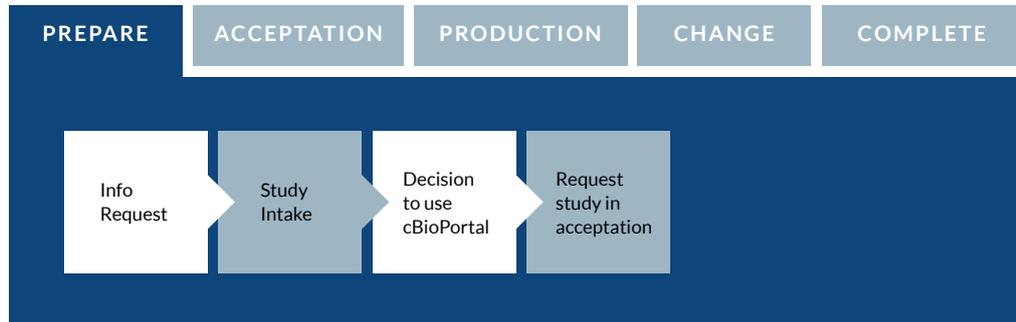


STEP-BY-STEP GUIDE
cBioPortal

V1.0 | 26 may 2020



PREPARATION



INFORMATION REQUEST

If you want to know more about whether and how you can use cBioPortal for your research data, please contact the *Health-RI Servicedesk*. If you do not yet have a Health-RI self-service portal account, an account will be created for you, and your information request will be processed accordingly. You will be contacted by a member of the cBioPortal data team, who will gladly answer any questions you may have.

You can view demonstration studies in cBioPortal by visiting <https://www.cbioportal.org/>. For a more in-depth impression of cBioPortal, its application and uses, a demo can be given upon request.

STUDY INTAKE

Following the information request step, a member of the cBioPortal data team will discuss with you if and in which manner, your data types can be imported and modelled in the cBioPortal data-integration platform in order to view, query and analyze the available data.

It will also be established which types of queries and analyses can be performed within this data-integration platform, and which may need to be performed in another environment (more specialized queries). During this talk it will be discussed with you which steps need to be performed by you as a data owner, and which steps will be performed by the Data Team.

In the *“Data preparation”* step, you can already read about how you should prepare the data and what the requirements are.

DECISION TO USE CBIOPORTAL

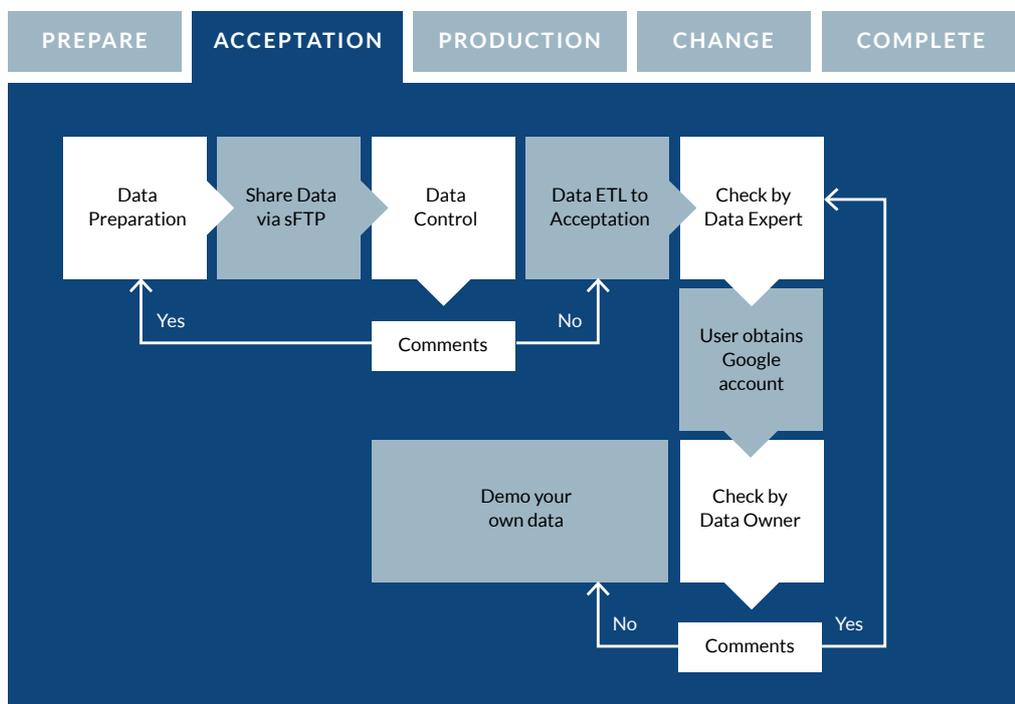
Following the information request and study intake, you are free to decide whether or not to make use of cBioPortal. If you do decide to use cBioPortal, you can request a for a new study to be made in the Acceptation environment of cBioPortal, as

described in the “Request study in acceptance”.

REQUEST STUDY IN ACCEPTATION

Once you have decided that you would like to make use of cBioPortal for your data, you can login to the Health-RI *Self-Service Portal*, select ‘Services and Request Forms’ and go to ‘cBioPortal’ and then enter ‘cBioPortal new study request’.

ACCEPTATION



DATA PREPARATION

During the initial preparation phase, you have been in contact with the cBioPortal data team to get a first impression of in cBioPortal and the data you would like to store (see ‘study intake’). Here you can find information on how you should prepare your data for cBioPortal.

Please read the information provided in the Data Requirements Document: <https://docs.cbioportal.org/5.1-data-loading/data-loading/file-formats>. Click on the data types that are applicable to you (e.g. segmented data or methylation data). Prepare a tsv file according to the instructions from this link.

If you want to see examples, you can find demonstration studies by visiting: <https://www.cbioportal.org/>. The data preparation files for these studies can be found here: <https://github.com/cBioPortal/datahub/tree/master/public>.

As an example, in the MSK-IMPACCT study: http://www.cBioPortal.org/study?id=msk_impact_2017, clinical, CNA and mutation data has been uploaded for 10.000 patients.

In here, https://github.com/cBioPortal/datahub/tree/master/public/msk_impact_2017, you can investigate for yourself how the data files were prepared for the upload, based on the requirements set by the Data Requirements Document.

SHARE DATA VIA SFTP

Once the data has been prepared, the data should be transferred in a safe and secure manner using the TraIT 'secure File Transfer Protocol' (sFTP). An account for the sFTP, in case you do not have one yet, can be requested via the Health-RI *Self-Service Portal*.

To request an sFTP account for file transfer, login into the Health-RI *Self-Service Portal*, select 'Services and Request Forms' and scroll down to 'cBioPortal' and click on 'request sFTP account'. After the request has been made, you will receive login details and a manual explaining how to set up and use an sFTP client like FileZilla or Cyberduck to transfer the files directly to the safe environment of the data team. Please note that if you have received an sFTP account for any other Health-RI tool (e.g. tranSMART) in the past already, that the same account can be used for this step.

After uploading the data, you must update the status of your initial cBioPortal study request in order to inform the data team that the files have been uploaded.

DATA CONTROL

After the Data Team receives a notification that data has been delivered to the safe environment of the, they will make an internal working directory for your study on the secure ETL server (Extract, Transform and Load). In this working directory, initial data checks are performed to make sure the data is ready to be loaded to cBioPortal.

The data checks include investigating whether the data has been delivered correctly according to the Data Requirements Document (see "*Data preparation*").

COMMENTS

In case errors are found in the data, feedback will be provided. Small issues which do not require significant work will be taken care of by the data team. Errors such as missing required fields are errors that the data team often cannot solve by themselves. In these cases, you – the data owner – will have to correct the data or the files with the feedback provided by the data team. The feedback will be accurate enough for you to spot and fix the problem.

If the data is in good shape and the data forms seem complete, then the data team will send out an update that the data was received in good order, and that the next steps can now be taken.

DATA ETL TO ACCEPTATION

The data provided has been delivered correctly according to the Data Requirements Document (see “*Data preparation*”). Next, the data will be processed by the data team on the ETL server (Extract, Transform, Load), after which it will be uploaded to the cBioPortal acceptance server. This is a stable and secure cBioPortal environment with the purpose of modelling and iterating your data.

CHECK BY DATA EXPERT

After the data has been loaded to the cBioPortal acceptance server, the initial load result is checked by the data loader. In case anything went wrong with the upload, it can be corrected. If everything went correctly, the Data Owner will be notified.

USER OBTAINS GOOGLE ACCOUNT

In order to login in to cBioPortal, the user is required to have a Google account. In case the user doesn't have a Google account yet, they can create one at accounts.google.com/signup.

After the data owner has received a notification that the study is ready in acceptance, access can be granted to other users. Access to the study on acceptance is arranged via the Health-RI *Self-Service Portal*. A request for access can be done by selecting 'Services and Request Forms', scrolling down to 'cBioPortal', and then clicking on 'cBioPortal access to existing Study'. Fill out the form, make sure you select the correct study name, the correct environment (acceptation) and the correct role.

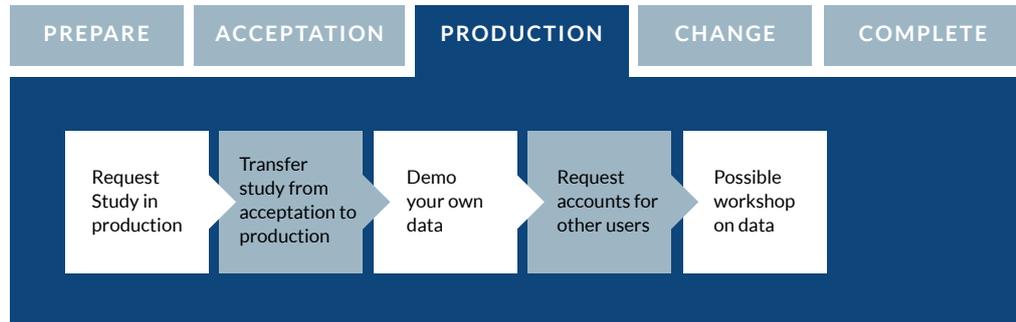
CHECK BY DATA OWNER

After the data has been loaded and checked, the data owner is given access to the study in the acceptance environment for an initial inspection of their own data in cBioPortal. Any comments or adjustments that fit within the cBioPortal model will be picked up by the data team to improve the representation of your data in cBioPortal.

COMMENTS

Any feedback on the study as it is currently displayed in cBioPortal? No problem; the data team is glad to help you incorporate the feedback into the data loading process. Note that extensive changes to the data mapping or to the source data might also require you as data owner to adjust the provided data.

PRODUCTION



REQUEST STUDY IN PRODUCTION

After the study has been modelled in the acceptance environment and you are content with the representation in cBioPortal, the next step is to move the study to the production server. The production server is a stable environment where versions of the data are loaded to use in your research.

To move your study from the acceptance to the production server, a formal request has to be made to the data team. The request is made by logging into the Health-RI *Self-Service Portal*, select 'Services and Request Forms' and go to 'cBioPortal' in order to enter a 'cBioPortal new study request'. This time, in the form, the request can be made to load your data to the production environment of cBioPortal.

TRANSFER STUDY FROM ACCEPTATION TO PRODUCTION

After the data team receives the request to load the current version on acceptance to production, the data will be loaded to the production server. After the data loading, the data loader checks whether the data was loaded correctly. When the study is available in production, access will be given to the PI/delegate and a notification will be sent.

DONE

You can now make use of your study in cBioPortal. Access for other users can be requested through the Health-RI *Self-Service Portal*. In case you as the data owner, or other users would like to do a quick 30-minutes tutorial on cBioPortal, in which you will learn all the basics, then you can download the Training Handout material that is available upon request.

DEMO YOU OWN DATA (OPTIONAL)

Next, you are free to demo your data in cBioPortal to your peers. If needed, a member of the data team can be present at the demo in order to provide further explanation or clarification if necessary.

CHANGES IN STUDY

It is always possible that data updates become available over time or errors in the provided input data are spotted and adjustments are needed.

In order to make changes or adjustments to an existing study existing in cBioportal, a request can be made by logging in to the Health-RI *Self-Service Portal*. Select 'Services and Request Forms' and go to 'cBioportal' in order to fill in a 'cBioportal update existing study' form. In the remarks please indicate that the study needs adjustments. The data team will get in touch after receiving the request.