

abYsis v4

Installation and Data Loading

Table of Contents

Introduction	3
Linux Server Setup	3
Migrating to a New Version	4
Sequence and Structure Export	4
Installing on the same server	5
Installation	7
Setup	7
Configure	7
Configuration file details	8
Install	9
Securing abYsis	13
Loading Third Party Data	14
IMGT/GENE-DB (Germline) dataset	14
Obtaining the data	14
Loading the data	14
IMGT/LIGM-DB (Large-scale) dataset	15
Obtaining the data	16
Loading the data	16
OGRDB (Germline) dataset	17
Obtaining the data	17
Loading the data	18
Deleting IDs for a given Data Source	19
Post-translation Modifications	20
Configuration file	20
Displaying the new annotations	21
Appendix. Installation Error Codes	22

Introduction

Currently, we support version 9 of the Rocky or RHEL operating systems.

You must use one of these.

Do NOT try to copy and paste commands from the PDF version of this document! Characters such as '-' and spaces may not translate properly and this could lead to strange error messages and the commands not working.

This guide assumes:

- You setup the server as described in the **Linux Server Setup** section
- **Hardened versions of Rocky and Red Hat can cause problems** during installation and abYsis cannot support the additional installation complications that may be experienced on these types of servers.
- The user performing the installation has a basic knowledge and experience of Linux systems and can edit files, navigate directories and files, and run scripts and commands.

Linux Server Setup

- A. The server you wish to install abYsis on will need to be running Rocky 9.x or RHEL 9.x or above. 'Minimal' is recommended – everything else you need will be installed by our install process.
- B. The server minimal requirements are 2-core, 80GB free disk space, 4GB memory. Recommendations are 4-core, 100GB+ and 8GB memory.
- C. This server will require access to the internet during installation to download OS packages and the abYsis software and database.
- D. The person doing the installation must be able to log in as root.

The installation script will carry out the following Linux actions.

- *create a standard user which will manage the abYsis installation.*
- *install additional Linux packages and Perl packages when not already available.*
- *create directories and files*
- *configure, enable and start the **apache web service***
- *set **SELINUX** to permissive mode*
- *disable the machine **firewall** as your machine will be within your company firewall*
- *add a **systemd service** to automatically start the **postgres** database*

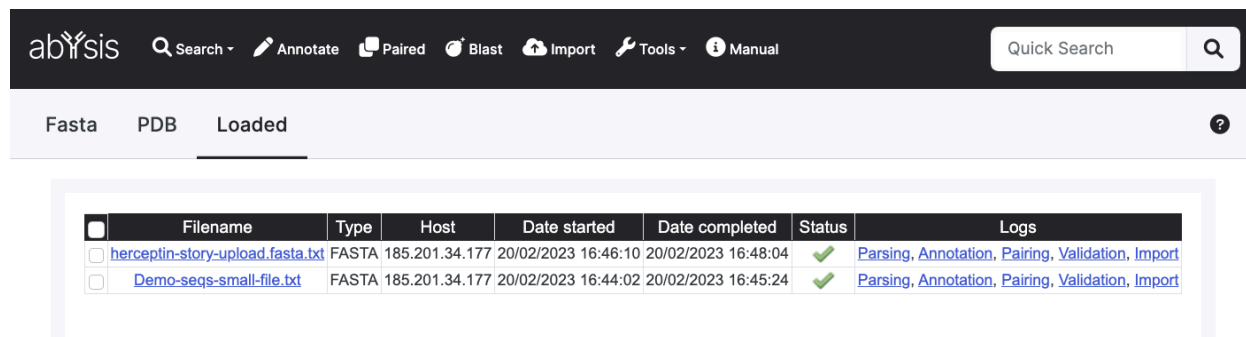
Migrating to a New Version

If you have an earlier version of abYsis installed please first read this extra section carefully to understand how to migrate to the new version.

Sequence and Structure Export

This new installation will build a brand new database so any sequences added to an existing system using the **Import** facility **will not be ported** to the new installation.

If you have imported sequences you will see Filename entries in the Loaded Tab of the Import Function. If there are none showing this means no sequences are loaded.



The screenshot shows the abYsis web interface. At the top, there is a navigation bar with the abYsis logo and several menu items: Search, Annotate, Paired, Blast, Import, Tools, and Manual. A search box is also present. Below the navigation bar, there are three tabs: Fasta, PDB, and Loaded. The Loaded tab is selected. Below the tabs, there is a table with the following columns: Filename, Type, Host, Date started, Date completed, Status, and Logs. The table contains two rows of data:

Filename	Type	Host	Date started	Date completed	Status	Logs
herceptin-story-upload.fasta.txt	FASTA	185.201.34.177	20/02/2023 16:46:10	20/02/2023 16:48:04	✓	Parsing , Annotation , Pairing , Validation , Import
Demo-seqs-small-file.txt	FASTA	185.201.34.177	20/02/2023 16:44:02	20/02/2023 16:45:24	✓	Parsing , Annotation , Pairing , Validation , Import

If you have sequences (or structures) that you wish to preserve use in the new installation you must follow the steps:

- Go to the **Import** section and select the Loaded tab
- For sequences select **Prepare Fasta**. For structure select **Prepare PDB**.
- Wait while the file builds (can take several minutes if a lot of data has been uploaded)



The screenshot shows a section titled "Download uploaded data". Below the title, there are two buttons: "Building file..." and "Prepare PDB".

- When the button changes to **Download**, click and the data will be downloaded to your local computer

Download uploaded data

Download Fasta

Prepare PDB

- Keep the file safe. You will be able to use it to upload the data to your new installation

Installing on the same server

If you are installing this new version of abYsis *on the same server* as your existing installation then follow these steps to deactivate the old version.

- Go to the root directory of the installation. It is usually something like

```
cd /home/abysis/abysis_x.y.z
```

- Go into the database directory and switch off the database connection. Your existing abYsis installation will still be intact, but switched off.

```
cd database_server
```

```
sudo -u abysis ./stop.sh
```

- If you are using a version earlier than v3.4.0 then run the following command to disable the server link to the web page.

```
sudo mv /var/www/html/abysis /var/www/html/abysis.orig
```

- For the new install it is advisable to use a different port number. To see the port number of your disabled PostgreSQL use the following command.

```
grep ABYSIS_PGPORT /home/abysis/abysis_x.y.z/abysis.cfg
```

Note: If you intend to run 2 versions simultaneously you must use a different `ABYSIS_WWWROOT` name in your `abysis.config` file for your new installation. For instance

Existing: `ABYSIS_WWWROOT=/abysis`

Existing URL: `http://yourserver/abysis`

New: `ABYSIS_WWWROOT=/abysisv4`

New URL: `http://yourserver/abysisv4`

Do **not** nest your new installation inside an existing one. **For instance, the following will cause errors.**

Existing: `ABYSIS_WWWROOT=/abysis`

Wrong: `ABYSIS_WWWROOT=/abysis/v4`



Correct: `ABYSIS_WWWROOT=/abysisv4`



Installation

Setup

1. Log in to your server as **a user that has full sudo permissions** (we advise against using **root**) and go to the user's home directory

```
cd
```

2. Download the Installation package from our distribution site, substituting `USERNAME` and `PASSWORD` for the username and password you were provided for downloading abYsis.

```
curl -O https://USERNAME:PASSWORD@info.abysis.org/distribution/InstallScripts.tar
```

You may need to install curl if you are using a minimal install OS.

```
sudo yum install -y curl
```

3. Unpack the installation package to your home directory.

```
tar xf InstallScripts.tar
```

You may need to install tar if you are using a minimal install OS.

```
sudo yum install -y tar
```

4. Enter the installation directory.

```
cd InstallScripts
```

Configure

5. The configuration file `abysis.config` contains settings for the installation.

Edit `abysis.config` to set options appropriate to your installation. You can use `vi`, `nano` or any other text editor. On a minimal install OS you may need to install the editor.

```
sudo yum install -y <editor package>
```

To configure the installation you will need to either:

- Set `DOWNLOAD_USERNAME` and `DOWNLOAD_PASSWORD` in `abysis.config` with your the username and password provided to you for downloading the abYsis code and database (recommended)

Or

- If your server does not have internet access:
 - manually download the code (`abysis.tar.bz2`) and database (`abysis_data.tar.bz2`) from
<https://USERNAME:PASSWORD@info.abysis.org/distribution/abysis.tar.bz2>
https://USERNAME:PASSWORD@info.abysis.org/distribution/abysis_data.tar.bz2
 -
 - locally copy the files onto the server
 - set both `SOFTWARE_DOWNLOAD` and `DATA_DOWNLOAD` in `abysis.config` with the full paths to the local files.

Configuration file details

For a fuller understanding, here is a listing of settings and options contained in `abysis.config`. Under normal circumstances the only configuration settings you would need to change have already been explained elsewhere in this document.

ABYSIS_USERNAME the abYsis system user (**default: `abysis`**) that will own the installed software. If it does not exist then it will be created. It can be an existing user, but the user cannot be root.

ABYSIS_DIRECTORY the directory for the abysis installation. The abYsis system user must have full access to this directory (**default: `/home/abysis`**).

WEB_DIRECTORY the directory within web root (`/var/www/html`) where the system will reside – this must begin with a forward slash `'`

e.g. `/myabysis` will install the abYsis website at `http://yourserver/myabysis` (**default: `/abysis`**).

DOWNLOAD_USERNAME/DOWNLOAD_PASSWORD set your username and password that was provided to you for downloading the abYsis software and database.

UPLOAD_USER/UPLOAD_PWD The upload facility on your abYsis server can be protected by a username and password. You can change this or keep the default (**default: UN:abysis and PW:upload**).

SOFTWARE_DOWNLOAD if you needed to manually download the abYsis software (bz2) file, it would need to be located in a directory accessible to the abysis user. Manual download is generally not recommended and by default is left blank. (**default: blank**).

DATA_DOWNLOAD if you needed to manually download the abYsis database file (bz2) file, enter the full path to its location here. Manual download is generally not recommended and by default is left blank. (**default: blank**).

CONFIGURE_HTTP configuration is required in order to run CGI software; set to 'no' if your system administrator wishes to set this manually (**default: yes**).

DISABLE_SELINUX setting security-enhanced linux to permissive is required for this installation. If your system administrator wishes to do this manually then set to 'no' (**default: yes**).

DISABLE_FIREWALL configuration of the firewall may be required if it is set up. Choose 'yes' to do so automatically or 'no' to do so manually (**default: yes**).

SOFTWARE_VERSION this should already be set to the version you are installing.

POSTGRESQL_PORT the port allocated to the instance of PostgreSQL. This may require changing if you have multiple instances of abYsis running, or if you already have an instance of PostgreSQL (**default: 5432**).

SYSTEM_USER only set this to yes if you wish the abysis user to also be a system user.

DISTRIBUTION_DIR where installation logs are put.

(**default: /root/InstallScripts/install_abysis_4.x.x/abysis_install.log**)

Install

6. Start the install. Please run this in the background by ending the command with '&' in case your connection to your server times out.

```
./install_abysis.sh &
```

7. The install process can take up to 2 hours and the screen may appear inactive at times.

If installation error codes occur please look up the code with information in the Appendix. You may be able to fix the issue easily.

During the installation the server will be checked for compatibility. Required operating system packages will be updated or installed. The software will then be downloaded and installed followed by the database files. Third party software such as PostgreSQL will also be installed and will generate a lot of messages to the screen - do not be concerned about this. The database will then be populated and finally, the web server will be configured.

8. During the installation a directory for logs and backups will be created.

```
/root/InstallScripts/install_abysis_4.x.x
```

This directory will contain the logs from the installation and backups of files that may have been changed on your system such as:

- **abysis_install.log** – a log file of the installation process
- **httpd.bak** – the original /etc/httpd/conf/httpd.conf (if modified)
- **iptables.bak** – the original /etc/sysconfig/iptables (if modified)

9. On completion, the terminal window will present data such as:

```
===== Installation Completed =====  
===== Obtaining MAC address  
===== MAC ADDRESS: 01:23:45:67:89:AB  
===== Please send this MAC address to 'licence@abysis.org' to obtain a licence file.  
===== Your licence code will be emailed to you with instructions on how to deploy it.
```

```
Public visibility of the installation
```

```
Public web site      : http://nnn.nnn.nnn.nnn/abysis/ is [is not] visible
```

```
Local network location of the installation
```

```
Private web site    : http://172.17.0.1/abysis/ is [is not] up
```

```
--- Completed: macip
```

```
===== SCRIPT COMPLETE =====  
==== End of script ./install_abysis.sh
```

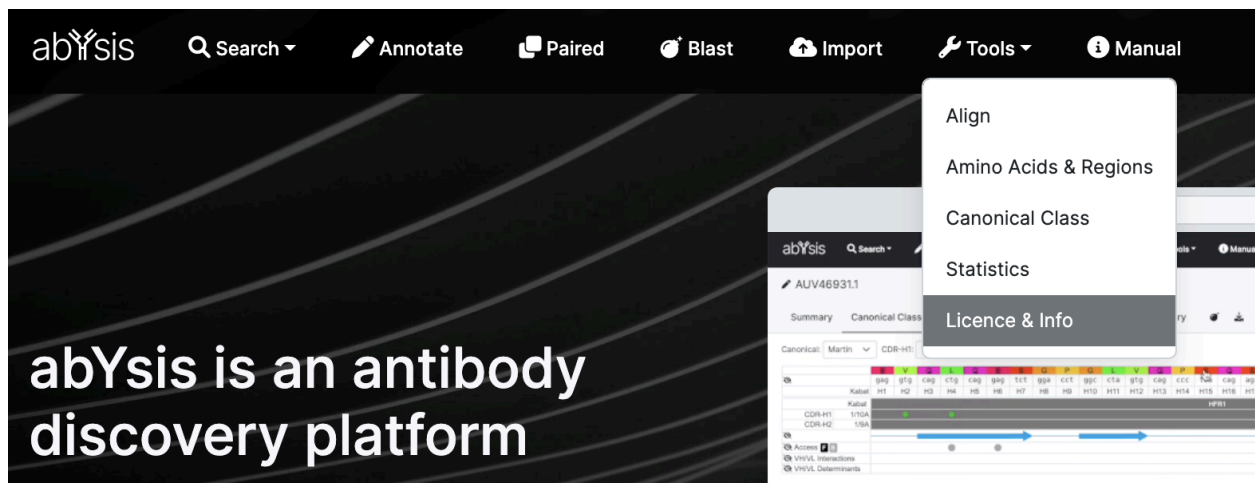
On one of the final lines printed to the screen, you will be given the machine's MAC address that will look something like this:

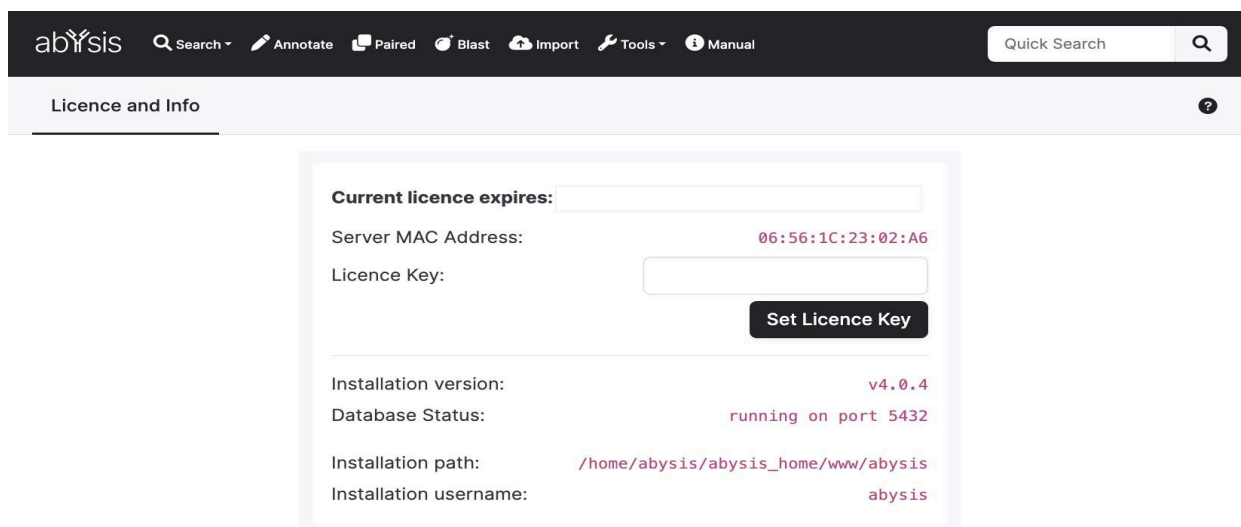
```
01:23:45:67:89:AB
```

If for some reason you cannot see this. There are two other ways of getting the MAC address:

- From the terminal do the following two commands,

```
cd /home/abysis/abysis_home/www/abysis  
./getmac.pl
```
- Alternatively, on the **Licence & Info** page of your abYsis installation you will see your MAC address displayed as in the figure below:





To obtain a licence key to activate your installation, **please send this MAC address to licence@abysis.org**. Your licence will be sent to you with instructions on how to activate your installation.

When you have received a licence key activate it through the abYsis web interface on the **Licence & Info** page of your abYsis installation.

10. You will also be presented with the public and private/local URL of your server and whether your website is accessible or not. Example below.

Public visibility of the installation

Public web site : http://nnn.nnn.nnn.nnn/abysis/ is not visible

Local network location of the installation

Private web site : http://nnn.nnn.nnn.nnn/abysis/ is up

These network permissions are controlled by your IT administration and are not associated with abYsis. If you cannot see them or are unsure, you should contact your IT administrator. You may wish to ensure that they are appropriate for your requirements such as changing the public web site to be not visible for instance.

11. If an installation fails and you wish to contact, please email and include a copy of the abysis_install.log (Section 7 above) so that we can analyse.

The default email address is `licence@abysis.org`

Securing abYsis

When abYsis v4 is installed it is by default available on an HTTP connection and protected by whatever firewall the host server sits behind. If you wish to add additional security then we recommend reviewing the document "Securing abYsis v4" which can be found on the abYsis Info web page (<https://info.abysis.org>).

Loading Third Party Data

Beyond the data sources distributed with abYsis it is possible to load other resources. Use of third party resources should be independently validated by the user.

All of the commands below should be run by a **user with full sudo permissions**.

IMGT/GENE-DB (Germline) dataset

This section describes loading **IMGT germline** sequence data into abYsis.

It is assumed that you have completed the abYsis installation and have the database running.

Obtaining the data

You will need the `imgtrefseq.fasta` file downloaded to your computer.

IMGT data are available from locations such as the EBI, where files can be found in the following directory at the time of writing:

```
https://ftp.ebi.ac.uk/pub/databases/imgt/
```

Germline dataset:

```
https://ftp.ebi.ac.uk/pub/databases/imgt/ligm/imgtrefseq.fasta
```

Loading the data

To parse and load the IMGT germline sequence data change to the **manage_data** directory and run the update command for IMGT germline data.

```
cd /home/abysis/abysis_home/manage_data
./update_imgtgene.sh /tmp/imgtrefseq.fasta &
```

The script will take several minutes to run.

Note: If you wish to delete any existing IMGT germline upload before running this script,

see the section **Delete all Source IDs for a given Data Source** below.

When the script completes, and if you are finished loading all your data, (i.e. not planning to load other databases described in this section) you will need to refresh the germline and BLAST databases and site statistics. Go to the *website directory* and run the build script.

```
cd /home/abysis/abysis_home/www/abysis
```

```
./build.sh
```

Do not do this step if you intend to immediately upload another dataset, such as IMGTLIGM-DB. This step should be done after all updates have occurred.

Note

The detailed steps that are carried out by **update_imgtgene.sh** are:

- Create a working directory for temporary processing files
- Parse the **imgtrefseq.fasta** FASTA file. No check is made to see what's already been loaded, so all the sequences are output to XML format.
- Validate XML file
- Archive XML file and logs
- Import XML and archive import log
- Any Source IDs already loaded will be overwritten. Any Source IDs that are no longer listed by the Data Source will persist in the database. To avoid this you can delete the previous upload before doing the new upload
- Remove the working directory

IMGTLIGM-DB (Large-scale) dataset

This section describes loading or updating the large scale sequence data from **IMGTLIGM-DB** into abYsis.

It is assumed that you have completed the abYsis installation and have the database running.

Obtaining the data

You will need the `imgt.dat` file downloaded to your computer.

IMG T data are available from locations such as the EBI, where files can be found in the following directory at the time of writing:

```
https://ftp.ebi.ac.uk/pub/databases/imgt/
```

Germline dataset:

```
https://ftp.ebi.ac.uk/pub/databases/imgt/ligm/imgt.dat.Z
```

Loading the data

To parse and load the IMG T data into abYsis, change to the `manage_data` directory and run the update command for IMG T/LIG M-DB data

```
cd /home/abysis/abysis_home/manage_data
```

```
./update_imgt.sh /tmp/imgt.dat &
```

Due to the large number of sequences in `imgt.dat.z`, the script will take about a day to run for the first import of this data. Subsequent updates will only import *new* sequences and should be faster.

Note: If you wish to delete any existing IMG T upload before running this script, see the section **Delete all Source IDs for a given Data Source** below.

When the script completes, and if you are finished loading **all** your data (i.e. not planning to load other databases described in this section), you will need to refresh the BLAST databases and site statistics. Go to the *website directory* and run the build script.

```
cd /home/abysis/abysis_home/www/abysis
```

```
./build.sh
```


Note

The **update_imgt.sh** script uses the abYsis EMBL-format parser. The detailed steps carried out by **update_imgt.sh** are:

- Create a working directory for temporary processing files
- Source IDs previously loaded in the database are skipped if the database is connected (to run without checking previous uploaded IDs, set the **\$UDB** flag to 0 in **manage_data/common.sh**)
- Compute and add standard annotations (numbering, canonicals, regions assignment, protein/DNA alignments, human subgroup, etc.).
- Pair chains if the Source IDs are linked via IMGT/LIGM-DB:partner references both chains are classified (not necessarily numbered) one heavy, one light
- Validate final XML file
- Archive XML file and logs
- Import XML and archive import log
- Any Source IDs already loaded will be overwritten. Any Source IDs that are no longer listed by the Data Source will persist in the database. To avoid this you can delete the previous upload before doing the new upload
- Remove the working directory

OGRDB (Germline) dataset

This section describes loading **OGRDB germline** sequence data in to abYsis.

It is assumed that you have completed the abYsis installation and have the database running.

Obtaining the data

You will need the OGRDB germline files in json format downloaded to your server. They can be found at

https://ogrdb.airr-community.org/germline_sets/Human

https://ogrdb.airr-community.org/germline_sets/Mouse

For each you want to access the AIRR (JSON) files from the appropriate links either through a standard web browser or a Linux terminal command such as curl. It is suggested that you create a directory in the /tmp directory for these files to simplify access for the following loading step.

```
mkdir /tmp/ogrdb
```

```
cd /tmp/ogrdb
```

Loading the data

To parse and load the OGRDB germline sequence data change to the **manage_data** directory and run the update command for OGRDB germline data.

```
cd /home/abysis/abysis_home/manage_data
```

```
./update_ogrdb.sh /tmp/ogrdb &
```

The script will take several minutes to run.

Note: If you wish to delete any existing OGRDB germline upload before running this script, see the section **Delete all Source IDs for a given Data Source**.

When the script completes, and if you are finished loading all your data (i.e. not planning to load other databases described in this section), you should refresh the germline and BLAST databases and site statistics. To do this, go to the *website directory* and run the build script.

```
cd /home/abysis/abysis_home/www/abysis
```

```
./build.sh
```

Do not do this step if you intend to immediately upload another dataset, such as IMGT/GENE-DB. This step should be done after all updates have occurred.

Note

The detailed steps that are carried out by **update_ogrdb.sh** are:

- Create a working directory for temporary processing files
- Parse the json files found in the provided directory. No check is made to see what's already been loaded, so all the sequences are output to XML format.
- Validate XML file
- Archive XML file and logs
- Import XML and archive import log
- Any Source IDs already loaded will be overwritten. Any Source IDs that are no longer listed by the Data Source will persist in the database. To avoid this you can delete the previous upload before doing the new upload
- Remove the working directory

Deleting IDs for a given Data Source

In certain circumstances you may want to reload a given data source. For example it might have been updated by the provider (most commonly EMBL or IMGT).

To ensure that you have no stale accessions, data or both retained and not overwritten during the reload, you can manually delete ALL the Source IDs for a data source type with the following command:

```
cd /home/abysis/abysis_home/manage_data  
./delete.sh 'IMG/GENE-DB'
```

Deleting all the Source IDs for a data source may take up to 10 minutes.

Post-translation Modifications

This section explains how to customise the annotation of sequences with site information such as post-translational modification sensitive sites and default cutoffs in the humanization tab heatmap.

Annotations can be provided based on regular expression matches against the sequence.

Configuration file

The sequence annotation configuration file is found at:

```
/home/abysis/abysis_home/static_data/CONFIG_FILES/ptms/ptms.yaml
```

This is a normal text file that can be edited using any text editor. You will see that the file is divided into a number of entries.

The top element is a unique id number for the given annotation. It is an integer followed by a colon. This is followed by entries as follows:

1. **description:** This is followed by a text description for the annotation which is displayed when you hover over the icon on the annotated sequence.
2. **short:** This is a short text label for the annotation. It must not contain any spaces and is used internally
3. **icon:** This defines a font-awesome icon to be displayed on the sequence where the site is found. See <http://fontawesome.github.io/Font-Awesome/icons/> for a complete list of icons.
4. **annotator:** Currently this should always be set to 'Regex' to indicate that Perl regular expressions are used to indicate an annotation site. In a future version of abYsis this will allow an external program to be called
5. **regex:** This is a Perl regular expression to indicate the sequence pattern being annotated. The site annotated is indicated with a `()` capture group. There should only be one of these in a given Regex. In order to capture overlapping patterns the non site part of the regex should be wrapped in lookahead and lookback modifiers: `(?= pattern)` and `(?<= pattern)` This ensures that only the site is consumed on matching. For example, if you want to match the pattern `'([ST]) . [RK]'` – i.e. **S** or **T** followed by **X** followed by **R** or **K**, then the sequence

'STRK' should match twice, at both **S** and **T**. This is achieved by modifying the regular expression to '`([ST]) (?= . [RK])`'

6. `onlysurface`: This is a Boolean indicator (`1=TRUE`, `0=FALSE`). If set then the annotation will only be displayed for residues that are, on average, on the protein surface.

Displaying the new annotations

The new annotations will be displayed immediately for sequences that are entered via the **Import** tab or the **Annotate** tab on the abYsis frontend.

However, the annotations on sequences already in the database are pre-calculated and will not be displayed until they are updated. To do this go to the `manage_data` directory and run the command for updating PTMs.

```
cd /home/abysis/abysis_home/manage_data
./update_database_PTMs.sh
```

Notes

You will need to ensure that the **\$UDB** flag is set to 1 in `manage_data/common.sh` in order to update existing entries rather than delete them.

Appendix. Installation Error Codes

0001 The directory 'abysis_install' has not been created in the abysis installation directory. To create this directory manually do the following;

```
mkdir /home/abysis
```

change access rights to 777

```
chmod 777 /home/abysis
```

Re-run the installation script

```
./install_abysis.sh &
```

0002 /usr/bin has not been found in the path. This is required for standard OS commands. Please add /usr/bin to the path. You can then re-run the installation script.

0003 /bin has not been found in the path. This is required for standard OS commands. Please add /bin to the path. You can then re-run the installation script.

0004 /sbin has not been found in the path. This is required for standard OS commands. Please add /sbin to the path. You can then re-run the installation script.

0005 The abysis home directory (e.g. /home/abysis) does not exist. Create the directory

```
mkdir /home/abysis
```

and make abysis the owner:

```
chown abysis:abysis /home/abysis
```

You can then re-run the installation script.

0006 The abysis user does not exist. Create the abysis user and ensure that the abysis user has ownership of the abysis user's home directory (see 0005). You can then re-run

the installation script.

0007 Access rights on the abysis user's home directory (e.g. `/home/abysis`) are incorrect. The abysis user must have read and execute rights for home. To do this, type (e.g.):

```
chmod a+rx /home/abysis
```

You can now re-run the installation script.

0009 Access rights on the abysis user's home directory (e.g. `/home/abysis`) are incorrect. Other users must have read and execute rights for the home directory. To do this, type (e.g.):

```
chmod a+rx /home/abysis
```

You can now re-run the installation script.

0010 You must run the script from the abysis user's home directory. Change directory to the home directory and re-run the installation script.

0011 You must install the EPEL Yum repository before proceeding. See instructions at: https://fedoraproject.org/wiki/EPEL/FAQ#How_can_I_install_the_packages_from_the_EPEL_software_repository.3F

0013 The appropriate operating system has not been found. abysis requires Rocky version 9 or RedHat version 9. Halt installation and install the correct operating system before continuing.

0017 SELINUX has not been disabled. SELINUX must be disabled prior to installation. Ensure that, in the file `/etc/selinux/config`, SELINUX is assigned as follows:

```
SELINUX=permissive
```

or

```
SELINUX=disabled
```

Use a text editor to edit the file accordingly. Ensure that the file has been updated using

```
cat /etc/selinux/config
```

You can now re-run the installation script.

0024 The httpd configuration file has not been copied to install directory. Issue the command

```
cp /etc/httpd/conf/httpd.conf ~/InstallScripts/install_abysis_4.x.x/httpd.bak
```

to save the configuration file. Check that the installation directory contains the file httpd.bak. You can now re-run the installation script.