D4.2 Data management plan

777363 – DRIVE

Development of Robust and Innovative Vaccine Effectiveness

WP4 – Framework for analysis and study reports

Lead contributor	Tom De Smedt (3 – P95)
	tom.desmedt@p-95.com
Other contributors	Margarita Riera (3 – P95), Topi Turunen (1 – FISABIO), Mónica Vázquez (1 – FISABIO), Roberto Bonaiuti (4 – UNIFI)

Due date	31 DEC 2017
Delivery date	22 DEC 2017
Deliverable type	R
Dissemination level	PU

Description of Work	**Version**	**Date**
	V1.0	22 DEC 2017

Document History

Version	Date	Description
V0.1	10 NOV 2017	First Draft
V0.2	27 NOV 2017	Comments from core group incorporated
V0.3	06 DEC 2017	Comments from all task partners incorporated
V0.4	11 DEC 2017	Draft
V1.0	22 DEC 2017	Final Version

Table of contents

1 Introduction and aim

The main objective of the DRIVE project is the development of a governance model to allow the collaboration between the different stakeholders, through public-private partnership, in order to enable the development of a sustainable network of vaccine effectiveness studies.

The DRIVE Description of Action (DoA) includes a Data Management Plan (DMP) as deliverable 4.2, as part of WP4, and together with the DRIVE Consortium Agreement provides the general framework regarding data management, data protection, data ownership, accessibility and sustainability requirements.

Overall, the DMP provides a description of the data management that will be applied in the DRIVE project including:

- A description of the data repositories, who is able to access the data, and who owns the data.
- The main DMP elements for each of the studies contributing (or sharing data) to DRIVE.
- The time period for which data must be stored.
- The standards for data collection, validation evaluation.
- The possibilities of and conditions for sharing data.
- The implementation of data protection requirements.

As the DMP is an evolving document, some of the aspects may be further described and/or updated in later versions of the document.

The DMP will be updated over the course of the project whenever significant changes arise, such as (but not limited to):

- Addition of new data
- Changes in consortium policies (e.g. new innovation potential, …)
- Changes in consortium composition and external factors (e.g. consortium members and/or associated partners joining or leaving).

In summary, the DRIVE DMP gives guidance and provides an oversight of general data management, while each study needs to provide specific data management information including, but not limited to, data capture systems, data analysis systems, data protection and data privacy measures, including description of de-identification of data sets and access rules. In cases where the research results are not open access a justification needs to be provided.

2 General principles

This is the initial DMP for DRIVE. The DMP is a working document, that will evolve during the DRIVE project, and will be updated to reflect project progress. Table 1 lists the deliverable version updates of the DMP for DRIVE.

Table 1. DRIVE DMP deliverables

Deliverable no.*	Deliverable name	WP no	Short name of lead participant	Type	Dissemination level	Delivery date**
4.2	Generic DMP	4	P95	R	PU	M6 (DEC 2017)
4.8	Mid-term DMP	4	P95	R	PU	M30 (DEC 2019)
4.9	Final DMP	4	P95	R	PU	M60 (JUN 2022)

DMP = Data management plan; WP = work package; R = Document, report; PU = public.
*According to DRIVE Description of Action document, page 28.
**Measured in months from the project start date (July 2017, Month 1)

The DMP follows the principles that research data are findable, accessible, interoperable and reusable (FAIR)[1].

The general principles on access rules are defined in the Consortium Agreement (Section 8 Intellectual property – Access rights).

3 Overview of data managers, data repository and access rules

Two data repositories/platforms will be used in the DRIVE project:

- One data repository (DRIVE research server) hosted by P95 which will be used to store all study related datasets produced and/or shared within WP7, with limited access.
- One platform, an electronic study support web application, which will be used to store the metadata related to all study datasets in the DRIVE research server. This platform will be developed during the first year of the DRIVE project, and will be described in detail in the next revision of the DMP.

The contact details for the data management team are described in Table 2.

Table 2. Data management team contact list

Responsability	Name	E-mail address
Data management compliance / Server admin	Tom De Smedt	Tom.desmedt@p-95.com

[1] European Commission Horizon2020 programme. Guidelines on FAIR Data Management, v3.0, 2016. (http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

Deputy data management compliance	Roberto Bonauiti	roberto.bonaiuti@unifi.it
Deputy data management compliance	Topi Turunen	Turunen_top@gva.es

The following generic user roles will be defined:

- Server administrator
- Study lead
- Data owner
- Data analyst

It is possible for a single person to take on multiple generic roles.

The study lead is responsible for the conduct of the study, ensuring the adherence to the protocol, statistical analysis plan (SAP) and study procedures.

The data owners are the database custodians and are the responsible at the local data sources that have the necessary data to participate in the study. The local data providers are responsible for extracting the requested data out of their database(s) according to the study protocol and study-specific DMP and importing the extracted data in the DRIVE research server using a secure connection.

The data analyst uses the imported data to perform the necessary data transformations following the protocol and pooled analysis SAP. The resulting output on aggregated data (tables, figures) is exported from the DRIVE research server by the data analyst for further analysis and communication as per protocol and SAP procedures, led by the study lead within WP7.

The server administrator of the DRIVE research server is responsible for the following tasks:

- Set-up and maintenance of the DRIVE server (permissions, security, logs, updates, etc.)
- Responsible for user management (registration of local data providers and data analysts, two-factor authentication, connection set-up and monitoring)
- Validate uploaded data from local data providers
- Transfer files between DRIVE server compartments
- Validate files flagged for export by data analysts
- Perform privacy assessments where necessary with every data transfer step (DPO certification necessary)

Access to the server will only be granted to DRIVE consortium members and associated partners that are part of WP7, or part of the quality and audit committee as defined in the CA. Access can be requested by completing an intake template document, detailing necessary information about who is requesting access (i.e. a data analyst, etc.) and the purpose (for which study, for audit purposes etc.). This document should be sent to the server admin, who will decide on granting the access together with the other members of the data management team. The intake document also contains contact details necessary for getting access with two-factor authentication. The intake template document can be found in Annex 1.

3.1 DRIVE research server

A secure repository to store datasets will be provided by P95. Within this repository, research sites will be able to upload study specific datasets in order to be able to perform pooled data analysis. This is a highly secure environment and network, with strict rules for data access. And the infrastructure is in accordance to the new GDPR guidelines on storing personal identifier data and in a processor-role, as well as storing anonymized data.

3.1.1 Specifications

The repository consists of:

- Dedicated secure virtual server on redundant cluster
  - Physical connectivity: 100Mbit/s
  - Traffic: 100GB/month
  - Hardware: 4 CPU cores
  - 8 GB RAM Memory
  - 500GB Storage Memory
  - Location: Belgium
- Backup storage platform and necessary licenses
- Installed software and licenses
  - Windows Server 2016
  - MS Office 2016
  - Microsoft SQL Server
  - Remote desktop licenses
  - R
  - SAS
- Monitoring of operating system
  - System update status
  - System log files
  - Access to system, scheduled batch job status
  - Memory
  - Disk space and status
  - Monitoring status
  - Backup status
  - Running services
  - Additional application maintenance on the webserver configuration and SSL certificate status
- Server availability Monitoring & Reporting (SLA)
  - 24/7 monitoring and reporting of system availability and security
  - Weekly server security scan and necessary corrective actions
- Two-factor authentication using DUO Access
- Additional server set-up documentation
  - URS document
  - Workflow specifications (functional, non-functional & design specifications)
  - Workflow validation

3.1.2 Procedures/tools for data accessibility/security

Details of all users of the repository must be registered with P95. Access will be restricted to a minimum number of people.

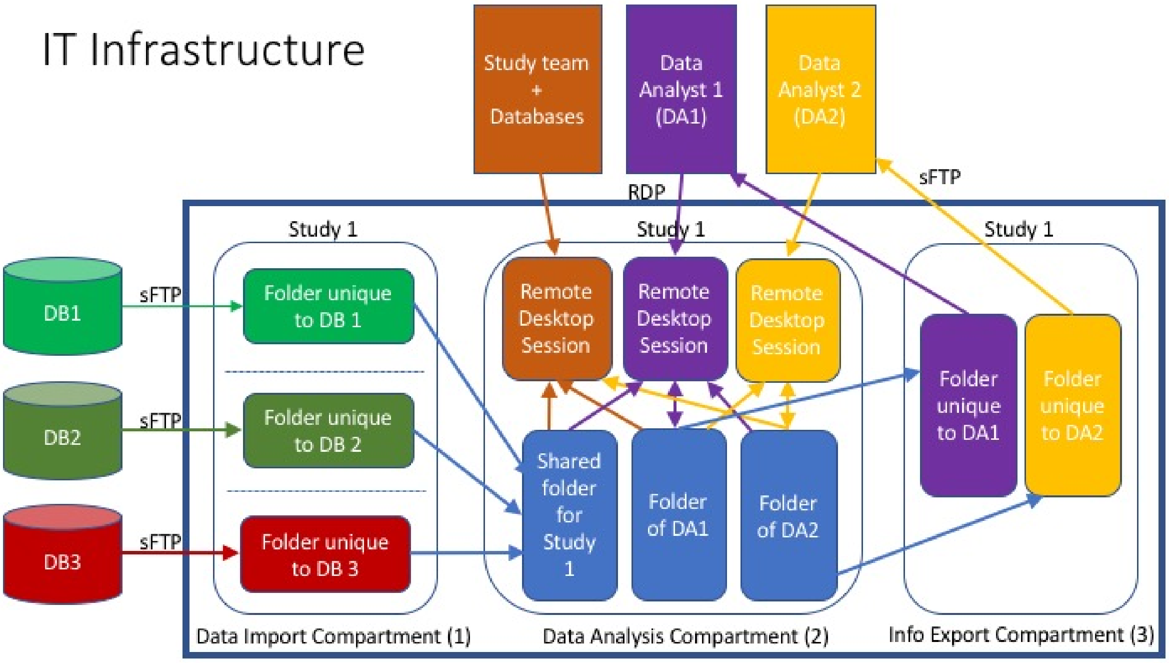
The architecture of the DRIVE research server is shown below (Figure 1).

The described IT architecture allows for all types of data (i.e. individual record level data, anonymised record level data and aggregated data) to be used and allows for multiple studies to be carried out at the same time.

The proposed IT architecture is:

- Scalable
- Secure
- Transparent
- Can use all formats of data
- Adheres to General Data Protection Regulation (GDPR)

Figure 1. Architecture of the DRIVE research server



The general architecture of the DRIVE research server has three compartments: the data import compartment, the data analysis compartment and the info export compartment. The DRIVE research server is only accessible through the secure file transfer protocol (with upload capability to the data import compartment and download capability out of the data export compartment) and the remote desktop protocol allowing (primarily) the data analyst role to log into the data analysis compartment. The transfer of any data between the different compartments is done solely by the server administrator role, where data privacy assessments should be carried out if deemed necessary. Every interaction on the DRIVE research server will be logged, and these logs will be accessible upon request.

Following the procedures specified in the study-specific protocol and SAP, the local data providers extract data out of their respective database(s), perform the necessary data quality checks, and upload the data through a secure connection using the secure file transfer protocol and store this raw data (this data can potentially take any form or file format, i.e. aggregated data or individual record level data; plain text data or relational database without any IT restriction from the DRIVE research server) in a folder unique to each local data provider in the data import compartment of the server.

The server admin checks the imported data from the local data providers, i.e. assess the data correctness and completeness, whether the data is of the correct type, correct numbers of records, the anonymization and/or aggregation was performed correctly, etc. If necessary, the server admin performs additional data privacy assessments.

When the uploaded data is correct following protocol details, this uploaded data from the local data providers is transferred by the server admin from the data import compartment to a shared folder unique to the study in the data analysis compartment. The server admin sets the permission to this data in this shared folder as read-only for the data analyst.

The data analyst can log in to the data analysis compartment using the remote desktop protocol and two-factor authentication. Each data analyst that has gone through the registration process and is accepted as a data analyst on the server, has its own unique folder(s) for each study where the data analyst is involved. The data analyst can transfer data from the shared folder unique to the study (and holding the data the server admin put there) to its own folder unique to the study. Following protocol and SAP details, the data analyst can perform data transformations using the appropriate software tools (f.e. R, Excel, Stata etc.) and perform quality control on the created code and files. After the data transformation step is complete, the data analyst can flag the resulting files (f.e. aggregated data, graphs, code) for export out of the DRIVE server and notify the server admin of the flagged files. The data analyst has read and write access to its own folders, has read access to folders from other data analysts in the study and has read access to the shared folder of the study.

Upon notification of flagged files by the data analyst, the server admin checks these flagged files: correct data following protocol details and if necessary perform a privacy assessment. When the flagged files are checked, these files are transferred from the data analysis compartment to the info export compartment in a folder unique to the data analyst.

The data analyst can export the files from the info export compartment of the DRIVE server to its own local computer using the secure file transfer protocol.

3.1.3 Duration of accessibility

Authorized users will have access to the DRIVE research server during the course of the DRIVE project, and as long as the user is a member of the project.

3.1.4 Back-up process & Disaster recovery

There will be on site and off-site backups of the server OS, the application code and data:

- Daily system block level encrypted backup
- Retention of 30 daily backups
- Reporting on yearly manual backup restore test of server and on deleted backups on request
- Backup locations:
  - Interxion Datacenter, Wezembeekstraat 2, 1930 Zaventem
  - Uniweb BVBA, ’s Herenweg 16, 1860 Meise

3.1.5 Archiving and preservation

All study related data in the DRIVE research server produced as the results of the DRIVE project will be kept for 5 years following the completion of the DRIVE project as specified in the Consortium Agreement (Article 18). The details of the physical location of the archive (if different from the location used during the project) after the completion of the project as well as access will be developed in future versions of the DMP.

3.2 Electronic study support web application platform

This platform will be developed during the first year of the DRIVE project. A full description of the platform will be added in the next version of the DMP.

4 Overview of data types generated and collected in DRIVE

DRIVE will generate/collect data containing personal identifiers. Within personal data, we can differentiate between individual level datasets and aggregated datasets. The individual level datasets should be sufficiently anonymized so there is no risk for re-identification using this data. These data may be primary data produced by consortium partners and associated partners or secondary data from existing registries or databases. All personal data will be generated within the activities of WP7.

The WP4 data management team will generate a meta-data repository of all datasets in collaboration with data owners and WP7 leads. This repository will be updated regularly.

5 Operational data management requirements for DRIVE research projects

Each individual study within DRIVE will need to provide a short study-specific DMP, either as part of the protocol or as a separate document.

In addition, IMI requires a dataset template to be completed for each dataset described in the protocol or SAP of each study. This dataset template requires at the minimum the following:

- Dataset reference and name
- Dataset description
- Standards and metadata
- Data sharing policy
- Ethics and legal issues
- Data protection, IP/copyright and ownership

This dataset template shall be completed when the dataset is first imported and the whole set of completed dataset templates shall be reviewed every six months.

5.1 Requirements for the short dataset specific DMP

All data owners will be required to fill in the metadata template shown in Table 3 below for each dataset. These metadata are specifications for data that provide the contextual information required to understand those data. The template will be made available in the server. Each completed table will be reviewed by the data management team for completeness, compliance with the DMP and with the Consortium Agreement.

Table 3 Metadata requested per dataset (adapted from the Data Management General Guidance of the DMP Tool)[2]

General Overview	
Title	Name of the dataset or research project that produced it
DRIVE task	DRIVE task/subtask where dataset was generated
Data owner	Names and addresses of the organizations or people who own the data
Identifier	Unique number used to identify the data.
Start and end date	Study start and end date
Time period covered by the dataset	Start and end date of the period covered by the dataset
Methods	How the data were generated (e.g. primary data collection, registry, study design, etc.), listing equipment and software used (including model and version numbers)
Type of data	Datasets containing personal data Datasets containing non-personal data
Processing	How the data have been altered or processed (e.g. normalized), including de-identification procedures
Source	Citations to data derived from other sources, including details of where the source data is held and how it was accessed
Funder	Organizations or agencies who funded the research, or indicate that the data owner funds the study
Content description	
Subject	Keywords or phrases describing the subjects or content of the data
Language	All languages used in the dataset
Variable list and codebook	All variables in the data files, with description of the variable name, length, type, values
Data quality	Description of data quality standards and procedures to assure data quality
Technical description	
File inventory	All files associated with the project, including extensions
File formats	Format of the file

[2] https://dmptool.org/dm_guidance#metadata

File structure	Organization of the data file(s) and layout of the variables, where applicable
Checksum	A digest value computed for each file that can be used to detect changes
Necessary software	Names of any special-purpose software packages required to create, view, analyse, or otherwise use the data
Access	
Rights	Any known intellectual property rights, statutory rights, licenses, or restrictions on use of the data
Access information	Where and how your data can be accessed by other researchers
Data sharing	Description of how data will be shared, including access procedures
Ethics and legal issues	Description of any ethics and legal issues associated with the dataset, if any

5.2 Responsibilities of the data owner

Data owners per study and dataset will be identified in the dataset metadata. The data owner of the respective datasets must ensure and is responsible to comply with all legal and ethical requirements for data collection, handling, protection and storage. This includes adherence to regulations, guidelines such as (but not limited to) the EU clinical trial directive 2001/20/EC, Good clinical practice (GCP) and Good Pharmacoepidemiology Practice (GPP), as applicable.

6 Sharing and secondary use of DRIVE generated or collected data

6.1 Procedures for making data findable

The information collected and updated via Table 3 will be available in the electronic study support web application. This will enable the easy identification of datasets available and identify the data owner.

6.2 Re-use of data within the DRIVE consortium

To achieve the objectives of DRIVE, it is imperative to follow the collaborative approach the partners agreed on when signing the consortium agreement. This includes the necessity to share data from the individual studies for the implementation of the DRIVE project, while respecting data protection and intellectual property of the partners' work. For those individual studies within DRIVE that need to use data generated in another DRIVE task, the metadata will contain the data owner contact details to whom a requester can reach out if they need to access the results.

6.3 Re-use of DRIVE results by third parties

For those external individuals/institutions wanting to use DRIVE generated or collected data during the course of DRIVE, the Data Management Team should be contacted (Table 2). Given the nature of the studies conducted in DRIVE, that aim to use secondary data to conduct pooled data analysis, only access to pooled aggregated results datasets will be considered, since individual study data are not owned by DRIVE. Giving access to external parties will be considered by the Steering Committee on a case by case basis. Access rules for the time after DRIVE termination will be worked out and described in the final DMP.

7 Protection of personal data

The collection of personal data will be conducted under the applicable international, IMI, and national laws and regulations and requires previous written informed consent by the individual, i.e., with public and commercial entities and if applicable outside the EU in countries with lower data protection standards. To obtain the agreement of participants of studies to use their data for secondary research, the following lines can be included in the consent form:

- I understand the information collected about me will be stored in secure database, which will be used for future research.
- I authorise the research to use my anonymised study data for additional medical and/or scientific research projects.

DRIVE may also utilize purely register-based data that is collected as part of routine surveillance or clinical practice. Because of the nature of these data and the large number of study subjects, it is often not possible to obtain informed consent in these cases. However, all the other considerations related to ethics, data security and protection apply.

DRIVE researchers commit to the highest standards of data security and protection in order to preserve the personal rights and interests of study participants. They will adhere to the provisions set out in the:

- General data protection regulation (GDPR), foreseen coming into effect in 2018[3]
- Directive 2006/24/EC of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communication services or of public communications networks[4]
- Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the

[3] http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32006L0024&from=en
[4] http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32006L0024&from=en

electronic communications sector (Directive on privacy and electronic communications)[5]

- Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data[6]

Prior to collecting, storing, and processing sensitive personal data, the consortium will seek approval of the applicable local and/ or national data protection authorities. Consent forms will contain information on how personal data will be managed. To secure the confidentiality, accuracy, and security of data and data management, the following measures will be taken:

- All personal data obtained in DRIVE studies will be transmitted to partners within the consortium only after anonymization. Keys to identification numbers will be held confidentially within the respective research units. In situations were re-identification of study participants becomes necessary, for example the collection of additional data, this will only be possible through the research unit and in cases where informed consent for such cases has been given.
- Personal data are entered to secure websites. Data are processed only for the purposes outlined in the patient information and informed consent forms of the respective case studies. Use for other purposes will require explicit patient approval. Also, data are not transferred to any places outside the consortium without patient consent.
- None of the personal data will be used for commercial purposes, but the knowledge derived from the research using the personal data may be brought forward to such use as appropriate, and this process will be regulated by the Grant Agreement and the Consortium Agreement, in accordance with any generally valid legislation and regulations.

The following points to consider will guide the protection of data within the DRIVE project: (i) The entity providing personal data to the project shall verify that:

- the initial collection of these data has been compliant with the requirements of the original purpose
- the collection and the provision of the data to the project meets all legal requirements to which the entity is subject
- further storage and processing of the data after completion of the research project is in compliance with applicable law

(ii) The entity which provides personal data to the project shall document any restriction of use or obligation applicable to these data (e.g., the limited scope of purpose imposed by the consent form). The entity which uses personal data in the project shall be responsible to ensure that it has the right under the applicable data protection and other laws to perform the activities contemplated in the project.

Personal data shall always be collected, stored, and exchanged in a secure manner, through secure channels.

8 Ethical aspects

DRIVE partners and associated partners are required to adhere to all relevant international,

[5] http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32002L0058&from=en

[6] http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31995L0046&from=en

IMI, and national legislation and guidelines relating to the conduct of studies. Research ethics in DRIVE are described in detail in deliverable 6.2 (Report with the definition of the ethics policies handbook collection).

9 List of abbreviations

CA	Consortium Agreement
DMP	Data Management Plan
DoA	Description of Action
DRIVE	Development of Robust and Innovative Vaccine Effectiveness
EU	European Union
GDPR	General Data Protection Regulation
IMI	Innovative Medicines Initiative
SAP	Statistical Analysis Plan
WP	Work Package

Annex 1. Data access request form

DRIVE Research Server application instructions

The DRIVE Research Server maintained by P95 CVBA in the DRIVE Research Consortium allows researchers to work together in a secure environment on studies developed in the DRIVE Research Consortium with the possibility to work with distributed data.

Access to the DRIVE Research Server is secured by two-factor authentication using Duo Access using a mobile phone running either the Android or iOS operating system or using a desktop application.

Please fill in the form and send it as a PDF by email to tom.desmedt@p-95.com (cc roberto.bonaiuti@unifi.it and turunen_top@gva.es) and add [DRIVE RESEARCH SERVER ACCESS REQUEST] to the subject heading, or send it by regular mail to the address below.

Tom De Smedt
Server Admin DRIVE Research Server
P95 CVBA
Koning Leopold III laan 1
3001 Heverlee
BELGIUM
Email: tom.desmedt@p-95.com
Tel: +32 (0) 472 21 65 65

DECLARATION REGARDING THE ACCESS TO THE DRIVE RESEARCH SERVER

The user requests access to the DRIVE Research Server using the two-factor authentication procedures set forward by the DRIVE Consortium in WP4.

Rationale to request access

User Responsibility Statement - The signatory hereby commits to use the access to the DRIVE Research Server and the accompanying two-factor authentication procedure only in the framework of the project and only to carry out its assigned tasks as described in the study protocol and related documents.

The signatory declares to adhere to the project study protocols and to be aware of the security measures for the use of the DRIVE Research Server. The signatory is aware that no data can be copied or used for other purposes. This includes a commitment not to transfer the access credentials to anyone else inside or outside his/her organisation and to take the necessary steps to prevent the access credentials from being misused in any way.

In case the signatory detects any use which is not compliant with the terms above he/she commits to inform **immediately** the DRIVE Server Admin by email to tom.desmedt@p-95.com

APPLICANT INFORMATION

First Name	***Last Name***
Organization	***Department***
Street	***Postal Code***
City	***Country***
Telephone Number	***Email Address***
Date of Application	***Signature***