Submitting read data

**Our public documentation is available in:
https://ena-docs.readthedocs.io/en/latest/reads.html**

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

[2] Data providers should register projects/studies: Once logged in, go to the 'New Submission' tab and register one or multiple sequencing studies, using the 'Register study' option.

[3] Register samples: At the same time of project registration, or at a later stage, samples that are being sequenced should be registered. To register samples, a reporting standard needs to be selected. The choice of a standard depends on sample provenance, i.e. the type of sample being described. A list of available sample reporting standards can be found here:
http://www.ebi.ac.uk/ena/submit/checklists

Note 1: By clicking on the checklist ID (ERC0000XX) you can browse pages where the standard requirements are described, i.e. which fields are mandatory, recommended or optional and whether or not allowed values are controlled by a list, free-text or any other format.

Examples of sample reporting checklists developed with the COMAPRE framework are:
Influenza virus : http://www.ebi.ac.uk/ena/data/view/ERC000032
Virus pathogen: http://www.ebi.ac.uk/ena/data/view/ERC000033
Sewage: http://www.ebi.ac.uk/ena/data/view/ERC000036
Parasites: https://www.ebi.ac.uk/ena/data/view/ERC000039

In addition there is a GMI (Global Microbial Identifier) sample reporting checklist that should be used for reporting prokaryotic pathogens:
http://www.ebi.ac.uk/ena/data/view/ERC000029

Note 2: For most pilot projects, where sequencing methods, platforms and analysis tools are being compared by a number of partner institutes, usually the same sample(s) are being sequenced. In this case only the institute(s) that is/are circulating the sample to others should register the sample(s). Please do not replicate the same physical sample by registering it multiple times. At the time of data submission (see under [4]) sequencing experiments can be pointed to any samples registered by yourself or by another institute.

[4] Upload and submit NGS data: Log into your account, go to the 'New Submission' tab and select the option 'Submit sequence reads and experiments'. The first step of your submission is to upload data files. You can use the '[Webin file Uploader](#)', linked from the interactive submission portal. However, there are other ways of uploading data via FTP or Aspera that are described here: https://ena-docs.readthedocs.io/en/latest/fileprep/upload.html

If you are doing the data submission at a later stage than the project registration, the next step of your data submission would be to select the study you have registered earlier. If you have already registered samples too, you can skip this step now and you will be forwarded to a step to describe the files you have uploaded. Here you will need to provide library, instrument and data file details by uploading a spreadsheet or by editing a table in the interactive interface. It is here where you also point to already registered samples. If the samples have been registered as part of the same submission session you can select them from the [Sample reference suggestions] field. If you have registered your samples earlier or you want to associate your sequences with samples registered by another institute (see Step [3] Note 2), you can add them to the 'Sample reference' field or to the spreadsheet you upload.

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 MD5 checksums please see here:
[https://ena-docs.readthedocs.io/en/latest/fileprep/preparation.html#registering-the-md5-checksum-with-ena](#)
https://ena-docs.readthedocs.io/en/latest/faq/runs.html#wrong-md5-value-registered-register-a-new-one

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 by using the 'Webin submission portal: [https://www.ebi.ac.uk/ena/submit/webin/](#), logging into your Webin-NNNN account and going to the 'Run files' tab. The data can only be accessed in a data hub if the 'Active status' shows 'File 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 [datasub@ebi.ac.uk](#) if you have any issues with the submission and quote 'COMPARE' in the subject line.