



VETBIONET

Veterinary Biocontained facility Network for excellence in animal infectiology research and experimentation

Deliverable D5.2

Creation of the Data Management Plan (DMP)

Due date of deliverable: M6

Actual submission date: M8

Start date of the project: March 1st, 2017

Duration: 60 months

Organisation name of lead contractor: EAAP

Revision: V1

Dissemination level	
Public	x
Confidential, only for members of the consortium (including Commission Services)	
Classified, as referred to in Commission Decision 2001/844/EC	

Table of contents

Summary	3
VetBioNet Data Management Plan	4
1. Data Summary	5
2. FAIR Data	7
3. Allocation of resources	10
4. Data security	11
5. Ethical aspects	11
6. Other issues	12

Summary

Objectives:

The H2020 topic under which VetBioNet has been funded (INFRAIA-01-2016-2017) participates per default in the Open Research Data Pilot which aims to improve and maximize access to and re-use of research data generated by H2020 projects.

The Deliverable 5.2 “Creation of the Data Management Plan” is aimed at raising awareness of project partners on Open Research Data and making sure that the data supporting VetBioNet publications are made available in open access repositories.

Rationale:

Deliverable 5.2 describes the data VetBioNet will generate, whether and how they will be exploited or made accessible for verification and re-use, and how they will be curated and preserved.

D5.2 also describes the roles and responsibilities of project partners as far as publication, management and preservation of open data is concerned. The deliverable is written in a plain language and will be updated in time with the periodic reviews of the project conducted by the European Commission.

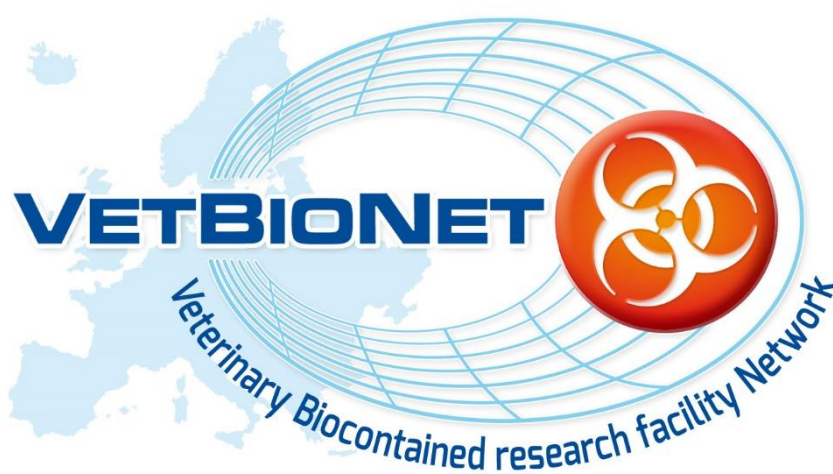
Teams involved:

EAAP, INRA, INIA, EMC, MRI, FLI, IT

VETBIONET

DATA MANAGEMENT PLAN

VERSION N. 1



1. Data Summary

1.1 Purpose of data collection/generation and its relation to the objectives of the project

The H2020 topic under which VetBioNet has been funded (INFRAIA-01-2016-2017) participates per default in the open access to research data pilot which aims to improve and maximize access to and re-use of research data generated by H2020 projects. This data management plan is part of this pilot and describes the data VetBioNet will generate, whether and how they will be exploited or made accessible for verification and re-use, and how they will be curated and preserved.

VetBioNet will collect and generate data to advance research on epizootic and zoonotic diseases, with the final objective of strengthening the present European capacity and competence to meet the challenges of (re)emerging animal infectious diseases. VetBioNet will collect data generated from its Joint Research Activities (JRA), which are designed to improve the scientific and technological standards of the integrated services provided by the network infrastructures. For this, the [FAIR principles](#) (Findability, Accessibility, Interoperability, and Reusability) and [ARRIVE guidelines](#) for scientific reporting will be adopted. Moreover, VetBioNet will also collect data from Transnational Access (TNA) research projects on a voluntary basis, namely when the owners of the data at stake agree to make them available and with a possible delay of 2 years after the end of their TNA projects.

1.2 Types and formats of data generated/collected

VetBioNet JRA will generate data of several types, including phenotypic, genotypic and sequencing data from pathogens and hosts. JRA and Networking Activities (NA) will also deliver novel guidance documents, training tools, harmonized protocols and other scientific information of interest for the research community. The final research data produced by JRA will be curated in a harmonized way to fit the purpose and stored in publicly available standard data repositories (e.g. Gene Expression Omnibus, Sequence Read Archive, etc.) or in the dedicated VetBioNet database, facilitating sharing of information among project partners and finally the wider community of scientists and end-users through specifically designed interfaces.

Research data (phenotypic, genotypic, sequencing and other kinds of data) generated by researchers outside the VetBioNet consortium through TNA Activities will be covered by data

management on a case-by-case basis, as these data may be required to be protected by confidentiality terms specific to each researcher selected for conducting a TNA project within VetBioNet. To stimulate the public sharing of these data, specific agreements will be sought with researchers conducting TNA projects - allowing them to openly share only a part of their research data and with a possible delay of 2 years after the end of their TNA projects.

Further details on the types and formats of data to be generated or collected by VetBioNet are included in the table here below:

Type	Source	Volume	Format
Spreadsheet	JRA and TNA Activities (Laboratory experiments / simulation / observation / compilation)	Small files, up to few megabytes per experiment	XLSX CSV
Image	JRA Activities (Laboratory experiments)	Few terabytes per experiment	RAW TIFF
Sequencing data	JRA Activities (Laboratory experiments)	200GB per sequence	FASTQ FASTA SAM/BAM GFF/GTF BED VCF

1.3 Existing data re-used, origin of data, expected size and who will find data useful

VetBioNet will use existing genomic data and related publications from NADIR (FP7 project, GA n. 228394). Particularly, data generated by NADIR experiments on fishes, chickens and sheep could be re-used. These data will not be available for use outside the VetBioNet network unless the owners give permission.

Overall, the data collected, generated and made openly available by the project will originate from VetBioNet JRA and NA, and possibly from VetBioNet TNA Activities and NADIR final results. It is expected that VetBioNet will not publicly share some experimental data to be obtained through WP7 (*Experimental models for animal infectious diseases*) and WP8 (*Development of novel analytical tools and reagents to help interrogate the host pathogen*)

interaction) activities, because these data could be included at a later stage in scientific publications or used for pursuing patents.

It is expected that, due to the nature of the studies, mainly small size data sets will be collected, at the exception of imaging files, sequencing data and data from instrumented behavioral and physiological monitoring. The largest amount of raw data is expected to come from imaging and NGS (next generation sequencing) genomic data.

The data that VetBioNet will make openly available will be useful for the veterinary field, particularly for the scientific community interested in animal infectious diseases, characterizing models and livestock production. The data will also be useful to policy makers, funders and industry who have an interest in biosecurity for animal diseases and zoonoses through sustainable and safe production of livestock species for the public good.

2. FAIR Data

2.1. Making foreground data findable, including provisions for metadata

VetBioNet final research data will be stored in publicly available standard data repositories or in the project dedicated database, which will be accessible through VetBioNet website (<http://www.vetbionet.eu/>). The data stored in VetBioNet database will be searchable with metadata and a standard identification mechanism will be used to describe the data.

The naming convention followed will be:

[work package].[task].[(JRA/TNA id)].[text description].[version].[format]

Search keywords will be provided to optimize possibilities for re-use of data and raw data will be curated accordingly. The keywords to be chosen will relate to animal species and pathogens analyzed by the project and followed methodologies.

Clear version numbers will be provided.

Metadata will be produced to describe how the data were analyzed and summarize important portions of data.

A simple form will be designed by WP7, and will be filled by partners involved in the development of experimental models for animal infectious diseases. Researchers will use the taxonomic international accepted names for pathogens, the common pathogen titration

methodologies expressing doses such as TCDI50, pfu, etc. The form will include the following fields: Animal species, Pathogen, Doses, Route of inoculation, Biosecurity requirements (following OIE classification), Mortality percentage, Clinical signs, Pathogenesis, Immunity and Reference to published data.

This form could show in summary the main project achievements, and could facilitate the search for information about animal models related to a particular pathogen, animal species, etc.

2.2. Making foreground data openly accessible

It is the goal of VetBioNet to share final research data (i.e., factual data on which summary statistics are based). Data sharing with third parties will be subject to a data-sharing agreement established by the IPUDC (Intellectual Property Use and Dissemination Committee). The agreement will indicate the conditions of use, criteria for access, and acknowledgements. Project participants who wish to withhold patentable or proprietary data can do so, and advice on this point will be given by the IPUDC.

Final research data will be made openly available only after one of the 3 following criteria are met:

- relevant scientific publications based on the data at stake have been accepted;
- a patent application has been published;
- 2 years after the project end.

Biosecurity reasons will be considered before making data openly available, as well as possible IP protection measures that will allow further exploitation of produced data. As explained in previous paragraphs, VetBioNet will also collect and publicly share data from TNA research projects on a voluntary basis, with the consent of the owners of the data.

Finally, VetBioNet partners are committed to give public access to the raw data that will not be subject to a patent application, at the latest two years after the end of the project. VetBioNet partners will autonomously store the raw data used to generate scientific papers in a repository of their choice.

A part of the data made openly available by VetBioNet will be accessible by means of MS Office applications. To access the statistical data publicly shared by VetBioNet, Graphpad Prism or similar software will be needed.

Documentation about the software needed to access the data will be provided upon request.

All freeware programs needed to access the project open data will be provided in the data repositories used by VetBioNet.

Where possible, open data and associated metadata, documentation and codes will be deposited in certified repositories which support open access.

Access to open data stored in the VetBioNet database and other (e.g. institutional) repositories will be provided to registered users (name, affiliation and email will be requested) who have accepted the terms of use.

Data that could raise societal concerns (such as biosecurity risks) will not be shared in open access mode.

The VetBioNet Executive Committee (i.e. the decision-implementing body of the project) will act as data access committee, and will be supported by two experts in social sciences and ethical evaluation.

2.3. Making data interoperable

Formats of data will be decided by the leaders of each Joint Research Activity Leader following, when possible, the guidelines already developed in the NADIR project and in accordance with the work of Networking Activity 3 (Best practices for biosafety, biosecurity and quality management in farmed animal high containment facilities). To ensure interoperability of data across the project and where applicable (e.g., data on gene sequences), project participants will upload basic datasets in standardized forms in a primary database as required by the journal in which they publish their results. Where adequate, data generated in VetBioNet will be defined according to the Animal Trait Ontology for Livestock (ATOL: <http://www.atol-ontology.com/index.php/en/>). ATOL is an on-going initiative and project participants will contribute to the development of the ontology when relevant. Most partners in the project are ISO 9001 certified.

Partners will agree on common terminology, key words, units and formats and apply those to the data and other documents to be incorporated in the consortium's database. Standard vocabularies for all data types present in VetBioNet data sets will be employed, to allow interdisciplinary interoperability.

2.4. Increase data re-use (through clarifying licences)

FAIR principles will be applied, also for optimal formatting of data.

Final research data will be made openly available through publicly available standard data repositories or the dedicated VetBioNet database only after one of the 3 following criteria are met:

- relevant scientific publications based on the data at stake have been accepted;
- a patent application has been published;
- 2 years after the project end.

The data produced and/or used in the project will be made openly available to third parties after the end of the project. Possible restrictions could be put in place due to the need to finalize patenting processes or publish scientific works based on project data.

Final research data stored in the project dedicated database will remain re-usable until the end of the project and for an additional 10-years term if adequate financial sponsors will be found. Raw and final research data stored in other open repositories will remain re-usable for a minimum of 10 years.

Each animal experiment conducted by VetBioNet will require approval from an ethical committee. In order to ensure data quality, guidelines will be produced by INRA and shared among all project participants to establish common procedures for acquiring, storing and amending data. Finally, [ARRIVE guidelines](#) will be adopted for ensuring metadata quality.

3. Allocation of resources

All data produced will initially be recorded in numbered workbooks, signed and dated and retained at each respective site, or stored on network attached file systems that will be regularly archived and automatically backed up. Data will be checked for quality and accuracy and all protocols recorded and adapted to clear guidelines. Where necessary, data will be archived in each institution's archive system. Data will be made fully available to the consortium in a timely manner after passing each partner's quality control criteria and after Intellectual Property considerations have been taken into account. To enable long-term accessibility and validation, data will be stored in formats that are open, non-proprietary, and in common use by the research community.

Each partner of VetBioNet will be responsible for uploading its own final research data on publicly available standard data repositories or the dedicated VetBioNet database. Management of final research data uploaded on the VetBioNet database is the responsibility of EAAP. Once final research data are collected and stored in the VetBioNet database (which will have a maximum capacity of 5 terabytes), there will be a monthly preservation cost of about 750 Euro.

Therefore, final research data will be preserved in one of the following databases: 1. publicly available standard data repositories (e.g. Gene Expression Omnibus, Sequence Read Archive, etc.); 2. VetBioNet database until the end of the project and for an additional 10-years term if adequate financial sponsors will be found.

Open raw data will be stored and preserved by the project partners who generated/collected them. Raw data will be retained and remain accessible for at least 10 years after completion of the project. Data storage facilities will be maintained in accordance with the manufacturer's warranty and guidelines and data backed up at regular intervals and stored safely and securely.

Moreover, project participants are encouraged to publish research data as supporting materials together with their publications, to facilitate preservation of data for future re-use by other projects or research initiatives.

4. Data security

If not stored in publicly available standard data repositories, VetBioNet final research data will be stored in the project dedicated database (primary site) and in a secondary site, which will be geographically distant from the primary one. Data backups will be then run from the secondary site, without any impact on the primary project database.

5. Ethical aspects

Ethical issues that can have an impact on data sharing will be analyzed in WP4 (*Ethical aspects, 3Rs and social impact*).

The project will not launch questionnaires dealing with personal data.

6. Other issues

Currently, VetBioNet partners do not make use of other national/funder/sectorial/departmental procedures for data management.