

D2.3 Electronic study support application

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), Roberto Bonaiuti (4 – UNIFI)

Due date	30 JUNE 2018
Delivery date	11 JULY 2018
Deliverable type	OTHER
Dissemination level	PU

Description of Work	Version	Date
	V1.0	05 JUL 2018

Document History

Version	Date	Description
0.1	31 MAY 2018	First draft
0.2	12 JUN 2018	Updated draft following internal feedback
0.3	18 JUN 2018	Draft for SC review
0.4	05 JUL 2018	Updated following feedback SC
1.0	05 JUL 2018	Updated following feedback TT and RB

Table of contents

<i>Document History</i>	2
<i>Table of contents</i>	3
1 Introduction	4
2 Specifications	4
3 Functionality	5
3.1 <i>Uploading data</i>	5
3.2 <i>Data upload check</i>	5
3.3 <i>Data visualizations</i>	6
3.4 <i>Data aggregations per view</i>	6
4 Glossary	7

1 Introduction

In the DRIVE project, influenza vaccine effectiveness (IVE) data from several independently operating national and regional study sites will be analysed together to obtain brand-specific IVE estimates.

The DRIVE Electronic Study Support Application facilitates this process. It is a web application accessible from any web browser serving the following purposes:

- Aiding the research sites in uploading their datasets to the DRIVE Research Server using a secure connection in a user-friendly manner
- Allowing research sites to have a quick glance at their uploaded data and check its correctness and completeness
- Summarizing the uploaded data in various high-level statistics

2 Specifications

Web application developed in R 3.4 using the additional packages shiny and highcharter (for more info see Glossary). Hosted on a P95-controlled server running shiny-server. The connections to the web application are ensured to be secure through the use of a SSL-certificate (Let's Encrypt Certificate Authority, see Glossary). Authentication is managed through the auth0 software (see Glossary).

There are two different tiers of users:

- General user: Every user can look at the “General View” with overview statistics on the of the aggregated uploaded data from all research sites
- Site-specific user: For each research site, each user could get accredited to also look at the lower-level statistics for each dataset from this research site

Upon loading the web application, the users wanting to log in are presented with a landing page where the username and password are to be put it. The web application checks the put in username and password combination with the combinations saved in a flat file on the server. The flat file looks like:

Username	Password	Type	Site
JohnDoe	xxxxxx	General	General
JohnSmith	xxxxxx	Site	Site1
JackPeters	xxxxxx	Site	Site2, Site3

Table 1 - Example of flat file for user authentication

The user authentication is ensured by the auth0 software installed on the server hosting the DRIVE Electronic Study Support Application

The application has two distinct views:

- “General” view: one aimed at general users and site-specific users with general high-level info on the uploaded data of all sites, i.e. how much datasets were uploaded in total, etc. and general info on the uploading process and about DRIVE
- “Site-specific” view: view that is site-specific and is only available to the user that have access to this site’s data, where the user can look at the datasets uploaded by that site, statistics on each dataset uploaded by that site and can upload new data of that site.

This application is also following the work on data protection detailed in D4.2 and D6.2.

3 Functionality

3.1 Uploading data

The web application allows the research sites to upload their datasets by clicking on a dedicated button. Uploaded data is subsequently securely stored on the DRIVE Research Server for further analysis. This functionality is provided on each “Site-specific view”

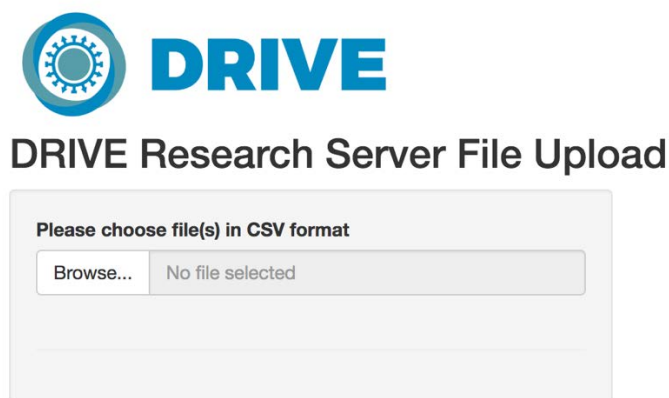


Figure 1 - Screenshot of DRIVE Electronic Study Support Application Upload functionality

3.2 Data upload check

Upon upload, the uploading research sites can immediately look at the uploaded data and adjust different parameters (headers, format, etc.) to ensure the upload data is correct and complete.

Upon upload the uploaded file is also checked to comply with the minimal dataset requirements (see D7.1 and D7.2). This step highlights the following non-compliances:

- Inconsistent naming of columns / unexpected columns
- Unexpected data type in a column

In order to facilitate the data collection at the research sites, the ESSA will provide an Excel template mimicking the minimal dataset requirements and example datasets for download by the users.

3.3 Data visualizations

Given the uploaded data by the research sites is following the minimal dataset requirements, they can be used to create the following data visualizations; either using a single uploaded dataset, or an aggregation using multiple uploaded datasets:

- Number of data subjects (table, time series line chart)
- Number of influenza positive / negative samples (table, time series line chart)
- Number of vaccinated / unvaccinated (table, time series line chart)
- Age distribution (bar chart; time series stacked bar chart)
- Sex distribution (bar chart; time series stacked bar chart)
- Symptom distributions (bar chart; time series stacked bar chart)
- Vaccine brands used & their distribution among data subjects (table, bar or pie chart; time series line chart)
- Proportion of patients with chronic conditions, pregnancy, frailty status (table, bar or pie chart; time series line chart)
- Proportion of missing data (in each variable) (table, bar or pie chart; time series line chart)
- Spatial distribution (map, bar chart by region)
- Indication of start and end of surveillance
- Hospital/site distribution (bar chart)
- Proportion of comorbidities (bar chart)
- Are all columns present in the uploaded data? (i.e. compliance with minimal dataset requirements)

3.4 Data aggregations per view

As discussed above, the two tiers of users are accredited to look at statistics based on different aggregations of the uploaded data. We have a general overview for all data combined and a general overview for each research site. We also have specific overview for each uploaded data for each site. Each of the views can use the data visualizations listed in Data visualizations

- General View
 - General overview for all datasets: calculated using the aggregation of all uploaded datasets across all research sites
 - General overview for each research site: for each research site, we use the aggregation of all uploaded datasets for that research site
- Site-specific View:

- Specific overview for each dataset: each uploaded dataset for the research site detailed in this site-specific view

4 Glossary

- auth0: widely used Auth0's Universal Identity Platform for web, mobile and IoT. <https://auth0.com/> (please see <https://auth0.com/about> for more info on copyright)
- Let's Encrypt is a free, automated and open Certificate Authority. <https://letsencrypt.org/>
- R – R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- shiny - Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie and Jonathan McPherson (2017). shiny: Web Application Framework for R. R package version 1.0.5. <https://CRAN.R-project.org/package=shiny>
- highcharter - Joshua Kunst (2017). highcharter: A Wrapper for the 'Highcharts' Library. R package version 0.5.0. <https://CRAN.R-project.org/package=highcharter>
- Minimal dataset requirements – DRIVE minimum dataset requirements as drafted by WP7.