Submitting read data

Our public documentation is available in: http://www.ebi.ac.uk/ena/submit/read-submission

For reporting data into the data hub, the following needs to happen:

[1] Data providers should register a submission account (Webin-NNNN) or login into an existing account here: https://www.ebi.ac.uk/ena/submit/sra/#registration

Once logged in, go to the 'New Submission' tab and click on "launch uploader". This opens a file uploader which you can use for your read submission at a later stage.

First continue with the next steps:

[2] Data providers should register projects/studies: On the 'New Submission' tab, select 'submit sequence reads and experiments' and click "next".

Either select an existing study or click on "create new study" and complete the details of your study to which the reads will be linked. Your study will appear in the list. Select it and click "next".

[3] Register samples: To register samples, a reporting standard needs to be selected. For purposes of this project, select "other checklists" and "ENA default sample checklist" and click "next".
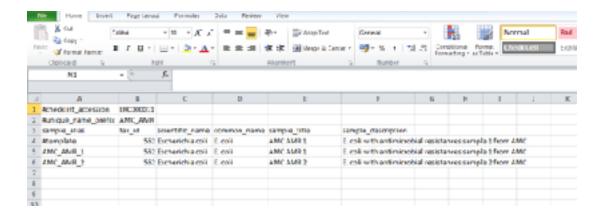
There are 2 ways to create samples. We recommend using the 2nd option.

Option 1:
Fill out the mandatory fields only and click "next". You can now indicate how many samples you want to create according to how many read files you wish to submit. Create 1 sample per read file. Click "add" and "next".

Option 2:
Click on "Download template". The file is a tab-delimited text file (.tsv). Open excel; go to "file" then to "open"; select the data template file; it will automatically import and convert to excel format by clicking "next" and "finish". Here is an example of what that should look like:

Make sure there are no extra spaces behind words and if needed, use dots – not commas for decimals. To upload the completed file, first go to "save as" in the excel menu and select "save as tab-delimited text file". This will automatically create the correct format (.txt). Please note that the taxonomic ID for Escherichia coli is 562. After conversion your file should look like:
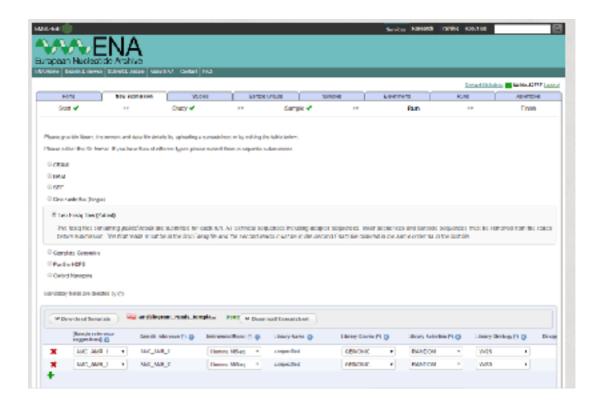

On your ENA/Webin account – go back 1 page by clicking "previous" and click "upload spreadsheet" to upload your ENA default sample checklist. Now click "next". The next page should now be completed automatically. Click "next" or "skip" to go to the final "run" page.

[4] Upload and submit NGS data: A file 'Uploader' is embedded into the interactive submission portal which you opened at the start of your submission process. However, there are other ways of uploading data via FTP or Aspera that are described here: https://www.ebi.ac.uk/ena/submit/uploading-data-files.

To use the embedded file uploader, you may need to install or update "Java Start App".
You will then access the Webin Java web start EBI uploader. Enter your Webin identifier ("username") and Webin password, then connect. Browse to the "Upload Directory" containing your files in gzip or fastq format. Select the files you want to upload. Click on "Upload" to start the transfer.

Having uploaded your reads, return to the "run" page:

Here you will need to provide library, instrument and data file details by uploading the spreadsheet (click "download template") or by editing a table in the interactive interface. We recommend using the spreadsheet as follows:

- Sample_alias: should be identical to the sample alias provided in the previous step
- Instrument model: the different instrument models are listed in the drop-down menu in the "run" webpage.
- Library source: the different sources of DNA are listed in the drop-down menu in the "run" webpage.
- Library selection: the different possible selection methods for library preparation are listed in the drop-down menu in the "run" webpage.
- Library strategy: the different possible library strategies are listed in the drop-down menu in the "run" webpage. For this project it should be "WGS".
- Library construction protocol: "SINGLE" or "PAIRED".
- Insert size: size of your library insert in bp.
- First file name: name of your forward fastq or gzip file. The name should correspond exactly to the name of the file uploaded via the Java web start uploader or FTP software.
- Second file name: name of your reverse fastq or gzip file.

- If a FTP or Aspera software is used to upload the files, an md5 checksum file must be provided for every file. This file is generated automatically when using the Java web start uploader.

Leave any additional data fields blank.

[5] MD5 Checksums
Note: If you are using FTP or Aspera for data upload, you would need to add MD5 checksums into the corresponding field. This is to check that data transfers have been completed successfully and/or the correct file has been uploaded. MD5 checksum is like a compact digital fingerprint of a file. Therefore, we need to know the checksum of the file as it is being transferred, i.e. if you are compressing a file for upload (as with fastq), then please provide the checksum for the file AFTER it has been compressed. For details on how to calculate MD5 checksums please see here: https://www.ebi.ac.uk/ena/submit/uploading-data-files#md5

Once all fields are completed, select the 'Submit' button.

After submission has been completed and as soon as data has been processed and archived all data will appear in the data hub.

Note: You can check the status of your submitted files in your Webin-NNNN account in the 'Run' tab. The data can only be accessed in a data hub if the status shows 'archived'. Note that processing and archiving can take a few days after successful submission, depending on data volume being processed at the same time.

Please email to datasub@ebi.ac.uk in case you had any issues with the submission and quote 'COMPARE' in the subject line.