Contract no. 101137248

LWNVIVAT

Limiting West Nile Virus impact by novel vaccines and therapeutics approaches

ACTION: HORIZON-RIA

CALL: HORIZON-HLTH-2023-DISEASE-03

TOPIC: Pandemic preparedness and response: Immunogenicity of viral proteins of viruses with epidemic and pandemic potential

Deliverable D9.13: Data Management plan v1

Due Date of Deliverable	May 31st, 2024	**Completion Date of Deliverable**	May 31st, 2024
Deliverable leading partner	IrsiCaixa	**Author**	Jorge Carrillo, Cristina Gálvez and Anaí
WP Nº	9	**WP Title**	Communication, dissemination and exploitation

Project starting date	01/12/2024	**Project Duration**	48 months

Dissemination Level	**PU**	**Public**	**x**
	PP	**Restricted to other programme participants (including the Commission Services)**	
	RE	**Restricted to a group specified by the consortium (including Commission Services)**	
	CO	**Confidential, only for members of the consortium (including Commission Services)**	

This project has received funding from the European Union's Horizon Europe Research and Innovation Programme under Grant Agreement Nº 101137248

TABLE OF CONTENT

1 EXECUTIVE SUMMARY

This report is the first deliverable of Task 9.13 “Data Management” and describes the initial Data Management Plan (DMP) for the LWNVIVAT project, funded by the EU’s Horizon Europe Programme under Grant Agreement number 101137248. The purpose of the DMP is to provide an overview of all datasets collected and generated by the project, and to define the LWNVIVAT consortium’s data management policy regarding these datasets.

The LWNVIVAT DMP follows the structure of the Horizon Europe DMP template. It reflects the status of the data that is collected, processed, or generated, as well as the methodologies and standards followed. It specifies whether and how data will be shared and/or made open, and how it will be preserved.

This initial version of the DMP outlines the general policy and approach to data management in LWNVIVAT. This includes for example topics like data and metadata collection, publication and deposition of open data.

The DMP will evolve over the lifespan of the project. Future versions will refine and enhance policy aspects and provide more detail regarding the datasets produced and where they will be shared by the LWNVIVAT project.

2 DATA SUMMARY

2.1 What is the purpose of the data generation or re-use and its relation to the objectives of the project?

The LWNVIVAT project aims to develop a safe and effective vaccine against the West Nile virus (WNV), as well as therapeutic antibodies, addressing the urgent clinical need for protection, particularly among vulnerable populations like the elderly and immunocompromised individuals. The objective is to create tools that prevents WNV infection and neuroinvasion, while avoiding cross-reactivity with other flaviviruses. Using advanced computational tools, we will design immunogens targeting the WNV Envelope protein to elicit potent and broad neutralizing antibodies, that will be screened to selected those with the highest therapeutic potential. Additionally, innovative techniques like Virus Like Particles will be explored to enhance immune responses (both T and B) and improve vaccine efficacy. Testing will involve *in vitro* studies using human tonsil organoids and *in vivo* studies in mice and swine.

2.2 What types and formats of data will the project generate or re-use?

To achieve these objectives, the LWNVIVAT project will collect and generate various types of data:

- **Computational modelling:** computational modelling will be used to design the immunogens to target WNV. Data will be stored as .pdb (protein structure), .pkl (estimates), .xls/.ods (row-based epitope description) and .csv (residue sequences and analyzed) formats.

- **Protein characterization**: different assays to characterize proteins will be used to understand the interaction affinity, kinetics, oligomerization and stability of WNV proteins. These assays include chromatography, biacore and nanoDSF for which raw data will be stored as .astra6, .blr and .pan formats. For all the assays the analyzed data will be stored as .csv.

- **Flow cytometry**: Flow cytometry data will be generated to analyze VLP produced from insect and mammalian cells; to describe immunophenotype and assess the production of cytokines from human tonsils challenged with vaccine candidates; and to evaluate the cellular phenotype of vaccine-induced CD4+ and CD8+ T cells. Raw data files in flow cytometry instrument will be saved as (.fcs) format. The intermedium analyses, considering gating strategy, will be processed in FlowJo software and stored as .wsp. Finally, the definitive output with the corresponding cell populations and phenotype will be saved as .wsp and .xslx.

- **ELISA**: ELISA assays will be used to characterize antibody:antigen/VLP interaction, antigen specific humoral responses and to quantify cytokine production in cell culture supernatants as well as in plasma/serum samples form immunized and WNV challenged mice. Raw and analyzed data will be stored as .csv or .xls.

- **Electronic microscopy**: VLP particle formation and characterization will be performed by electron microscopy. Raw data will be stored as .emi, .ser, .dm3, .dm4, .bcf, .mrc, .hdf5, .msa, .rpl, .raw or .tiff and processed data and analyzed data as .tif, .png, or .jpeg.

- **Western blot**: The analysis and characterization of WNV immunogen production, antigenicity and incorporation into VLPs will be performed by western blot. Raw data will be stored as .tiff and processed data as .tiff, jpeg or .png.

- **ELISPOT**: ELISPOT assays will be used to evaluate the magnitude of T-cell responses to WNV antigens in both human tonsil samples, and in immunized and WNV challenged mice. Data are retrieved from the ELISpot counter in .png files and data are transferred to Excel files for integration in overall analyses (.xlsx,) and jpg for plate image capturing.

- **Neutralization**: viral neutralization from generated WNV vaccine will be assessed by neutralization assays. Both raw and analyzed data will be stored as .csv.

- **Mice/swine data**: clinical parameters from animal studies (e.g. weight, behavior, and health status.) will be collected in .xls format.

- **Histological data**: histological assays will be used to analyze target tissue damage, including blood brain barrier integrity, and local virus replication in human and mice samples. Data will be collected and analyzed in .ndpi, .tif, .jpg, or .png format.

- **rt-qPCR**: detection of cytokines in mouse brain to compare the expression of general cytokines, chemokines, and biomarkers of neurovascular damage and neuroinflammation will be performed by RT-qPCR. Data will be stored as .ixo or .csv.

- **Surveys**: different surveys will be conducted related to vaccine literacy of primary and secondary students and the impact of the education activities provided by teachers. The raw data will be collected in .csv and .xls, and the analyzed data will be saved in .xls format.

- **Focus groups**: this dataset will contain data related with policy recommendations and a research agenda to improve the control and prevention strategies against emerging infectious diseases of European countries. The data will be collected in .mp4 and .pdf formats.

- **Compressed data sets**: large data sets will be stored as compressed files using .zip or .rar format.

2.3 Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded

Several tasks within the project will utilize data already available in public repositories, namely:

- Uniprot and NCBI: it will be used to perform the analysis of WNV sequence diversity and to analyze the cross-reactivity between WNV epitopes against the proteome of others flavivirus.
- Genebank: This resource will be accessed to obtain and validate WNV antigen sequences and variations.
- Protein Data Bank (PDB): 3D structure of WNV proteins will be obtained from PDB.

2.4 What is the origin/provenance of the data, either generated or re-used?

The data will primarily originate from experimental assays, computational modeling, and surveys conducted as part of the project. Additionally, some data will be sourced from public repositories, such as Uniprot, Protein Data Bank or Genebank.

2.5 What is the expected size of the data that you intend to generate or re-use?

The exact size of the data is currently unknown. However, based on initial experiences with storing results from various measurements, we anticipate that the main contributors to data size will be images from electron microscopy, computational modeling, and flow cytometry. As such, we expect the total data size to range between 500GB and 1TB.

2.6 To whom might your data be useful ('data utility'), outside your project?

The data generated by the project will be valuable to various stakeholders outside of the project, including:

- Clinicians: preclinical validation of a WNV vaccine provides clinicians a robust basis for future studies. This enables the medical community to respond more effectively to future outbreaks, potentially saving lives and enhancing public health.
- Immunologists and Infectious Disease Investigators: Researchers studying WNV and vaccines will benefit significantly from the data. It will provide valuable insights into the efficacy and mechanisms of action of potential vaccines, aiding in the development of novel strategies for combating WNV infections.
- Policy makers: The results of the present proposal might be of interest to policy makers at local, regional, national or international levels to support decisions on health and research priorities.
- Industry: Vaccine and diagnostic kits developers will be also interested in the results of the current project and on the final assets that will emerge from the research activity performed in LWNVIVAT.

3 FAIR DATA

3.1 Making data findable, including provisions for metadata

LWNVIVAT is committed to maximizing the openness of data while respecting intellectual property rights, personal data protection, and privacy considerations. Data affecting personal confidentiality and privacy will not be publicly shared.

For all other cases, data and metadata utilized in publications will be openly accessible through reputable disciplinary repositories such as Immport, Protein Data Bank (PDB), or Electron Microscopy Data Bank (EMDB). Alternatively, multidisciplinary repositories like CORA.RDR (Catalan Open Research Area. Repositori de dades de Recerca) or Zenodo will be considered. Details about which repository is used will be further elaborated in subsequent versions of this data management plan. In all cases, datasets and its associated metadata will be identified with a PID (Persistent Identifier for Data, e.g. DOI).

Additionally, open code will be made available on GitHub.

Rich metadata will accompany the data. However, due to variability among different data types, it may not be possible to find a single metadata standard that fits all datasets. Therefore, a pragmatic approach will be adopted to establish a common and minimal metadata schema. This schema will apply to datasets published in data repositories, with potential extensions for data-type-specific standards like those of the Protein Data Bank. The common metadata schema will be based on the flexible and widely adopted Dublin Core Schema, also endorsed by the European repository OpenAIRE. It will minimally include:

- Title
- Descriptions
- Files
- Upload type
- Publication date
- Creators
- License
- PID
- Keywords
- Related identifiers
- Grants

When allowed by the repository, a Readme.txt file will be provided with additional information to better describe the datasets. Additionally, keywords will be selected using controlled vocabularies, whenever feasible, tailored to the specific data type. This approach aims to enhance the discoverability and potential reusability of the datasets.

3.2 Making data accessible

Throughout the project duration, data will only be available to researchers participating in the project. Data transfers between project partners will be done through a secured cloud platform (e.g. OneDrive), accessible only to specific users.

Open-accessible results, devoid of intellectual property rights (IPR) conflicts or sensitive information, will be promptly shared to maximize the impact of LWNVIVAT. Selected data and findings will be disseminated to the scientific community and stakeholders through publication in peer-reviewed journals, conference presentations, and trusted data repositories. These repositories will assign persistent identifiers to ensure long-term accessibility.

For restricted access data (e.g. WP4), access will be controlled by a Data Access Committee, composed by researchers involved in data collection and analysis. This committee will be responsible for approving access to datasets in accordance with a Data Access Agreement.

Access to the data will be facilitated through various software and methods, including text readers, Alphafold, PyMOL, MeshLab, BioEdit, Biacore T100 evaluation software, Astra software, PR PANTA analysis, FlowJo, ImageJ, LightCycler 480 software, and VLC Media Player. These tools are either open-source or widely accepted within the scientific community, ensuring accessibility for those seeking to utilize the data.

3.3 Making data interoperable

LWNVIVAT is committed to fostering interoperability among datasets, enabling seamless data exchange and re-use across researchers, institutions, organizations, and countries. To achieve this, we will adhere to established standards for data formats, prioritizing compatibility with available (open) software applications. This approach will facilitate the integration of our datasets with those from diverse sources.

Data produced by LWNVIVAT will be stored in non-proprietary formats whenever possible, including .docx, .xlsx, .csv, .pdb, .txt, and PDF. In cases where non-proprietary formats are not feasible, community-accepted formats like .fastq for molecular data or .fcs for flow cytometry will be utilized.

The following formats are recommended:

- .txt/.csv/.xml/ (for data)
- .odf/.odc/.pdf/.ppt (for reports/presentations)
- .tiff (for images)
- .jpg/.png/.gif/.svg/ (for graphics)

Additionally, each published dataset will be cross-referenced with associated publications and related datasets. Controlled vocabularies such as Medical Subject Headings (MeSH) and Library of Congress Subject Headings (LoC) will be applied wherever possible to standardize terminology and facilitate data discovery and integration.

3.4 Increase data re-use

LWNVIVAT is dedicated to facilitating the re-use of data beyond the confines of the project. To this end, LWNVIVAT partners are considering various open licenses, including those within the Creative Commons family, to enable specific licensing of datasets while ensuring their re-usability.

Open access to data generated across all work packages will be provided upon publication, guided by the principle of "as open as possible, as close as necessary."

All data produced within LWNVIVAT holds the potential for re-use by other projects once the primary goals of LWNVIVAT have been accomplished. Generated data adhere to standard experimental and technical protocols, allowing for re-analysis if new analytical approaches emerge. Furthermore, the intention is for the data to remain re-usable indefinitely.

Quality control procedures employed for data generated within LWNVIVAT adhere to industry standards and represent the latest advancements in the field. These procedures will be thoroughly documented in open publications. Since raw data will also be shared, researchers will have the flexibility to apply alternative quality control protocols as needed.

4 OTHER RESEARCH OUTPUTS

Other research outputs generated within the project will include vaccine prototypes, protein expression vectors, recombinant proteins, and antibodies. These outputs will only be shared once the corresponding patents have been filed and will be governed by a Material Transfer Agreement, with some restrictions that are yet to be defined.

The management of these outputs will be addressed in future Data Management Plans as needed.

5 ALLOCATION OF RESOURCES

IrsiCaixa will lead coordination of data management. However, it is the responsibility of each work package leaders to oversee the coordination, storage and preservation of the data produced in accordance with this management plan

- IrsiCaixa: Jorge Carrillo
- BSC-CNS: Alfonso Valencia
- UM: Mireia Pelegrin
- UCPH: Henrik Kloverpris
- CRG: Carlo Carolis
- TUBS: Michael Hust

Regarding the publication of results, two main routes to open access are suggested and recommended in Horizon Europe projects by the European Commission and in the context of LWNVIVAT:

- Self-archiving/'green' open access: Authors or their representatives deposit the published article or the final peer-reviewed manuscript in an online repository either before, at the same time as, or after publication. Some publishers may require an embargo period before granting open access.

- Open access publishing/'gold' open access: Articles are immediately published in open access mode. In this model, publication costs are often covered by authors through Article Processing Charges (APCs), shifting the financial burden away from readers.

In the context of research funding, open access requirements do not mandate the publication of results. The decision to publish remains at the discretion of grant beneficiaries. Open access considerations only arise if publication is chosen as a dissemination method.

6 DATA SECURITY

Data security is the responsibility of each partner, with the following general recommendations provided:

- Back up files/data regularly and ensure that at each time at least two copies of the data are stored at two different sites.
- Enable computer firewalls and maintain up-to-date and operational anti-malware software.
- Users must have access to the computers and/or servers via individual user accounts and not shared accounts.
- Collaborative networks/platforms/Intranets: Permission-controlled files so that users, depending on their status, can “read only”, “write”, or “execute” files. Computers connected networks should not store sensitive data, unless data is encrypted, to minimize network vulnerabilities.
- Cloud-based storage is useful as a secondary or tertiary storage location for your files. Research Data repositories are recommended in this sense. However, this via is not recommended for sensitive or confidential data that will be encrypted or treated anonymously.

In LWNVIVAT, minimal sensitive data will be collected. Personal data (e.g., from WP4) will be stored in a pseudonymized manner on the institutional servers of the respective work package leaders. Long-term storage is ensured on internal partner servers for at least 10 years, supplemented by storage in external repositories.

7 ETHICS

All human data and/or biological material will adhere to the general Data Protection Regulation (GDPR) (Regulation (EU) 2016/679), and different safeguards will be applied. Selected personal data not belonging to special categories, will be collected but limited (respecting “data minimization principle”) to only those relevant and necessary for the project. Partners involved in human studies will exclusively handle pseudonymized data.

Approval from Institutional Review Boards (IRBs) will be obtained to conduct the studies outlined in this project. Participants will provide informed consent, wherein they will be informed about the use of their data in the project and provided with a contact email should they wish to withdraw from the study. Participants will be made aware that data collected prior to withdrawal may still be used in research projects, but no new data will be collected.

Each participating institution conducting research activities involving personal data holds responsibility for ensuring the protection of personal data.

All external services will act as data processors and the controller of the dataset will be responsible to make sure to revise the material and data transfer agreements to be compliant with the appropriate national and international legislation.

8 OTHER ISSUES

The guidelines from einaDMP have been used in the creation of this DMP (https://dmp.csuc.cat/). The data management department of IrsiCaixa (the coordinating institution in LWNVIVAT) will also participate in the revision of the document and in future updates.