboratory. |
|---|---|

| | |
|---|---|
| How will the data be backed up?<br><br>*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]*<br><br>*Refer to institution-specific policies regarding backup procedures when appropriate.* | Cryo-EM data is backed-up on an external 600TB capacity RAID server available in our lab.<br><br>Other data (all data, except cryo-EM data) is automatically back-up when stored in the KU Leuven datacenters. Files are actively backed-up at different timestamps using "snapshot" technology on faculty network shares (J: Drive), where the previous versions of the changed files are kept online in a snapshot on the same storage system.<br><br>To prevent data loss or in case of "business continuity" or "disaster recovery", a mirror (exact copy) of all data is created in a second datacenter. Every hour the data is copied to the second datacenter allowing data restoration within the hour by the ICTS team in case the primary storage unit is corrupted. |
| 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<br>☐ No<br><br>If yes, please specify concisely:<br>All KU Leuven staff has 2 TB of storage on OneDrive, which is sufficient for all data (except for cryo-EM datasets). For cryo-EM datasets an external 600 TB server is available in our lab. Once results are published, the related data is compressed to reduce their size. In addition an extra 15 TB is available on the HPC of the KU Leuven and can be increased if necessary. |
| How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?<br><br>*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]* | All computers in the lab require to login with your personal KU Leuven password and account. If the person is not a member of the lab, the shared drive is invisible for that person. Data stored on the HPC is protected and inaccessible by others not related to the lab.<br>In addition, everywhere where a login is required a two-step authentication is required. |

---

[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? | The external 600 TB data server is already available in our lab hence no costs have to be made for cryo-EM data preservation. Data-storage on the HPC is 20 euro/TB/year. Currently we have 15 TB available which corresponds to 300 euro/year and is managed by funding by the host institute and other projects. Hard drives are also available for data storage and back up. |

| 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 (including cryo-EM metadata) will be retained for at least five years after the end of the project. |
| Where will these data be archived (stored and curated for the long-term)? | Data (except for Cryo-EM data) will be stored on the K: archive drive (managed by KU Leuven) which has an unlimited maximum size, meaning that all generated data can be easily stored there. In addition, the K: archive drive can only be accessed by the involved researchers. The Cryo-EM raw and (meta)data will be compressed and stored on our external 600 Tb server in the lab |
| What are the expected costs for data preservation during the expected retention period? How will these costs be covered? | The yearly cost for network share storage on the K: archive drive is 11.38 euro per 100 Gb of which Group biomedical sciences sponsors 50%, so the price paid by the lab will be 5.69 per 100 Gb. No costs are expected for storage on our external 600 Tb server. |

| 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.<br><br>*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, in an Open Access repository<br>☒ Yes, in a restricted access repository (after approval, institutional access only, …)<br>☐ No (closed access)<br>☐ Other, please specify: |
| If access is restricted, please specify who will be able to access the data and under what conditions. | Professor Kalyan Das will be the lead contact for any request to access the data from this project. |
| 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<br>☐ Yes, intellectual property rights<br>☐ Yes, ethical aspects<br>☐ Yes, aspects of dual use<br>☐ Yes, other<br>☒ No<br><br>If yes, please specify: |
| Where will the data be made available? If already known, please provide a repository per dataset or data type. | Cryo-EM structures and cryo-EM density maps will be made available in the Protein Data Bank (PDB) and the Electron Microscopy Data Bank (EMDB) respectively. Raw data will be made available upon request. |

| | |
|---|---|
| When will the data be made available?<br><br>*THIS COULD BE A SPECIFIC DATE (DD/MM/YYYY) OR AN INDICATION SUCH AS 'UPON PUBLICATION OF RESEARCH RESULTS'.* | Data will be made available upon publication of research results |
| Which data usage licenses are you going to provide? If none, please explain why.<br><br>*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.*<br><br>*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] | 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. |
| Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here.<br><br>*INDICATE WHETHER YOU INTEND TO ADD A PERSISTENT AND UNIQUE IDENTIFIER IN ORDER TO IDENTIFY AND RETRIEVE THE DATA.* | ☒ Yes<br>☐ No<br>If yes: Permanent identifiers will be added to the data upon deposition (in PDB and EMDB) |
| What are the expected costs for data sharing? How will these costs be covered? | Data will be shared in open access data banks such as PDB and EMDB. |

---

[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? | Data documentation, storage & back-up during this project will be the responsibility of the PhD researcher (Quinten Goovaerts) under supervision of Prof. Kalyan Das. |
| Who will manage data storage and backup during the research project? | Data documentation, storage & back-up during this project will be the responsibility of the PhD researcher (Quinten Goovaerts) under supervision of Prof. Kalyan Das. |
| Who will manage data preservation and sharing? | Prof. Kalyan Das will have the end responsibility for ensuring data preservation and will be the lead contact for any request to access the data from this project. |
| Who will update and implement this DMP? | The PhD researcher (Quinten Goovaerts) will be responsible to update and implement this DMP. Prof. Kalyan Das will bear the end responsibility for updating and implementing this DMP. |