

DMP - FWO - Fonds Wetenschappelijk Onderzoek - GOG3721N

The participants to the present project understand the value of FAIR Data Management and Open Access to Scientific Publications and Research Data. They are fully committed to abiding by the related FWO policies.

The present data management plan (DMP) describes the specific outputs of this project and how they will be made available to the community. All participants will be informed of updates in the DMP, and any new participant will receive training to ensure compliance with the consortium's data management conventions.

1. General Information

Name applicant	Alan Urban - alan.urban@nerf.be
FWO Project Number & Title	G0G3721N Understanding brain circuit dysfunction in amblyopia using large-scale multimodal recordings in a new visuomotor task applied to animal models and patients
Affiliation	NERF

Responsible: Daniel Hillier

2. Data description

Will you generate/collect new data and/or make use of existing data?	New data
Describe the origin, type and format of the data (per dataset) and its (estimated) volume	<p>Observational data</p> <p>Experimental data Digital images Video and audio files Electrophysiology data Electrophysiology data</p> <p>Simulation data Derived and compiled data</p> <p>Canonical data These datasets represent an important source of information for the laboratory of the PI (including future staff), for scientists, journalists and higher education teachers working in the field of neuroscience, but also for non-profit organizations and industries active in the field of science and technology.</p>

3. Ethical and legal issues

Will you use personal data? If so, shortly describe the kind of personal data you will use AND add the reference to your file in your host institution's privacy register.	No
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, add the reference to the formal approval by the relevant ethical review committee(s).	No
Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?	<p>Yes</p> <p>We do not exclude that the proposed work could result in research data with potential for tech transfer and valorization. Ownership of the data generated belongs to KU Leuven and VIB in accordance with the framework agreement of both institutes. VIB has a policy to actively monitor research data for such potential. If there is substantial potential, the invention will be thoroughly assessed, and in a number of cases the invention will be IP protected (mostly patent protection or copyright protection). As such the IP protection does not withhold the research data from being made public. In the case a decision is taken to file a patent application it will be planned so that publications need not be delayed.</p> <p>The use of our designs, codes and all biological materials will be subjected to the terms described in their respective MTAs.</p> <p>Specific examples (adjust as required):</p> <ul style="list-style-type: none"> - new algorithm for brain imaging - software analysis pipeline - new hardware solutions for ultrasound imaging in large animal

Do existing 3 rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?	No
---	----

4. Documentation and metadata

What documentation will be provided to enable understanding and reuse of the data collected/generated in this project?

Data will be generated following standardized protocols. Metadata will be documented by the research and technical staff at the time of data collection and analysis, by taking careful notes in the electronic laboratory notebook (E-notebook) and/or in hard copy lab notebooks that refer to specific datasets.

All datasets will be accompanied by a README.txt file containing all the associated metadata (see more details below).

The data will be generated following standardized protocols. Clear and detailed descriptions of these protocols will be stored in our lab protocol database, and published along with the results.

For computations ran in central NERF computing facilities (nerfcluster-fs), a json file is automatically generated for certain type of jobs as spikesorting, which contains metadata. This contains information as input file, location of output file, computation-date and parameters used along the calculation. Furthermore, in case of conda environments, an yml file is automatically generated which contains python packages and their versions. These files are saved in a directory choosen by an user, and automatically copied to a directory managed by the admin of the system. These metadata files augment the manner to reproduce results and a posrteriori understanding of data production.

<p>Will a metadata standard be used? If so, describe in detail which standard will be used. If not, state in detail which metadata will be created to make the data easy/easier to find and reuse.</p>	<p>No</p> <p>For computations ran in central NERF computing facilities (nerfcluster-fs), a json file is automatically generated for certain type of jobs as spikesorting, which contains metadata. This contains information as input file, location of output file, computation-date and parameters used along the calculation. Furthermore, in case of conda environments, an yml file is automatically generated which contains python packages and their versions. These files are saved in a directory choosen by an user, and automatically copied to a directory managed by the admin of the system. These metadata files augment the manner to reproduce results and a porteriori understanding of data production.</p> <p>These metadata files are placed in a folder which is indexed by the job-scheduler (SLURM), which is accompanied by the execution date. Therefore, the manner to search for metadata is slurmid-DD.MM.YYYY. This meta data is reachable by user at http://nerfcluster-fs:8080/singlejob , which is browsable from the nerf-network.</p> <p>Metadata will include the following elements:</p> <ul style="list-style-type: none"> • Title: free text • Creator: Last name, first name, organization • Date and time reference • Subject: Choice of keywords and classifications • Description: Text explaining the content of the data set and other contextual information needed for the correct interpretation of the data, the software(s) (including version number) used to produce and to read the data, the purpose of the experiment, etc. • Format: Details of the file format, • Resource Type: data set, image, audio, etc. • Identifier: DOI (when applicable) • Access rights: closed access, embargoed access, restricted access, open access. <p>Additionally, we will closely monitor MIBBI (Minimum Information for Biological and Biomedical Investigations) for metadata standards more specific to our data type.</p> <p>- We will use the NIFTI file format and its associated metadata for all imaging data including for imaging data, all information regarding the subject (animal ID), experimental conditions (setupID, physiological parameters when available such as</p>
--	---

	<p>temperature, ECG, EMG, ambient noise level, ambient light level), recording conditions (frame rate, processing pipeline).</p> <p>or specific datasets, additional metadata will be associated with the data file as appropriate. Give details as needed for the project.</p> <p>The final dataset will be accompanied by this information under the form of a README.txt document. This file will be located in the top level directory of the dataset and will also list the contents of the other files and outline the file-naming convention used. This will allow the data to be understood by other members of the laboratory and add contextual value to the dataset for future reuse.</p> <p>The metadata files automatically generated for computations executed at the nerfcluster-fs, are placed in a folder which is indexed by the job-scheduler (SLURM), which is accompanied by the execution date. Therefore, the manner to search for metadata is slurmlD-DD.MM.YYYY. This meta data is reachable by the user at http://nerfcluster-fs:8080/singlejob , which is browsable from the nerf-network.</p>
--	--

5. Data storage & backup during the FWO project

Where will the data be stored?	<p>The NERF storage system is composed of two subsystems, one is a dedicated archive system and the another one is for Work In Progress (WIP). The former is an object storage system, for which NERF has an active maintenance contract with the provider (Cloudian), while the latter is a filesystem based on openZFS.</p> <p>The archive system serves to store data that needs to be kept long term (years) due to legal requirements or for a later analysis. This system is so-called "nerfhf01".</p> <p>The WIP system offers a high throughput which suits best for highly demanding daily IO operations. This system is so-called "nerffs13"</p> <p>- Algorithms, scripts and softwares: All the relevant algorithms, scripts and software code driving the project will be stored in a private online git repository from the GitHub account of the team.</p>
How will the data be backed up?	<p>The storage system at NERF has multiple layers of protection to ensure long-term data retention.</p> <p>The WIP server (nerffs13) has a "twin" server located in a different data center which acts as a mirror of the former. This provides data backup in case of full failure of the nerffs13, whether caused for severe hardware issues or in case the entire data center is compromised. Furthermore, snapshots of the data are taken regularly that allow recovering from accidental corruption or deletion of data, which in combination with a RAIDZ2 (zfs-raid) configuration provides a strong data redundancy per server. Lastly, the system screens the data on a regular basis to avoid data corruption due to bit-rot.</p> <p>The archive system (nerfhf01) is a redundant system on itself, this is composed of several nodes distributed in multiple data centers. The nerfhf01 technology allows to have one entire node down and data is not compromised.</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>Yes</p> <p>The archive and WIP systems have roughly more than 1000TB of capacity each. Those, together with backup servers, can be expanded upon necessity and considering technical specs. This task is managed by the admin of the system, who also performs the upgrades and provides data storage monitoring and reporting</p>

<p>What are the expected costs for data storage and backup during the project? How will these costs be covered?</p>	<p>Based on the last two years expenses and data storage forecast, NERF costs for the storage system comprises the hardware itself, and license and maintenance costs. The former amounts to 45000€ per year and the latter to 5000€ per year. These costs are covered by the central NERF budget.</p>
<p>Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?</p>	<p>NERF servers are in imec campus at Leuven. Thus, we have strong network protection as provided by imec firewalls. Moreover, imec provides a dedicated VLAN for NERF, meaning that only registered devices can access to the NERF network from the imec campus.</p> <p>For users outside of the imec campus, a Cisco AnyConnect VPN can be used to access to the NERF network. The VPN login authorization is setup by two factors authentication for each user. This VPN is provided and maintained by imec.</p> <p>In addition to that network security, the access to our storage servers from user computers is via SMB protocol. Therefore, each research group at NERF has their own “SMB accounts” as setup in the storage server by the system admin.</p> <p>Consequently, whether a device in-imec-campust or out-imec-campus attempts to access to the NERF network and thereafter to NERF storage servers, security layers on the network side and server accounts have to be passed first. This strongly reduces the likelihood that unauthorized persons access to NERF data.</p>

6. Data preservation after the end of the FWO project

FWO expects that data generated during the project are retained for a period of minimally 5 years after the end of the project, in as far as legal and contractual agreements allow.

Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).	The minimum preservation term of 5 years after the end of the project will be applied to all datasets. All datasets will be stored on the university's central servers with automatic back-up procedures for at least 5 years, conform the KU Leuven RDM policy. The costs (€156 per TB per year for “Large volume-storage”) will be covered by the AU Lab. If applicable: Datasets collected in the context of clinical research, which fall under the scope of the Belgian Law of 7 May 2004, will be archived for 25 years, in agreement with UZ Leuven policy and the European Regulation 536/2014 on clinical trials of medicinal products for human use.
Where will these data be archived (= stored for the long term)?	Data that needs long term storage is stored in nerfhf01.
What are the expected costs for data preservation during these 5 years? How will the costs be covered?	That amounts roughly to 60000€ per year, and will be covered by the NERF central budget.

7. Data sharing and reuse

Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3 rd party, legal restrictions)?	No
Which data will be made available after the end of the project?	<p>Participants to the present project are committed to publish research results to communicate them to peers and to a wide audience. All research outputs supporting publications will be made openly accessible. Depending on their nature, some data may be made available prior to publication, either on an individual basis to interested researchers and/or potential new collaborators, or publicly via repositories (e.g. negative data).</p> <p>We aim at communicating our results in top journals that require full disclosure upon publication of all included data, either in the main text, in supplementary material or in a data repository if requested by the journal and following deposit advice given by the journal. Depending on the journal, accessibility restrictions may apply.</p>
Where/how will the data be made available for reuse?	Upon request by mail
When will the data be made available?	After an embargo period.
Who will be able to access the data and under what conditions?	<p>Whenever possible, datasets and the appropriate metadata will be made publicly available through repositories that support FAIR data sharing. As detailed above, metadata will contain sufficient information to support data interpretation and reuse, and will be conform to community norms. These repositories clearly describe their conditions of use (typically under a Creative Commons CC0 1.0 Universal (CC0 1.0) Public Domain Dedication, a Creative Commons Attribution (CC-BY) or an ODC Public Domain Dedication and Licence, with a material transfer agreement when applicable). Interested parties will thereby be allowed to access data directly, and they will give credit to the authors for the data used by citing the corresponding DOI. For data shared directly by the PI, a material transfer agreement (and a non-disclosure agreement if applicable) will be concluded with the beneficiaries in order to clearly describe the types of reuse that are permitted.</p>

<p>What are the expected costs for data sharing? How will these costs be covered?</p>	<p>It is the intention to minimize data management costs by implementing standard procedures e.g. for metadata collection and file storage and organization from the start of the project, and by using free-to-use data repositories and dissemination facilities whenever possible. Data management costs will be covered by the laboratory budget. A budget for publication costs has been requested in this project.</p>
---	--

8. Responsibilities

Who will be responsible for the data documentation & metadata?	Metadata will be documented by the research and technical staff at the time of data collection and analysis, by taking careful notes in dedicated files (txt, csv) managed by the URBAN Lab. In addition, as indicated in the section 11, jobs launched in the nerfcluster will automatically create metadata, where the responsible is the system administrator of NERF.
Who will be responsible for data storage & back up during the project?	As long as the data is in the NERF central storage system, the responsible is the system administrator of NERF, Guiliano MAGGI.
Who will be responsible for ensuring data preservation and sharing?	The PI is responsible for data preservation and sharing, with support from the research and technical staff involved in the project, from René Custers and Alexander Botzki for the electronic laboratory notebook (ELN) and from Raf De Coster for the KU Leuven drives.
Who bears the end responsibility for updating & implementing this DMP?	The PI is ultimately responsible for all data management during and after data collection, including implementing and updating the DMP.