

Variant Exporter User Documentation

**Author:** Melvin Luong

**Status:** Final

**Version:** 1.1

**Date:** 23rd July 2014

Introduction 3

Requirements 3

Download and Installation 4

Setup of Data Source Settings 8

Database File Settings 8

Database Server Settings 9

File Settings 10

Setup Gene and RefSeq 11

Add a new Gene 11

Edit an existing Gene 12

Remove an existing Gene 13

Disease Tags 14

View all Disease Tags to a selected Gene 14

Adding a Disease Tag on to a Selected Gene 15

Delete a Disease tags on a Selected Gene 16

Sending Data 17

Loading Data for sending 17

Loading data from Database File 18

Loading data from Database Server 19

Loading data from a File (.csv, .xls, .xlsx, .txt) 19

Loaded Data in the Send Data Tab 20

Invalid Data 20

Sending Data to Test server 21

Web Proxy Setup 23

Glossary 24

F.A.Q 25

# Introduction

The Variant Exporter is an application that will extract clinical variant data from various data sources such as database files, database servers and various file formats such as “.xls” spreadsheet file or a “.csv” comma separated values file. The data extraction process will validate and remap the data if necessary to comply with HVP standards, encrypted sent to HVP servers where the data will be decrypted, validated again and imported into the HVP Australian Node Database.

This user documentation will help guide you through the installation and setup the Variant Exporter application.

# Requirements

The following are the basic requirements in order to run the Variant Exporter application adequately:

* Preferably a multi-core processor (more cores increases the performance for larger dataset extraction).
* 1 GB of RAM or higher (the higher the better)
* Application needs at least 45 MB of free disk space.
* Windows XP or higher.
* Microsoft .NET 3.5 Framework installed.
* If reading from a data source that is in Microsoft Excel Spreadsheet “.xlsx” file format, then Microsoft Office 2007 or higher needs to be installed. If you are using “.xlsx” file and don’t have Office 2007 installed then you need to install a data connection driver: <http://www.microsoft.com/en-us/download/details.aspx?id=23734> (NB: you may need admin rights to your PC to install this).
* The PC you are installing on must have Internet connection.

# Download and Installation

Here are the steps to download, install and set up the new exporter:

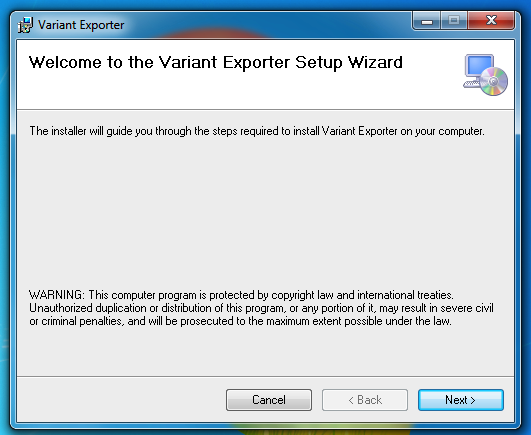
1. Download all the required the files:

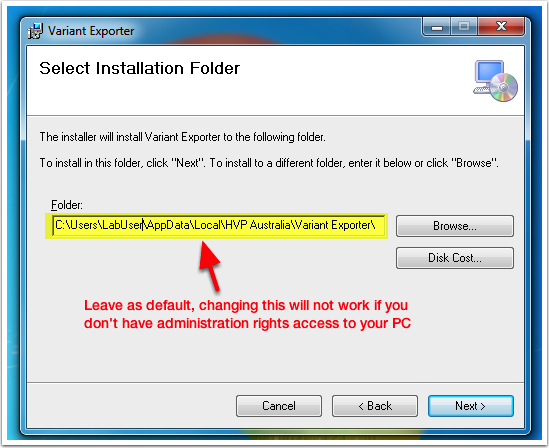
a) Download the installation file from the HVP servers.

b) You will also need a private key. This will be sent out separately with a link to download the file.

NB: The download links will be sent out by HVP. If you need to download the files again contact HVP for the links.

1. Run the installation file and follow the wizard. Just leave the settings to their default and click “Next” until you reach the end. (Unless you have administration rights to your PC it’s best to leave the installation path to the default user. If you are unsure, just leave it as default.)

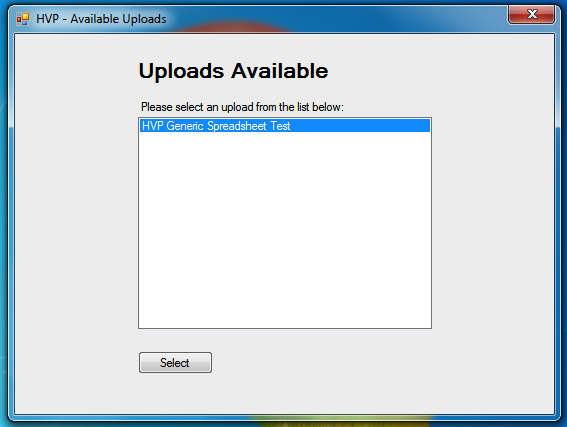




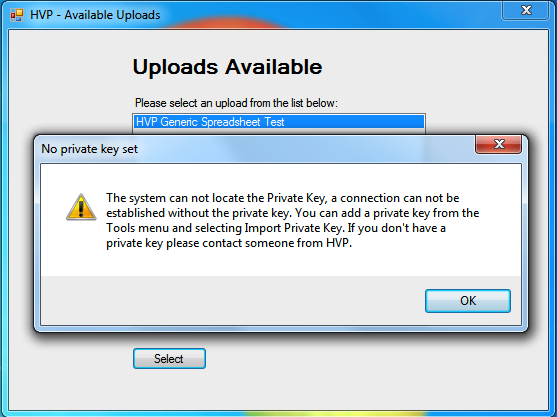
1. Once the installation has finished you will find a shortcut to the “Variant Exporter” application on your “Desktop” or in your “Start Menu “under “Programs”.

Now that you have the Variant Exporter application installed you will need to import your private key.

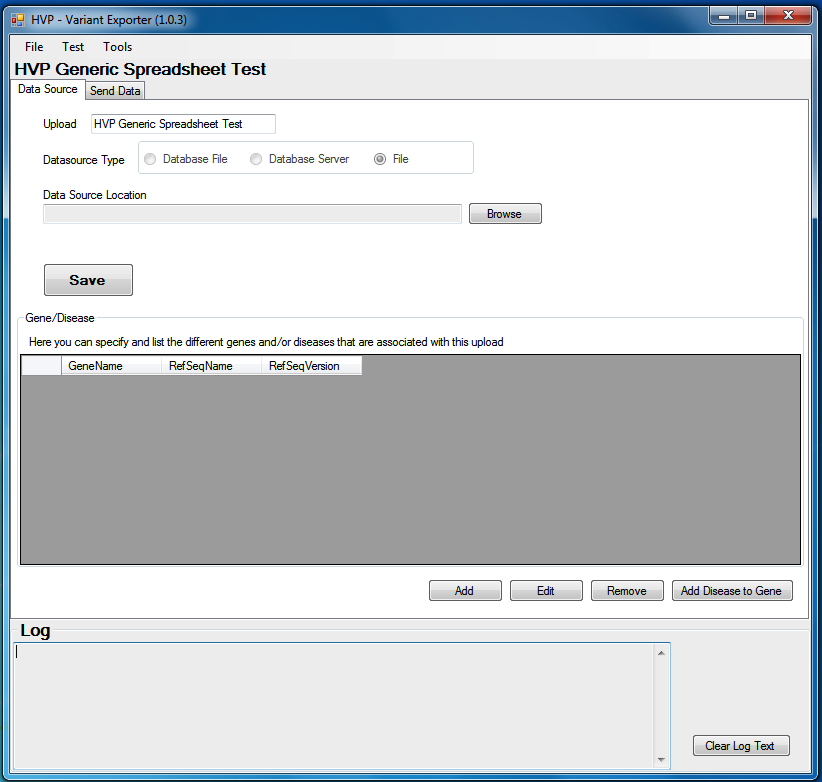
1. Start the Variant Exporter, once the application has loaded you will be presented with the “Upload” window with the list of uploads that are specifically available for your site.



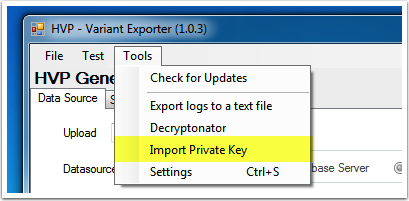
Select an upload and click the “Select” button. You should see a prompt about “No private key set”, click “OK” when you see that message.

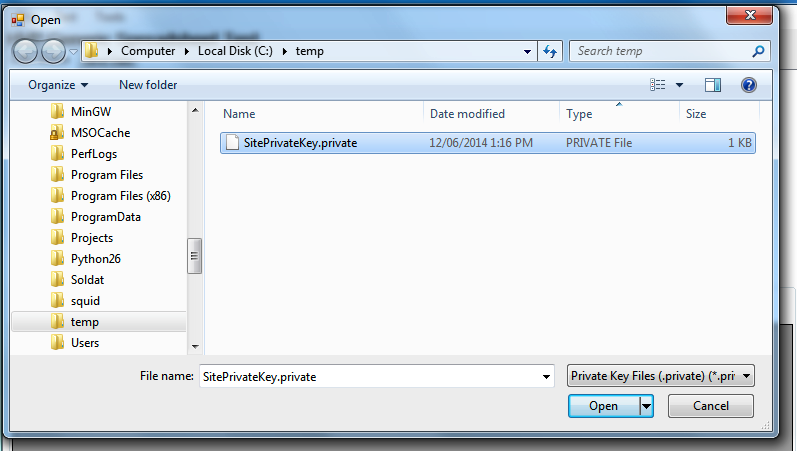


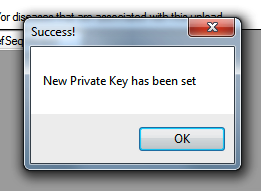
The next screen you will be shown is the main exporter interface.



1. To import the private key go to the “Tools” menu at the top of the Exporter and click on the “Import Private Key”. It will then ask for the location of the private key that you downloaded earlier. Find the location of that file and open it.







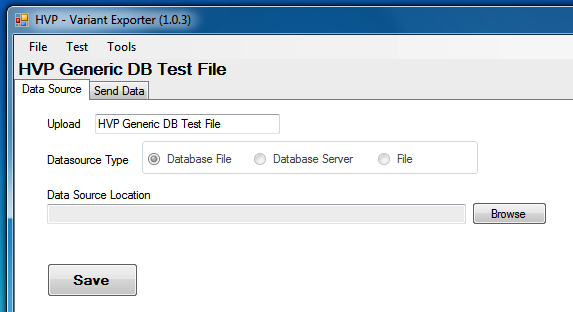
# Setup of Data Source Settings

Now that you have setup the private key the only thing left is to configure your upload data source and test the connections to ensure everything is setup correctly for you to begin sending data to HVPA Database.

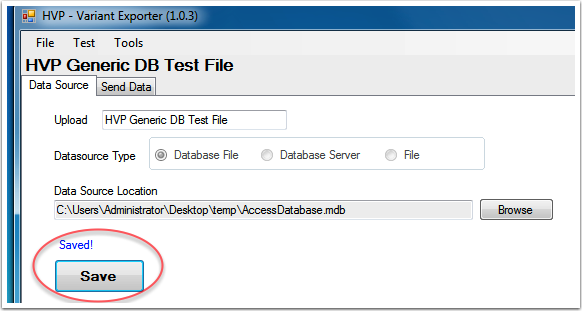
Depending on the type of data source you are using the settings in the “Data Source” tab may differ:

## Database File Settings

For a Database file, click on the “Browse” button next to the “Data Source Location” to select database file (e.g. “.db”, “.mdb”, “.accdb”) the exporter will be reading from. Click the “Save” button to save your settings when you are done.

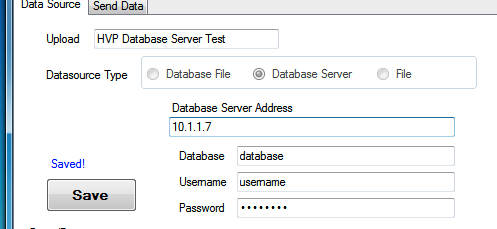






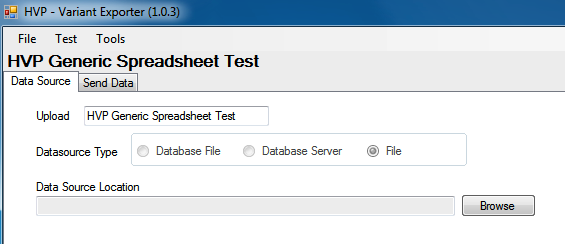
## Database Server Settings

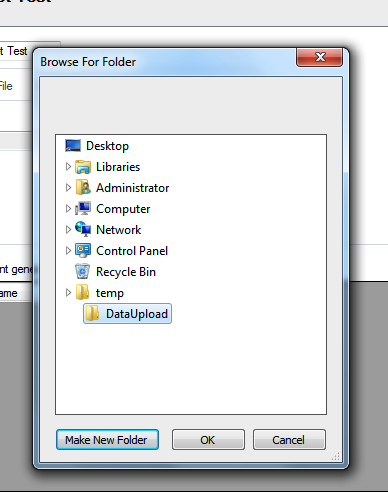
For reading data from a database server (e.g. Microsoft SQL Server, MySQL, PostGreSQL, Firebird, Oracle) you will need to enter in the connection settings (NB: You may need to contact your IT administrators to get the connections settings). Click the “Save” button to save your settings when you are done.



## File Settings

For a flat file format, you need to select a folder where the files will be kept as the exporter will read the files in that folder (only valid files with data it can read). Click on the “Browse” button next to the “Data Source Location” to select a data source folder. Click the “Save” button to save your settings when you are done.



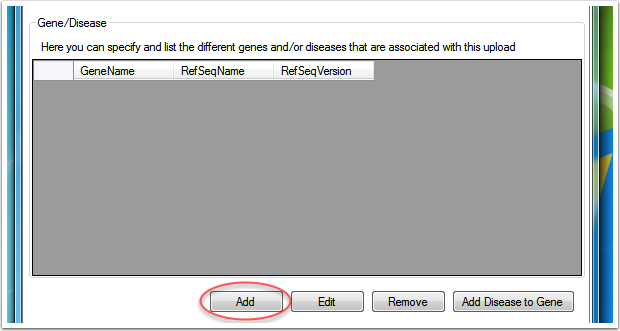


# Setup Gene and RefSeq

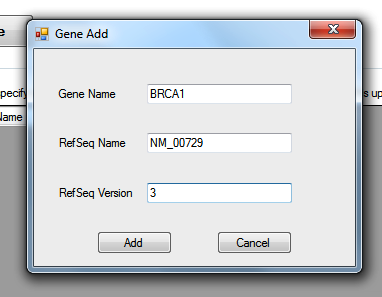
This table defines the Gene name, RefSeq name and RefSeq version of the uploading variant data being uploaded to. You need to list all of the genes included in the variant data upload.

## Add a new Gene

To add a list of Genes, RefSeq names and RefSeq versions to the upload click the “Add” Button under the “Gene/Disease” table.

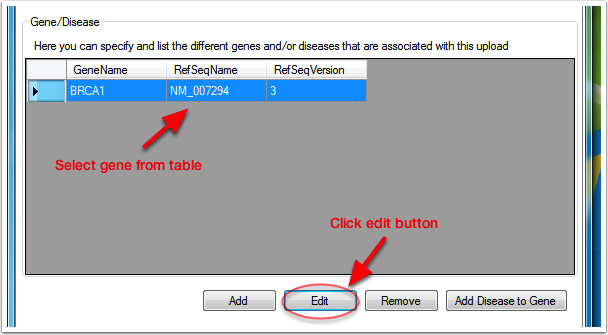


This opens up a new “Gene Add” window. From here you can enter in the Gene name, RefSeq Name and RefSeq Version. Click “Add” to save the Gene to the Gene table.

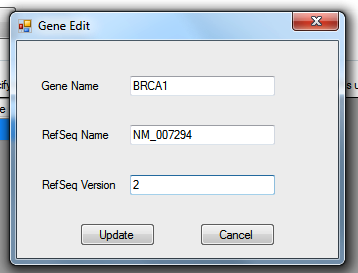


## Edit an existing Gene

If you want to make changes to an existing gene in table, select that gene by clicking on it in the table, then click the “Edit” button.

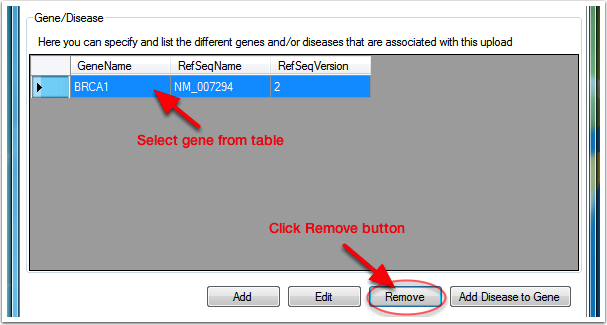


This will open the “Gene Edit” window and from here you can make any changes to the Gene and then click the “Update” button to save those changes.

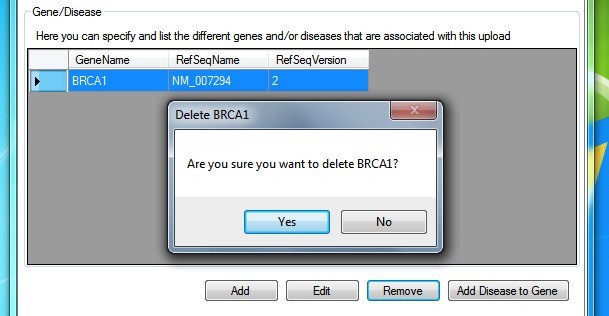


## Remove an existing Gene

If you want to remove an existing gene in the table, select that gen by clicking on it in the table, then click the “Remove” button.



You will get a prompt to confirm the removal of the selected gene. Clicking on “Yes” will remove that gene and clicking “No” will cancel the action.

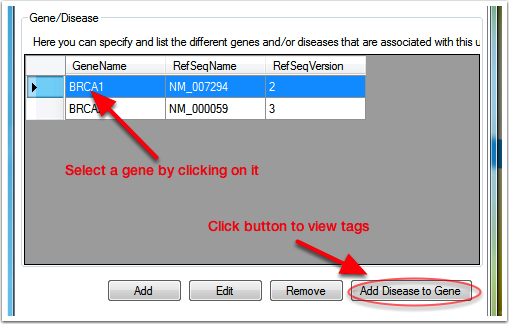


# Disease Tags

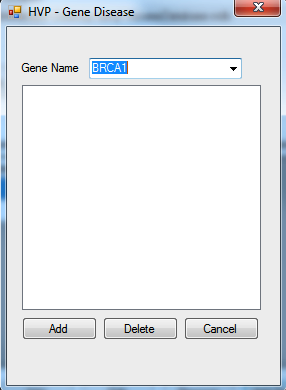
Disease Tags are used to enhance search queries on the HVPA Node Portal website. By tagging related diseases to genes it provides meta information for the gene record. You can add as many related disease tags to a gene as you like. This is an optional feature, but by providing the information here can greatly enhance the ability and searching capabilities for users who view the variant data on the HVPA Node Portal.

## View all Disease Tags to a selected Gene

To view all the disease tags to an existing gene first select the gene from the “Gene/Disease” table by clicking on the gene, once a gene is selected click on the “Add Disease to Gene” button.

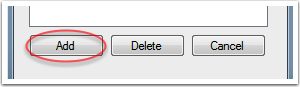


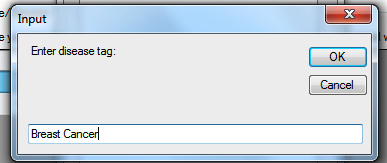
This should open a new window that will display all the current diseases that are tagged to the selected gene.



## Adding a Disease Tag on to a Selected Gene

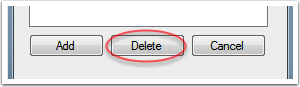
With the “Gene Disease” window open, click on the “Add” button on the button left. A new window will prompt for a “Disease Tag”. When you have entered a value into the field click “Ok” the tag will be added to the “Gene Disease” window list.

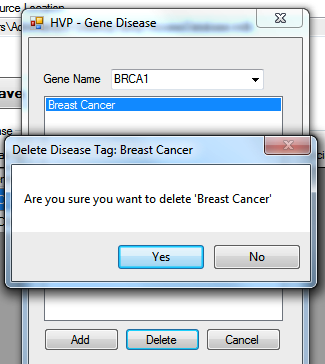




## Delete a Disease tags on a Selected Gene

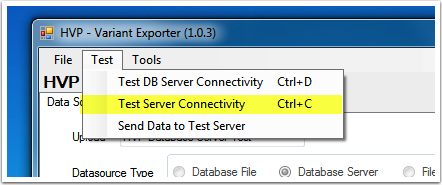
With the “Gene Disease” window open, select a tag from the list by clicking on the tag. With a “Disease Tag” selected click on the “Delete” button, you will be prompted with a confirmation before deleting the tag. Click on “Yes” and the tag will be deleted.





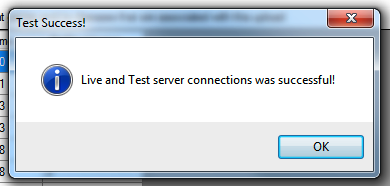
# Sending Data

Before sending any data you should check to see whether the application can communicate with the HVPA servers. To test this go to “Test” in the top menu and click on the “Test Server Connectivity” item.



This will then begin a test connection where the Exporter will send some dummy data to the server and await a response from the server to determine if it was successful or not. The connection test will test 3 things:

* Can the Variant Exporter connect to the Internet?
* Can the HVPA servers at the other end decrypt the sent data using your site’s private key?
* Are the HVPA live and test server up and running?

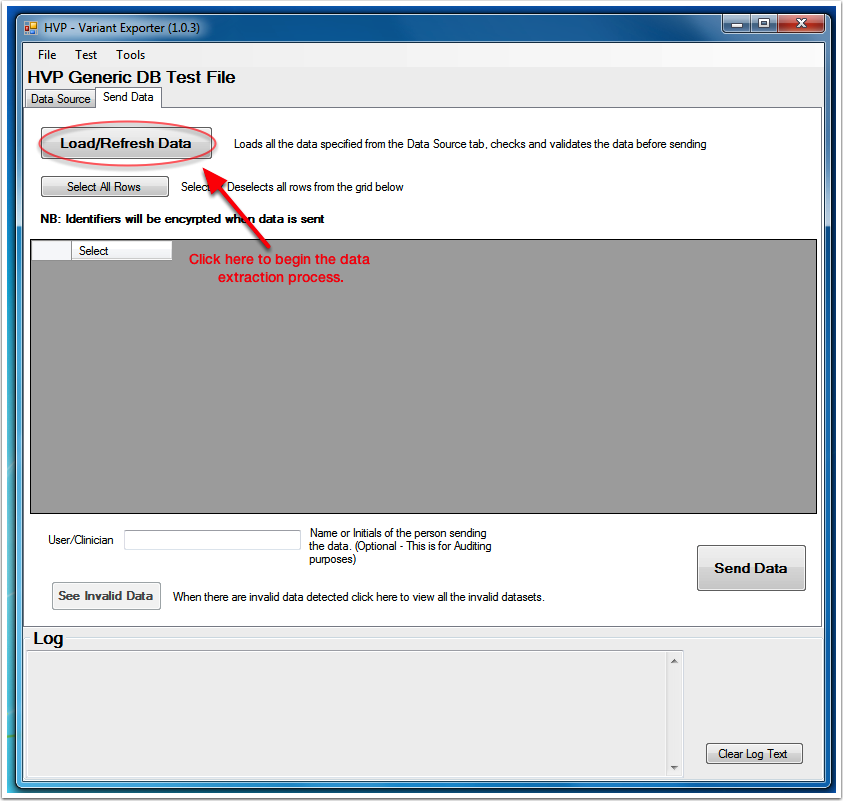


## Loading Data for sending

Once the private key has been imported to the Exporter and the upload data source settings have been configured the next step to before you start sending data is to load the data into the exporter. Depending on the data source type the process may differ.

## Loading data from Database File

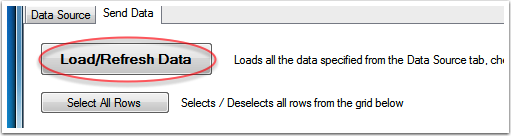
Go to the “Send Data” tab by clicking on the second tab.



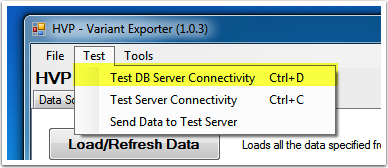
Click the “Load/Refresh Data” button. This will begin the data extraction process. Depending on how big the dataset is and the specs of the PC this may take quite some time. Please be patient and do not interrupt the application whilst it is loading.

## Loading data from Database Server

The Database server data extraction process works similar to the Database file method. Clicking on the “Load/Refresh Data” button the Exporter will begin to connect to the DB server and extract the data.

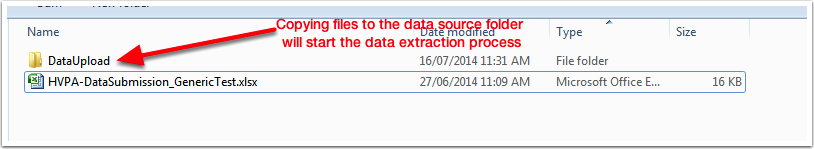


(NB: To ensure the exporter can correctly connect to the Database server and to test that the configuration settings for the Database server are correct, click on the “Test DB Server Connectivity” under the “Test” menu. This test will do a simple check that the exporter can communicate with the Database.)



## Loading data from a File (.csv, .xls, .xlsx, .txt)

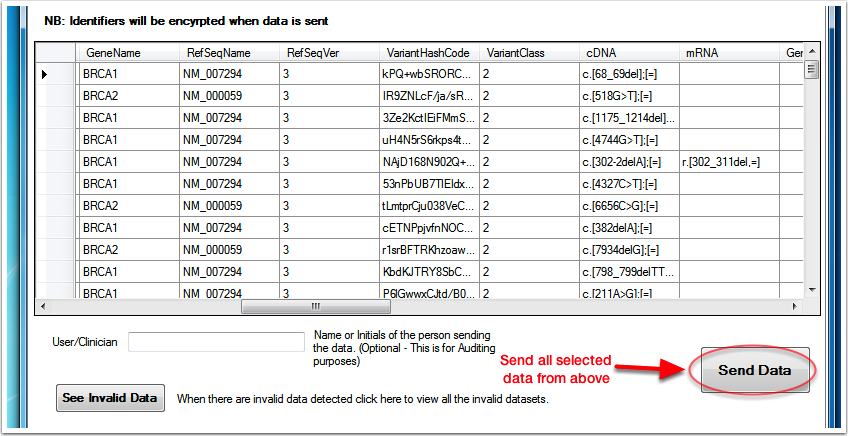
For file based uploads the workflow and concept is different from the Database File and Database Server method. With file based uploads you nominate a folder to put the files into. With the Exporter application running in the background copy the file or files with the data you wish to upload to the data source folder. The exporter should be able to automatically detect the files and instantly start the data extraction process.



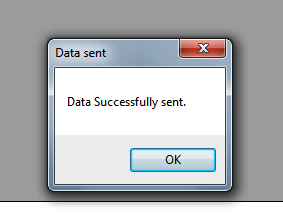
(NB: Because the Exporter can only focus on a single file upload at a time, if you have multiple files in the data source folder the Exporter will only load 1 file at a time with all the data extracted to the Send Data tab. Clicking on the Load button after sending data from a file will load up the next file in the folder.)

## Loaded Data in the Send Data Tab

Once the data extraction process has been completed you should see the “Send Data” tab table populated. Here you can inspect the data you sending to HVPA. You can manually select which records you want to send by checking the “Select” box. (NB: Unique Identifiers and Patient demographics displayed in table will not be sent, the details are displayed on the table for referencing)

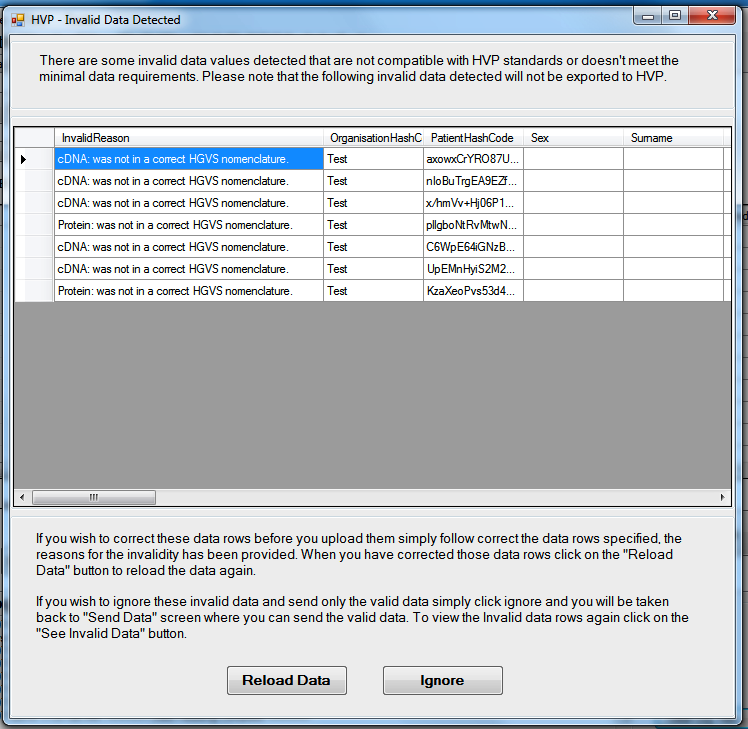


To send the data that has been selected in the table click on the “Send Data” button.

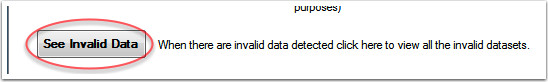


## Invalid Data

When the data extraction process finishes you may also be prompted with another window titled “Invalid Data Detected”. This window will contain a table with a list of invalid data. Invalid data are data that don’t meet the minimum required dataset or contain typos. Invalid records will not be sent.

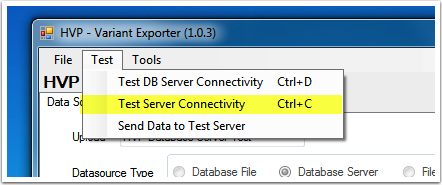


You are given the option to use this opportunity to go back and correct any typos or adjust values or simply ignore the invalid data and continue on. If you have closed the “Invalid Data” window and wish to see the list again click on the “See Invalid Data” button on the “Send Data” tab page to load it up again.

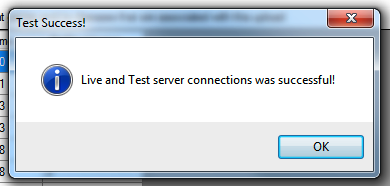


## Sending Data to Test server

Data can be sent to a Test server during the testing and development phase of the project. Data will be sent securely in the same manner as the regular sending method but instead the date will be sent to the HVPA test server where it can be verified by someone at HVPA.

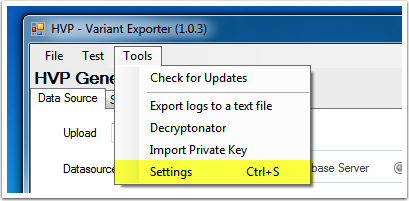


To send data to the Test server instead of the live server, go through the same procedures as you normally would to load the data into the “Send Data” tab. Once the data extraction process has finished and you have the data loaded and ready to send, go the “Test” menu and click on “Send Data to Test Server”. The data will be submitted to the test server.

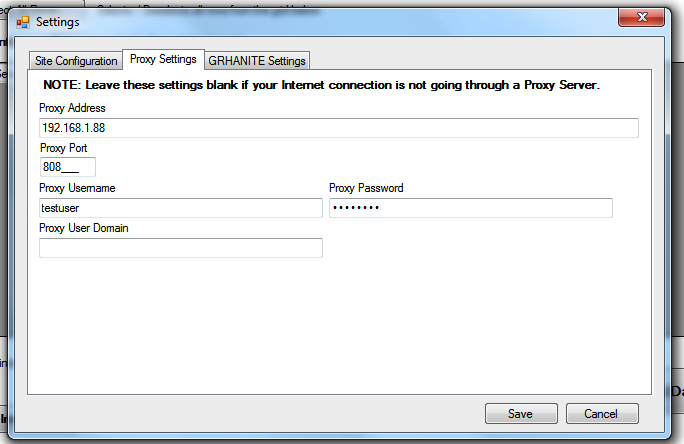


# Web Proxy Setup

By default the Variant Exporter should be able to detect your proxy settings, but if you need to manually set the proxy settings you can by going into the “Tools” menu and clicking on “Settings”.



This will open a separate “Settings” window. In this window click on the “Proxy Settings” tab. Here you can configure the Proxy Address, Proxy Port, Proxy User Domain, Username and Password.



When you finish configuring the proxy settings click on the “Save” button to save and close the Settings window.

# Glossary

Data Encryption/Decryption - This is a process of encoding any readable message/data into a format that can not be read by unauthorised parties. Decryption is the reverse process of deciphering the encoded message/data to readable form.

Data source - This is source of where the variant test data is recorded to. This can be a Database such as Microsoft Access, MySQL server or Oracle or it could be stored in a flat file such as Excel Spreadsheet, CSV or a text file.

Database File - A Database that is contained within a file. (e.g. Access database ".mdb" file, sqlite ".db" file, FileMaker ".fp7" file.)

Database Server - A Database that is deployed on server and requires a remote connection to access the data. You may need to contact you IT administrators to get the database connection details. (e.g: MySQL, MS SQL Server, PostgreSQL, Firebird, Oracle)

Flat File - A simple file that stores data typically in a table structured format such as ".xls" spreadsheet file, a comma-separated values ".csv" file, a extensible markup language ".xml" file or any sort of file that has a structured table element in it.

Private Key - An encryption key that is uniquely given to each site/lab. This key is used to securely encrypt the data that is sent from the users PC to HVP.

Upload – In context of this application an “Upload” is a custom configured plugin that defines the type of data source the exporter is reading from and any associated rules it applies to remap the data that is then compatible with the structure of the HVP database format. If you have multiple data sources that are structured or formatted differently you may have more than one upload. For example you have data that is stored in Database server and excel spreadsheet file, this that case you will have 2 uploads.

# F.A.Q

**Where do I download the Variant Exporter installation file and Private Key?**

After registering to upload data to HVP, you should received instructions on how to download and setup the application. But if you need to download the installation file and private key again, please contact [hvpa@hvpaustralia.org.au](mailto:hvpa@hvpaustralia.org.au) to get the files.

**The Exporter doesn’t seem to launch when I copy files to the data source folder?**

- Make sure the folder location in the “Data Source Location” are the same.

- Make sure you click save after setting the “Data Source Location” folder.

- Check the files you are copying are in the correct file format.

**The Variant Exporter fails the “Test Server Connectivity” test or can’t seem to connect to the HVP server?**

If you are unable to connect to HVP you need to check whether or not your Internet connection is going through a web proxy, you may need to check with your IT administrator to find out your proxy settings and proxy authentication.

The exporter should be able to automatically pick up and configure a web proxy if you are using one, but if you still encounter issues try manually entering in your proxy settings. (See Web Proxy Setup on how to configure your proxy settings)

**What happens to the Files in the folder after the data has been sent?**

In the case of File uploads, after a Files data has been successfully extracted and sent to HVP, the file will be moved to a sub-folder inside your data source folder called “sent”.

Files in the sent folder will also have a date stamp appended to the file name (e.g. “variant\_data.xls” will become “variant\_data(20140630143859).xls”), this makes it easier to see when the files where sent.

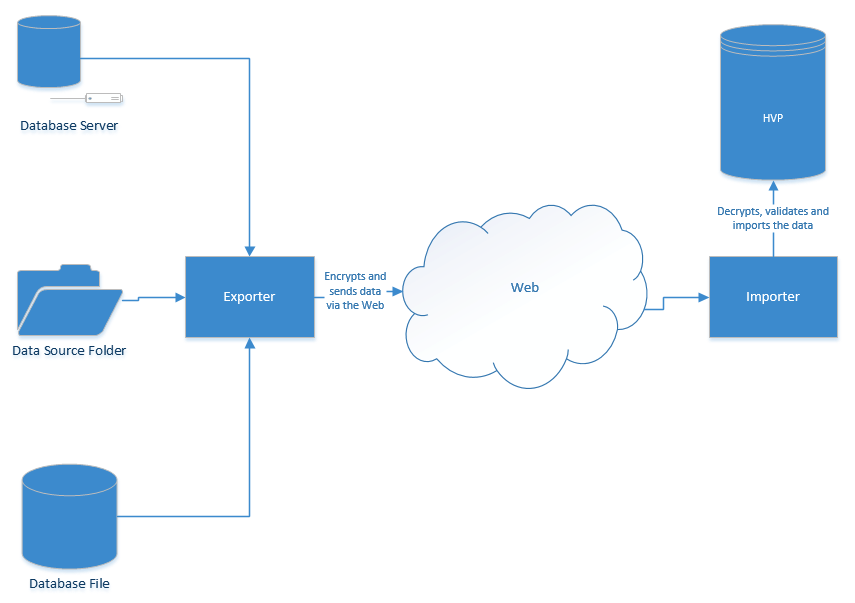
The timestamps units are grouped together in the following order: year, month, day, hour, minute and seconds. (e.g. 20140630143859 reads Year: 2014, Month: 06(June), Day: 30, Time: 14:38:59(HH:MM:SS).

So for example if you are putting your files on a folder on your “C:\” drive called “HVP Data Submissions” then files sent will be placed inside a “sent” folder.

(e.g. “C:\HVP Data Submissions\sent\”)

**What happens to the data I’ve sent?**

The data you send to HVP will be encrypted and then zipped out before they are sent to HVP. Once the files reach the HVP servers they are decrypted and validated again before the data is imported to the HVP Australia Node Database.



**How does the application keep track of what I’ve sent?**

The Variant Exporter application has a logging feature that can track what data has been sent to HVP therefore it can avoid resending data that has already been sent before, on the HVP server end it also does another check to ensure the data being uploaded is unique and not a duplicate of a past upload.

(NB: When sending data to the test server rather than the live server no data logging or tracking will be recorded.)

For “File” uploads there is no accurate way to keep track and log the data. Unique ID’s from files can’t be guaranteed to be unique as other files may reuse the same ID again. But in the event of a duplicate data submission to HVP servers, the server will be looking at the data to compare existing records to determine whether it is a duplicate submission based various unique ID fields (Patient ID’s, URN’s, Test ID’s, Lab ID’s), test dates, submission dates, sample source, patient age, patient gender etc.