Indian Biological Data Centre

Standard Operating Procedure (SOP) For Crop Phenome Data Submission to Indian Crop Phenome Database Version_2.0 2024

ICPD

ICPD: Table of Content

Overview of IBDC

Welcome to the data archive solutions covered by Indian Biological Data Centre (IBDC). This guide will be helpful in understanding the standard operating procedure for the submission of the crop phenotyping data to IBDC. Users are requested to devote a moment towards understanding the structure and mandate of portals developed for dedicated biological data archive before they proceed with the submissions. IBDC offers data submission services for diverse biological data types via its specialized data type-specific portals. Currently the IBDC provides active data archival services for nucleotide, phenome, metabolome and proteome data (Figure 1). IBDC operates nucleotide submissions through Indian Nucleotide Data Archive (INDA) and Indian Nucleotide Data Archive – Controlled Access (INDA-CA). While the crop phenome data can be submitted to other dedicated portal "Indian Crop Phenome Database (ICPD)". The metabolite, image and proteome data can be submitted to Indian Metabolome Data Archive (IMDA), Indian Biological Images Archive (IBIA) and Indian Proteome Databank (IPD) respectively. This guide provides the detailed standard operating procedures for the submission of crop phenotyping data to ICPD.

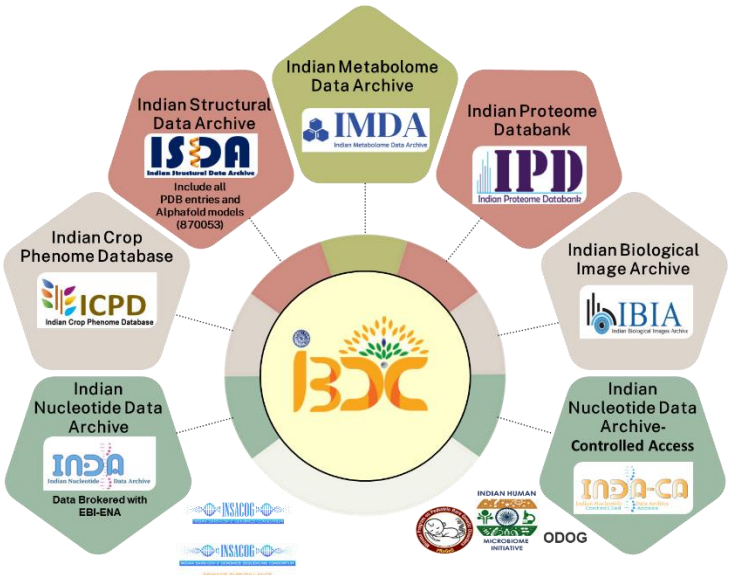


Figure 1. Active portals of IBDC based on different data types.

Introduction to Indian Crop Phenome Database

Indian Crop Phenome Database (ICPD) is a domain of the Indian Biological Data Centre (IBDC), Regional Centre for Biotechnology, Faridabad, INDIA, developed for the digitization of crop phenome data. Being the global agricultural powerhouse, the bulk of biological data generated in INDIA is associated with agricultural trials. Ironically most of the trial data has been inaccessible to other researchers, remains unpublished, and is lost as time passes. Therefore, ICPD would act as a single-stop user-friendly platform for freely archiving, organizing, analysing, and sharing the multi-crop phenome data following FAIR (Findable, Accessible, Interoperable, and Re-usable) data principles. We assign unique and persistent IBDC accessions to data submitted to ICPD. The portal can be accessed at https://ibdc.dbtindia.gov.in/icpd/

Figure 2. Snapshot of home page of ICPD.

The home page provides a comprehensive overview of the total number of submissions and offers user-friendly tools for submitting crop phenome data. Additionally, it includes features that allow users to browse and explore the submitted data, making it easy to access and analyze valuable information. This streamlined interface ensures that users can efficiently contribute to and benefit from the data repository.

Overview of metadata used in ICPD

Submissions made through ICPD are represented using number of metadata objects (Figure 3). A typical crop phenotypic or trial data is associated with a project (Phenoproject) containing several studies based on different crops, developmental stages, tissue, growth conditions and environmental treatments. Further each study has observation data on one to many traits analysed together. Therefore, ICPD data model consist of three main modules i.e. phenoproject, study and data file, which are further defined by several metadata sub-objects to provide detailed and comprehensive definition to each phenotypic data submitted to ICPD. The different modules of ICPD documentations are defined as follows:

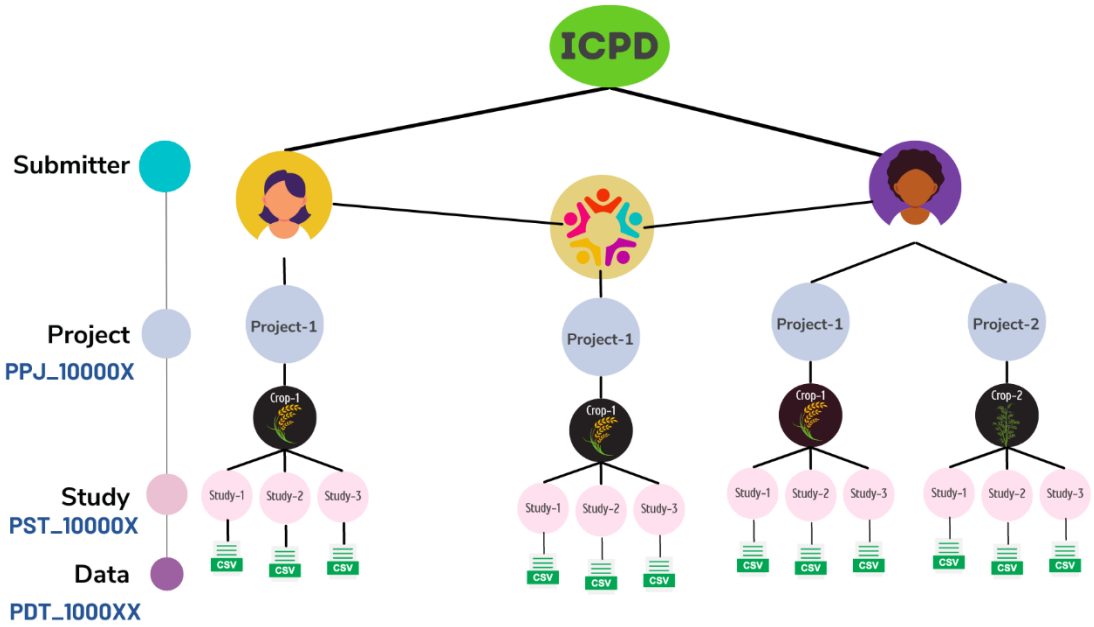


Figure 3. Different meta data objects used in ICPD.

- **Phenoproject:** A phenoproject are basically the research plan with defined aims and objectives. A single phenoproject can have single PI (one institute) to multiple PIs (consortium) belonging to grant funded programmes involving one to many crops. Phenoproject record provides users a single stop to find all the studies and data generated for that phenoproject.
- **Study:** A study comprises a set or series of experiments to understand single to different traits or measurements executed on the trial conducted on particular growth conditions and treatment. It's the sincere duty of the user to provide correct, comprehensive yet detailed description of the different elements (growth, genotypes/cultivars/tolerant or susceptible checks, treatments, traits etc) to define a study.
- **Data:** Data are the observation values recorded for single to multiple traits executed in a particular study.

Data types and modes of data submission

The phenome data submission format is based on the standardized data format used for MIAPPE (www.miappe.org) as well as GnpIS (https://urgi.versailles.inra.fr/Tools/GnpIS) full filling the FAIR data principles. Considering the heterogeneity in the terminology defining trait, tissue, developmental stage and methods, we have integrated the metadata with the standard ontology terms obtained from Planteome (https://planteome.org) and Crop Ontology for agricultural data (https://cropontology.org/). At present phenotyping data are submitted in .csv format along with all the important meta data associated with phenoproject and study. Currently two modes of submissions are offered by ICPD for the digitization of crop phenome data:

- **Web based submissions (Recommended)**
- **Template based submissions**

Although we recommend users to submit via the interactive web-based submissions route which are completed by filling out web-based forms directly in your browser but if required template-based route can also be used. The detailed submission process is explained in upcoming sections.

Getting started on submission

The overview of steps required for the initiate and complete the phenome data submission to ICPD are summarised in Figure 4.

First time users can select "User Registration" link on the top of the ICPD home page to enter the registration page and register an account, which is required for the submission of data to ICPD.

User Registration

User has to enter all the required details in the registration page including credentials, designation, organization, country and Orcid details (figure 5). The email id will be considered as the primary identification of the user, based on which the user will be given a unique user id. The user can set a secure password with format of one number, one uppercase, lowercase and at least 8 or more characters. After successful sign up, the account activation intimation will be sent to the registered Email. After account confirmation by the user, now the account will be accessible to the user.

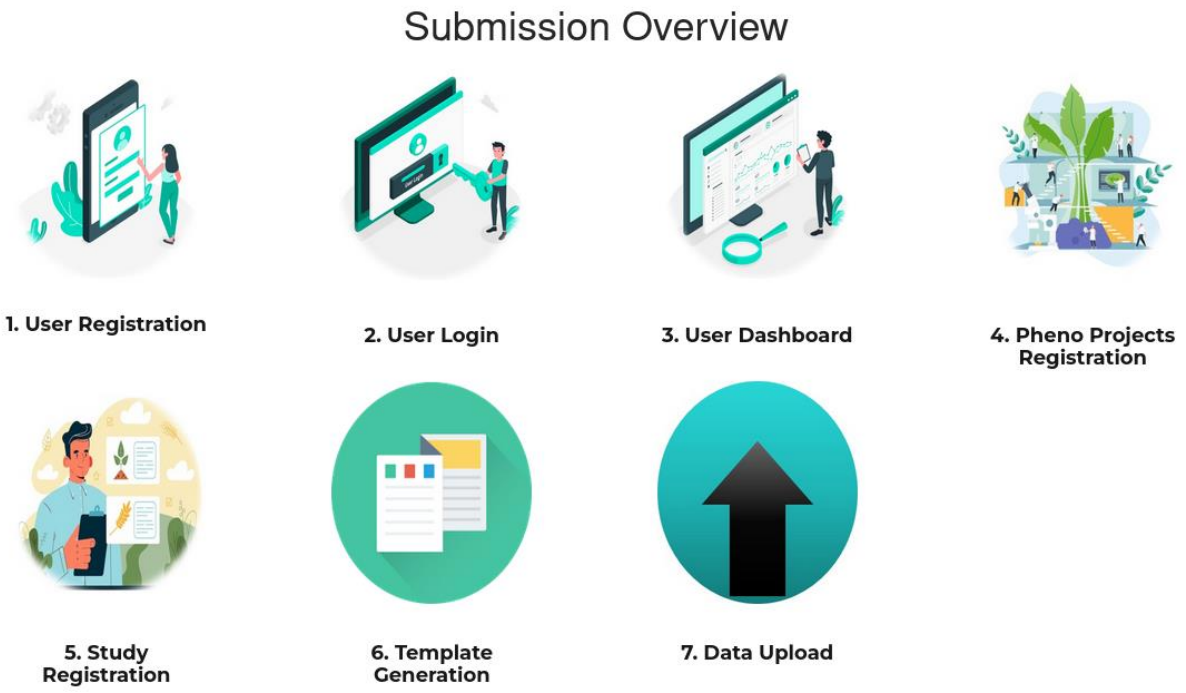


Figure 4. Steps involved in submitting data to ICPD.

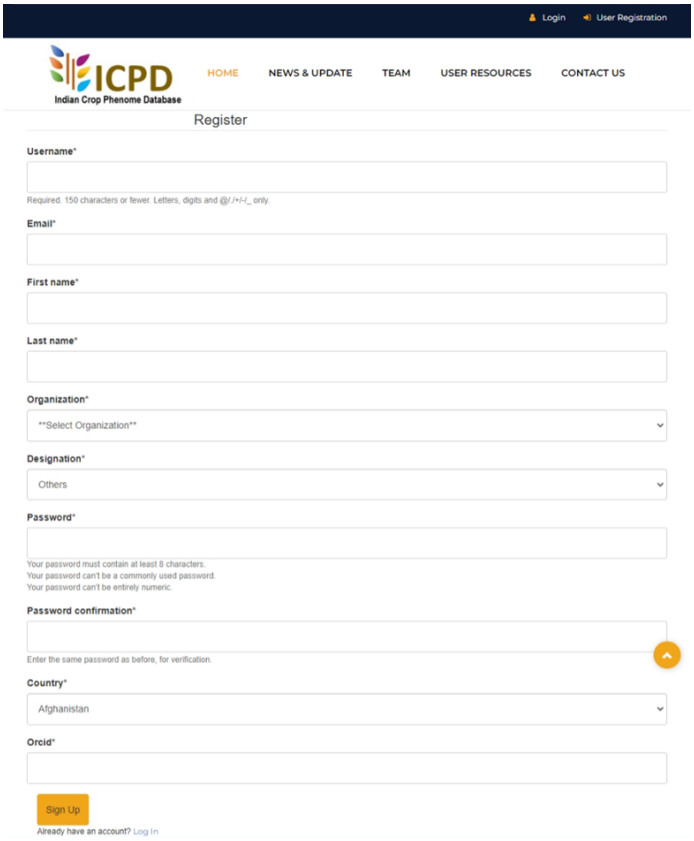


Figure 5. Snapshot of User registration page of ICPD.

User Login

The user can login to the submission portal by entering the login details provided at the time of registration. User has to use the specified username and the password set for the login. The options for forgot password, register now and resend activation link are also provided on the login page (figure 6).

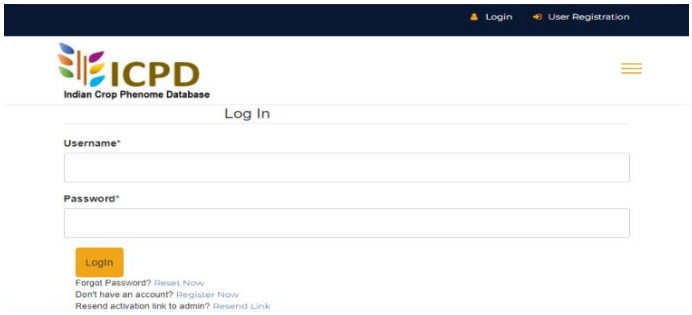


Figure 6. Snapshot of LogIn page of ICPD.

User Dashboard

The user dashboard (figure 7) consists of links to submit different parts of data submission on the left hand side of the dashboard and submission stats on the main part of the page. The dashboard of the user will also display the summary of details of previously submitted phenoproject, study and/or data by the user. User may submit their data by clicking on "Submit New Data" link. On the right top corner, the options related to user name and its profile are shown along with toggle menu and logout button. On the left panel, tools and user resources are listed. Discussion forum allows users to convey their suggestions and queries regarding the submission process. To allow users to browse and contribute to the ontology terms associated with traits, developmental stage, tissue and methods, 'Browse Ontologies' and 'Create Ontologies' options are provided. Then the detailed user manual for guiding users through the submission and ICPD can be downloaded by clicking on the 'Download User Manual' tab. At the bottom, the option for the template-based data submission is available. The upcoming sections of the manual deals with each options in detail.

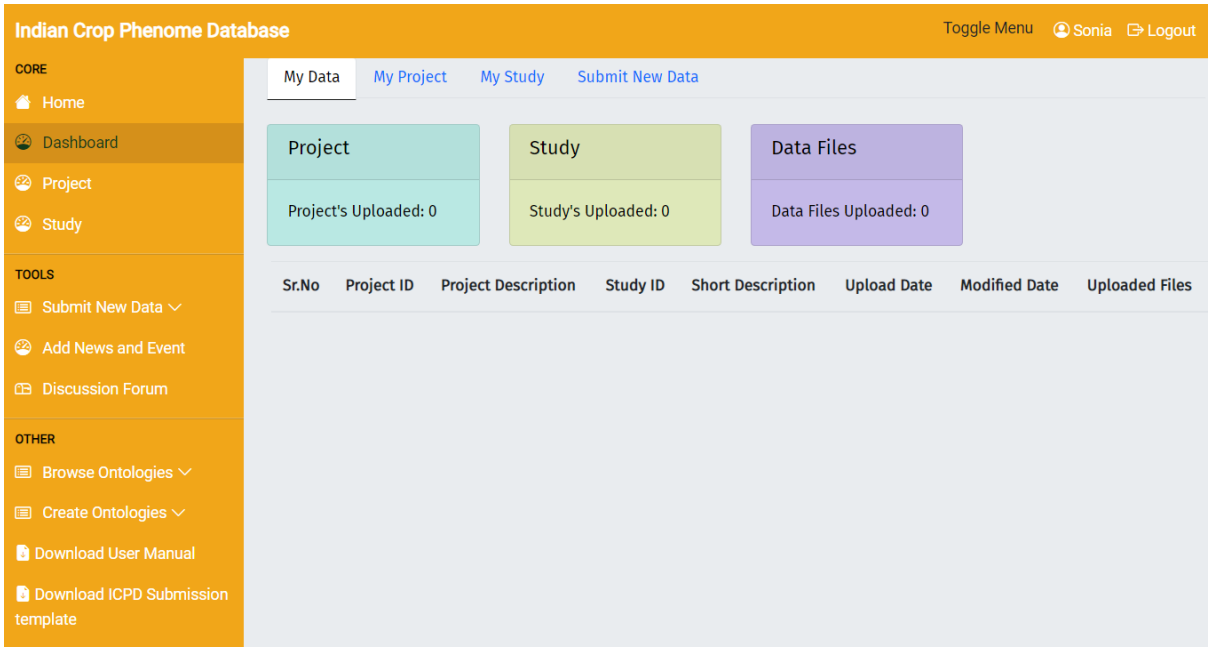


Figure 7. Snapshot of dashboard page of ICPD.

Web-Based Submission

The ICPD recommend users to submit the data through online web-based submission interactive interface which provide several automatic selection fields for recording the metadata of project and study associated with ontology terms. Also allows users to generate pre-filled templates in which only the observations have to be added on user's end. The web-based submission process consists of four steps highlighted in figure 4 ie register new phenoproject, register new study, generate and download template and upload data. Each step will be described comprehensively below:

1. Register New Phenoproject

Click on the 'Submit New Data' button on the user dashboard to initiate the step 1: new project registration. The phenoproject registration form consist of several fields required to comprehensively define a new phenoproject. All the fields marked with asterisk are mandatory and others are optional for the users to provide. This section is divided into project and crop specific form. In the project section, user need to fill out information regarding the title, grant number, funding agency, project description, associated publications, release date and project type. In the project description, users need to provide the abstract or the summary of the project proposed. Its optional to add the details of associated publications by providing the title, DOI or PMIDs. Users are provided with the option to set the data release date for the project and the date should not exceed more than five years from the date of registration. If the phenoproject has only one principle investigator (PI; user itself) then he has to select individual in the 'Project Type' field. Further if a project shares multiple PIs from different or same institution, user should select consortium in the 'Project Type'. As soon as consortium was selected, another section named 'Consortium Details' automatically adds and user need to provide the details regarding the name and affiliations of PIs. In the crop section, the scientific name and taxon_ID of crops analysed for trials in the project needs to be listed. In case the taxon_ID is not known, user can search the taxon_ID in the 'search TaxID' box. The snapshot of the 'Register New Phenoproject' is shown in figure 8. After entering all the sections, click submit button present on the bottom of the form. On successful registration, the user will be automatically directed to the next step and the message confirming the successful registration of the phenoproject along with the unique IBDC accession will be flashed on the top (Figure 8).

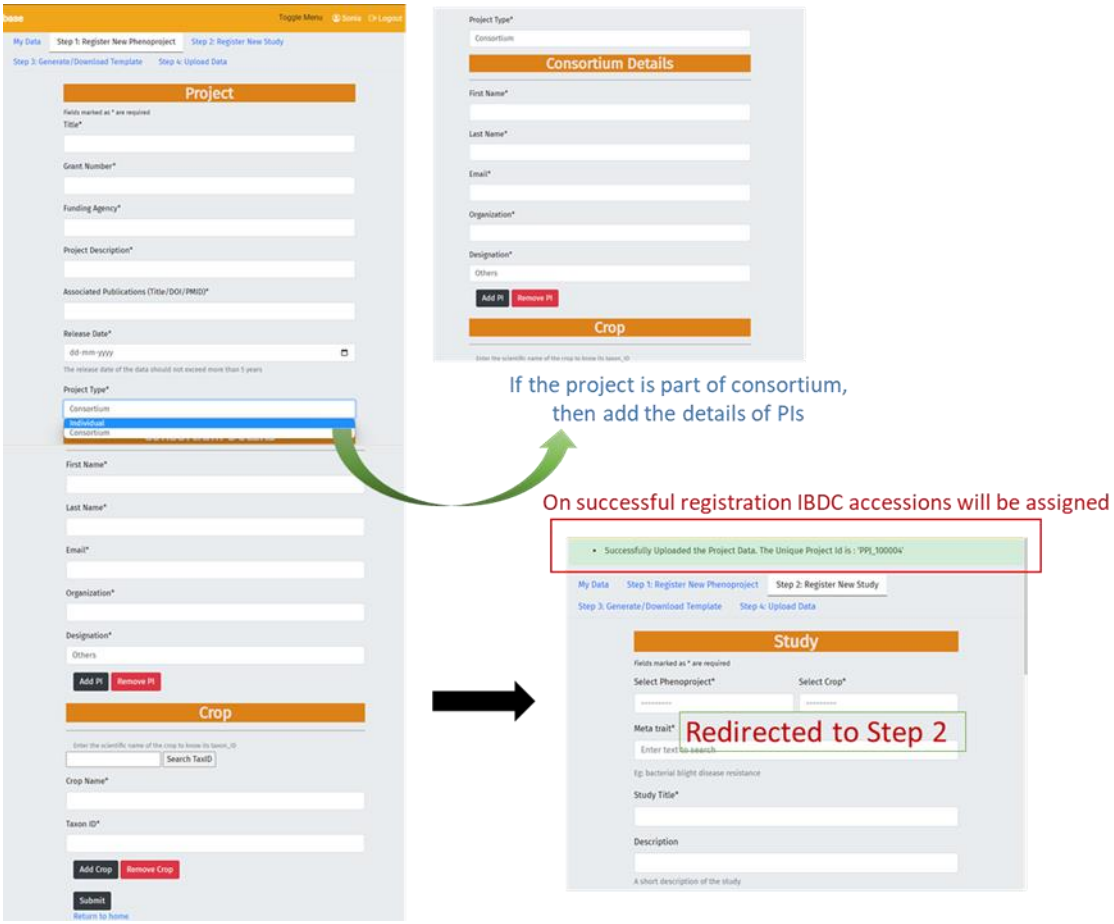


Figure 8. Snapshot of 'Step 1: Register New Phenoproject' page of ICPD.

2. Register New Study

After phenoproject registration, now user can add single to multiple studies in a single phenoproject. Click on the 'Register New Study' button to add a new study to newly registered or already existed phenoproject. For collecting the meta data that define a study, universal format has been developed by incorporating the ontology based documentation of different elements (meta-trait, development stage, plant tissue, traits and methods) of study. The ontology terms were extracted from the planteome (planteome.org) and crop ontology (cropontology.org) databases associated with different crop species. In case the user specified term is not annotated in the ICPD database, he can search the terms in 'Browse Ontology tool' and can create new term using the 'Create Ontology Term' tool present on the left panel of the user dashboard. This section is divided into two sub sections ie step 2: Register new study and step 2.1: study details. In step 2, user need to provide the details regarding the basic information of the study along with growth and environmental conditions. While the detailed meta-data regarding the treatments, traits

and authors will be entered in step 2.1. The list of fields in this section are shown in the figure 9.

In the first section ‘Study’, select the phenoproject_ID and crop from the dropdown. Here all the phenoprojects registered by the user are shown in the dropdown to assign the parent project to the study. After the selection of phenoproject, all the crops that are associated with that phenoproject are shown in the ‘Select Crop’ dropdown. Then user need to specify the meta-trait, study title, study description, its start and end date, location, experimental design, the growth environment and the accessions of tolerant and susceptible checks used in the study.

In the ‘Growth and Environmental Conditions’ section, the details regarding the temperature regimens, light intensity, relative humidity, and any other accessory conditions required to filled. After the complete submission of Step 2.0, the study is assigned with a IBDC study accession.

Under the ‘Treatments’ section of step 2.1, fill out the information regarding the treatment type (biotic/abiotic/no treatment), agent, description, qualifier, duration along with the development stage (type and select) at which the treatment was applied to the plants. In case of combined stress conditions where two treatments were given simultaneously to the plants, users are requested to provide the details by clicking on the ‘Add Simultaneous Treatment’ button.

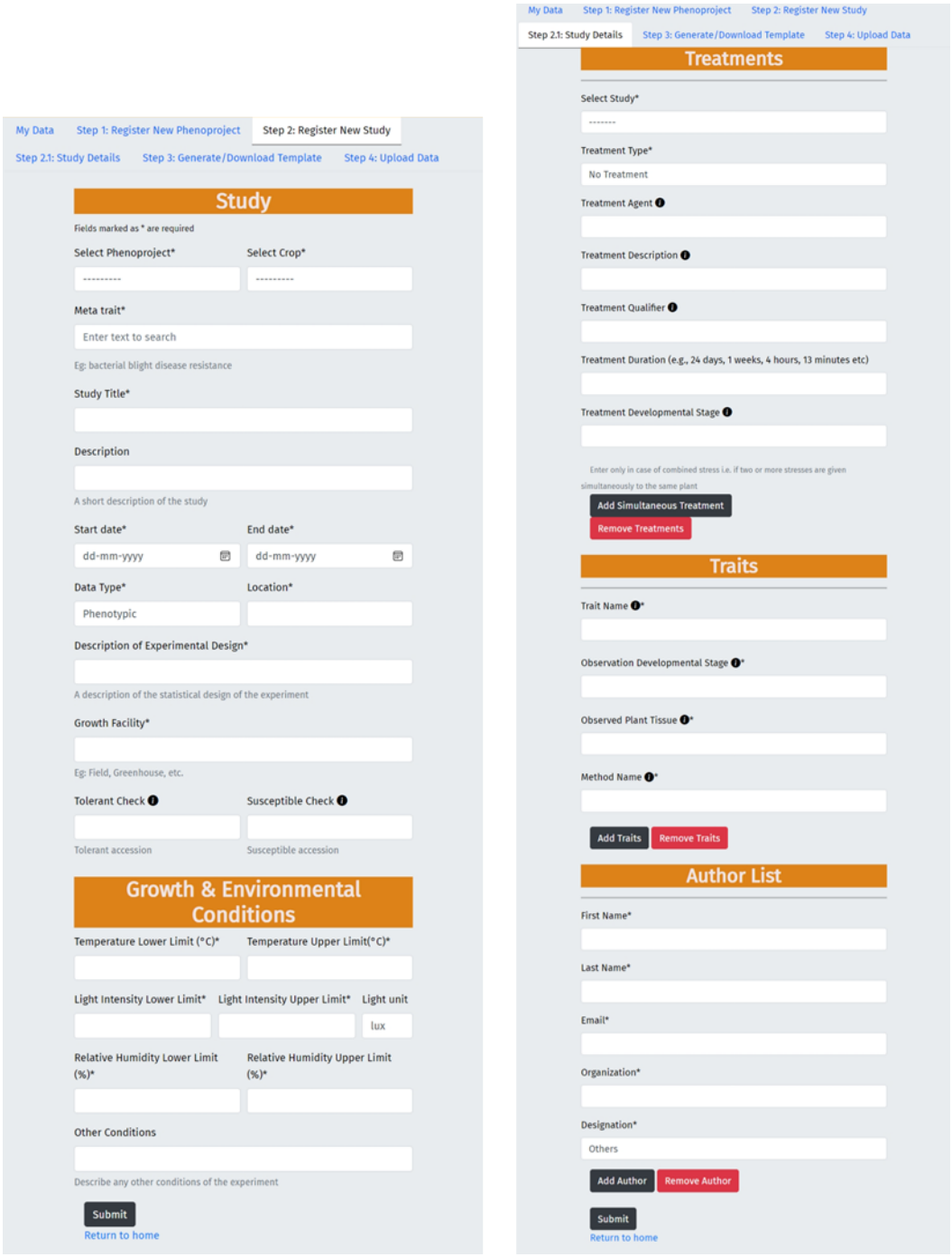


Figure 9. Snapshot of Step 2 of data submission.

Next, all the traits that has been recorded in the study are entered in the trait section by providing selecting the trait name, observation development stage, observed plant tissue and method name. all the above fields are associated with ontology terms and user has to type the terms and select out the relevant one from the options. In case the choice is not annotated here, please open the 'Create Ontology Tool' in the another tab and create the

term of your choice in the respective domain (meta-trait, trait, development stage, tissue and method).

At last provide the details of the authors who contributed in the study in the 'Author List' section and press submit.

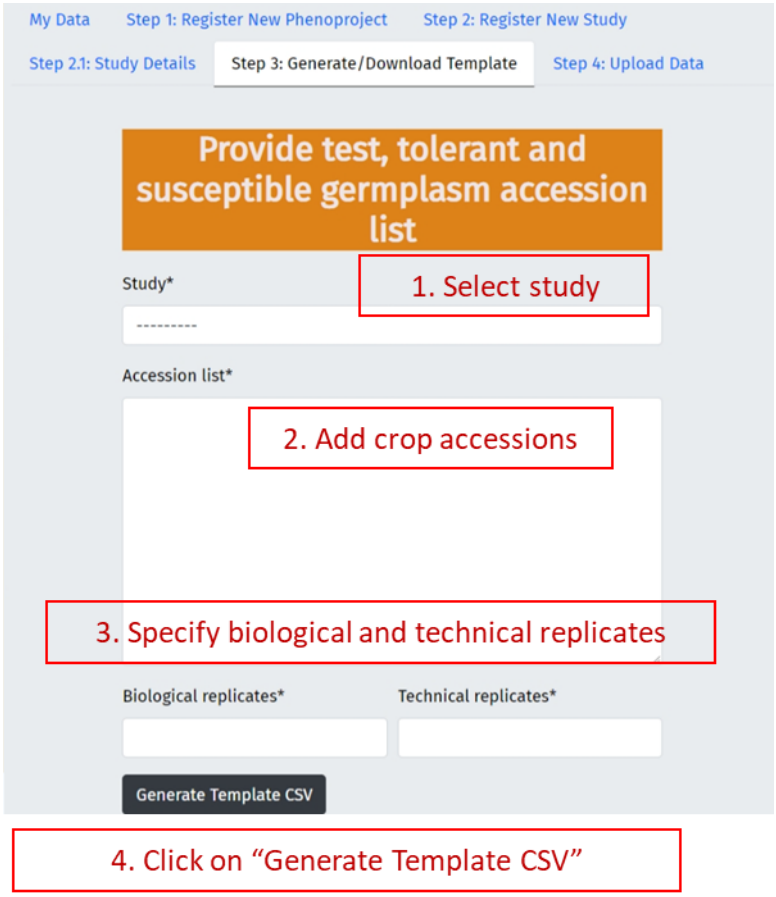
3. Generate and Download template

Once the study has been successfully registered, now user can generate and download the pre-filled data file by clicking on 'Step 3: Generate and Download Template' button on the user dashboard under submit new study (Figure 10). User need to select the study_ID and provide the list of TEST crop accessions (tested for a particular meta-trait) along with the tolerant and susceptible checks. Then specify the number of biological and technical replicates and click on the 'Generate Template CSV' button. Instantaneously, the template file will be downloaded (Figure 10). As shown in figure 10, the first nine columns are pre-filled in the template and depending on the number of biological (B) and technical (T) replicate the observations columns are labelled as B1T1, B1T2 and so on. These pre-filled templates can be used for recording observation during the trials also. User has to fill or paste the observation data into the observation columns carefully and then save the file.

Important Note:

- *Do not edit or change the format of the template file.*
- *The crop accession list should be unique and no accession should be present in the list multiple times.*
- *Do not leave any observation cell blank in the data file. Alternately use NA or '-' instead of missing or blank values in your data.*
- *Kindly ensure the list pasted in the 'Accession list' contain the names of tolerant and susceptible checks also.*

a



b

sr_no	study_Id	project_Id	meta_trait_name	treatment_agent	treatment_qualifier	trait_name	tissue	accession	b1_t1	b1_t2	b1_t3	b2_t1	b2_t2	b2_t3	b3_t1	b3_t2	b3_t3
1	PST_100006	PPJ_100011	drought tolerance	Canopy temperature	diurnal fluctuations	drought tolerance	stamen	A1	12	45	6	77	88	99	7	56	43
2	PST_100006	PPJ_100011	drought tolerance	Canopy temperature	diurnal fluctuations	drought tolerance	stamen	A2	3	9	6	5	12	11	21	23	4
3	PST_100006	PPJ_100011	drought tolerance	Canopy temperature	diurnal fluctuations	drought tolerance	stamen	A3	2	89	43	55	45	12	65	NA	54
4	PST_100006	PPJ_100011	drought tolerance	Canopy temperature	diurnal fluctuations	drought tolerance	stamen	A4	NA	9	56	78	65	45	87	54	43
5	PST_100006	PPJ_100011	drought tolerance	Canopy temperature	diurnal fluctuations	drought tolerance	stamen	A5	6	76	7	98	87	65	12	11	32
6	PST_100006	PPJ_100011	drought tolerance	Canopy temperature	diurnal fluctuations	drought tolerance	stamen	A6	6	6	6	23	98	78	45	34	1

Figure 10. a) Snapshot of Step 3. Generate and Download Template. (b) Snapshot of template file. The red box highlights the observation cells in which user need to enter the values.

4. Upload data

This is the final step in submitting the phenotyping data to ICPD. Click on the 'Step 4: Upload Data' button and select the study in which the data has to be submitted (Figure 11). Browse the filled data file and press upload.All new submissions are updated in the user dashboard instantaneously.

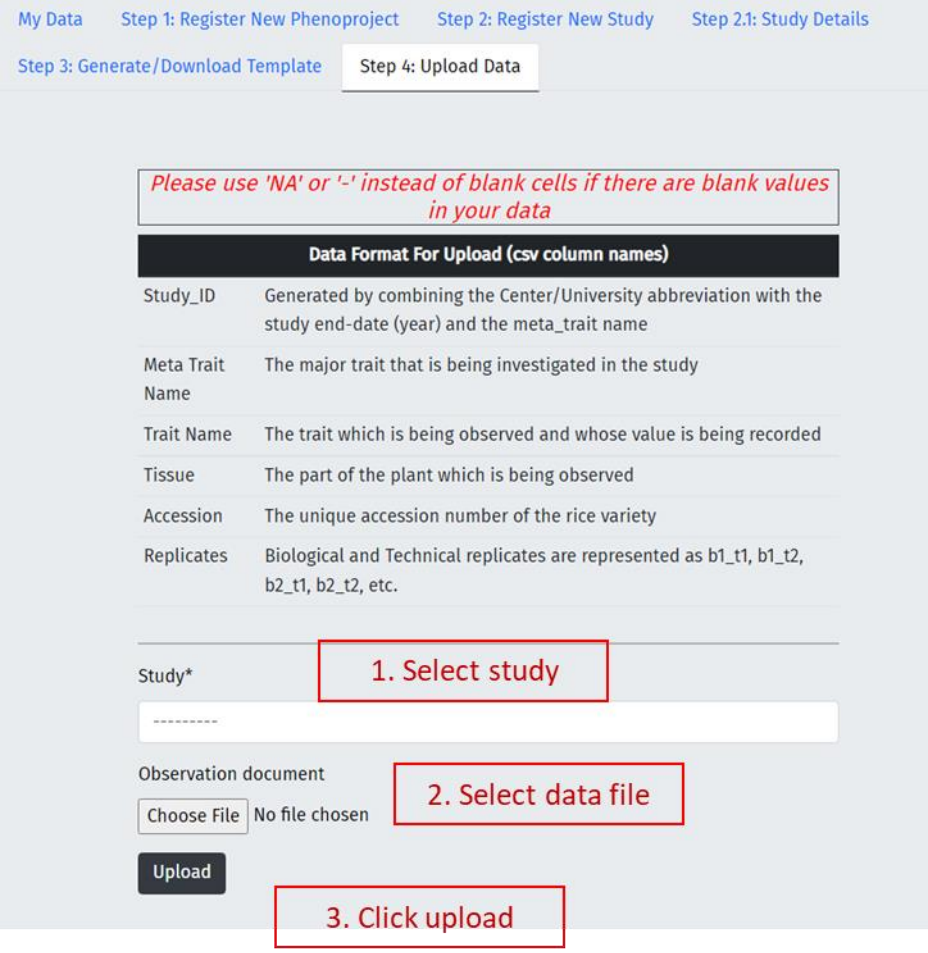


Figure 11. Snapshot of Step 4: Upload Data.

Submission of other supporting data

The functionality is recently updated as Step 5 “Submission of Supporting Data” to allow users to upload the supporting data files including the plant passport data. For uploading the plant passport details of the genotypes screened, first download the plant passport data submission template and select the study in which data has to be submitted (Figure12). Enter the details and upload it to the portal. Users can also submit the other additional information related to the study (eg soil or weather data) and upload its file by clicking the upload option and finally clicking the submit button. All new submissions are updated in the user dashboard instantaneously.

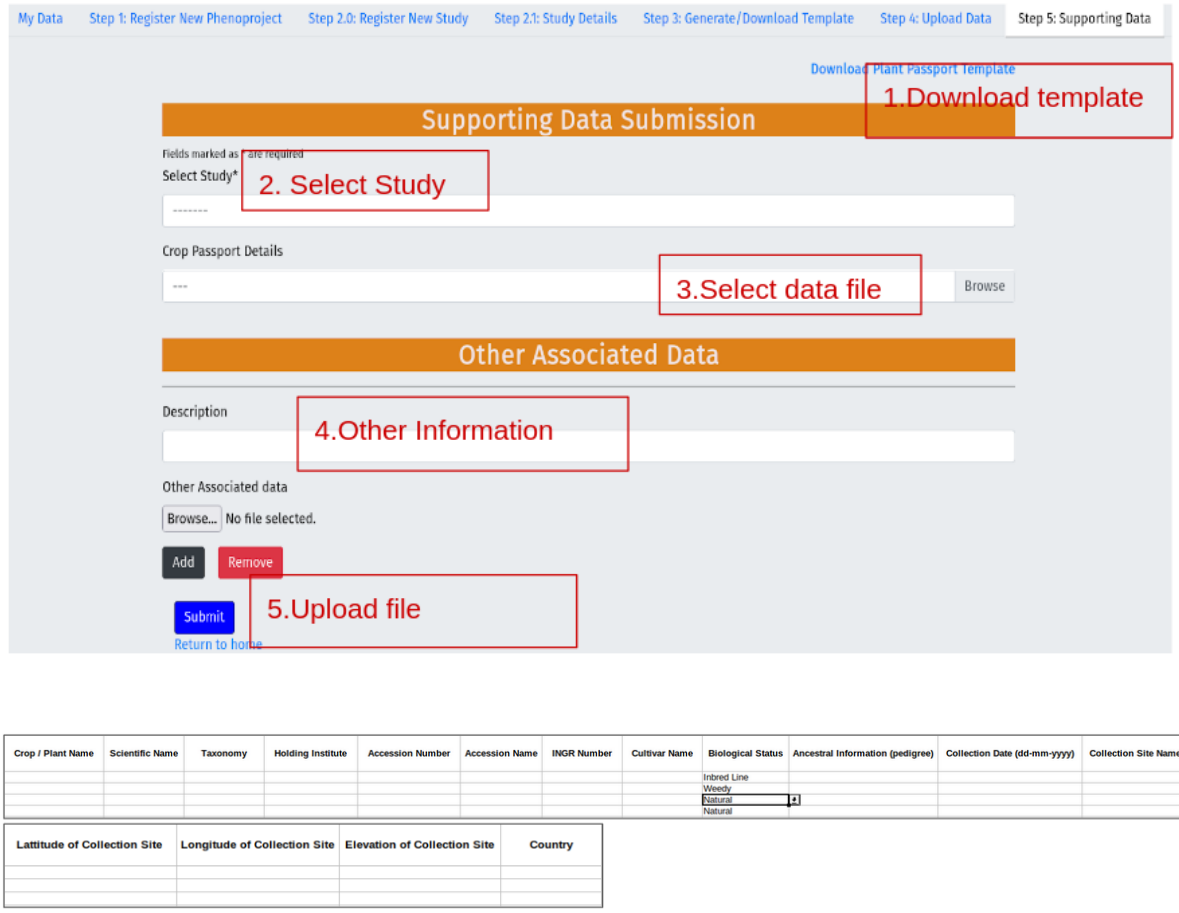


Figure 12. Snapshot of Step 5: Supporting Data submission and plant passport template file.

Template-based Submission

ICPD also allows the data submission by email through an alternate mode called '**Template-based submission**' in case the user is not able to submit by web-based interface. Although we strongly recommend the web-based interactive method for submission as it is very convenient and user friendly. Users are requested to download the template form and fill in the project and study-related metadata and required data in the data file before uploading to ICPD portal. The steps are summarised below (Figure 12):

1. Click on Download ICPD Submission template
2. Download the template file
3. Fill out the template sheet1-3 regarding the phenoproject, study and data.
4. Send the filled form to IBDC support: icpdicpdsupport@ibdc.rcb.res.in

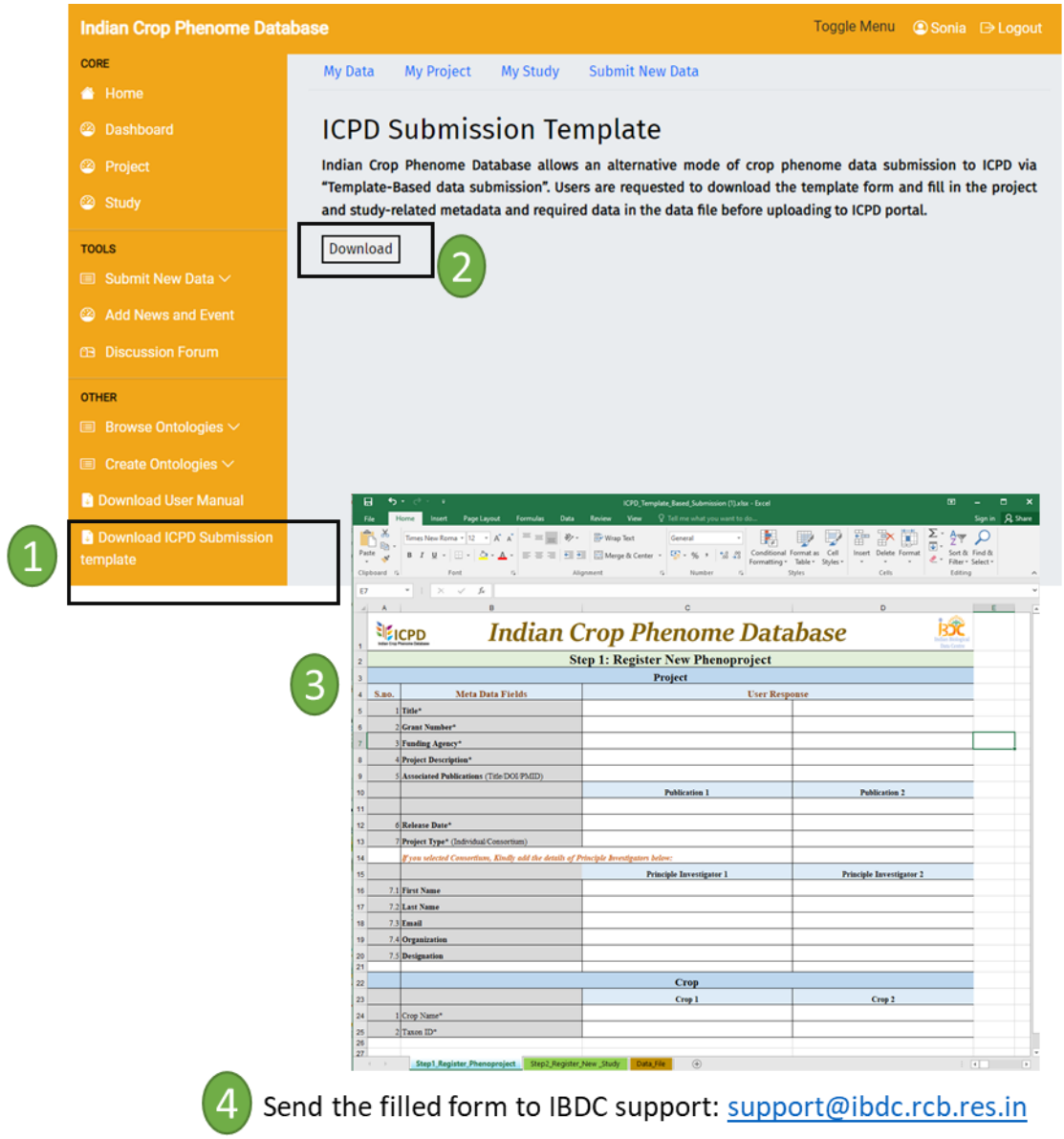


Figure 12. Snapshot of overview of template based submission option of ICPD

User Resources

My Data

The phenoproject, study and data files successfully submitted to ICPD can be viewed under the 'My data' tab of user dashboard (Figure 13). User can view and download the phenoproject and study meta-data. Similarly, the data files can also be downloaded anytime and anywhere.

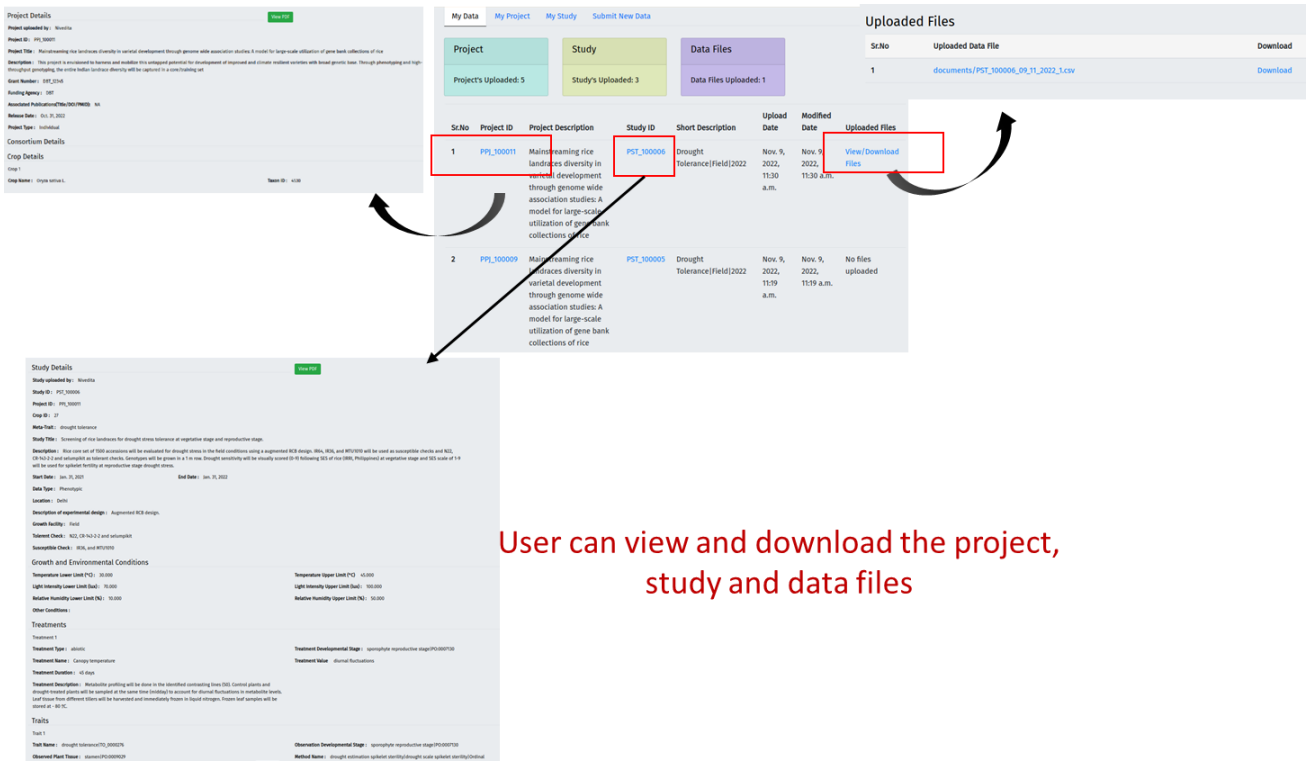


Figure 13. View and download meta-data and data files options in the user's dashboard.

Browse Data

The phenoproject, study and data files successfully submitted to ICPD can be Browse under "Browse data" option in home page (Figure 14). User can view and download the phenoproject and study meta-data. Similarly, the data files can also be downloaded anytime and anywhere.

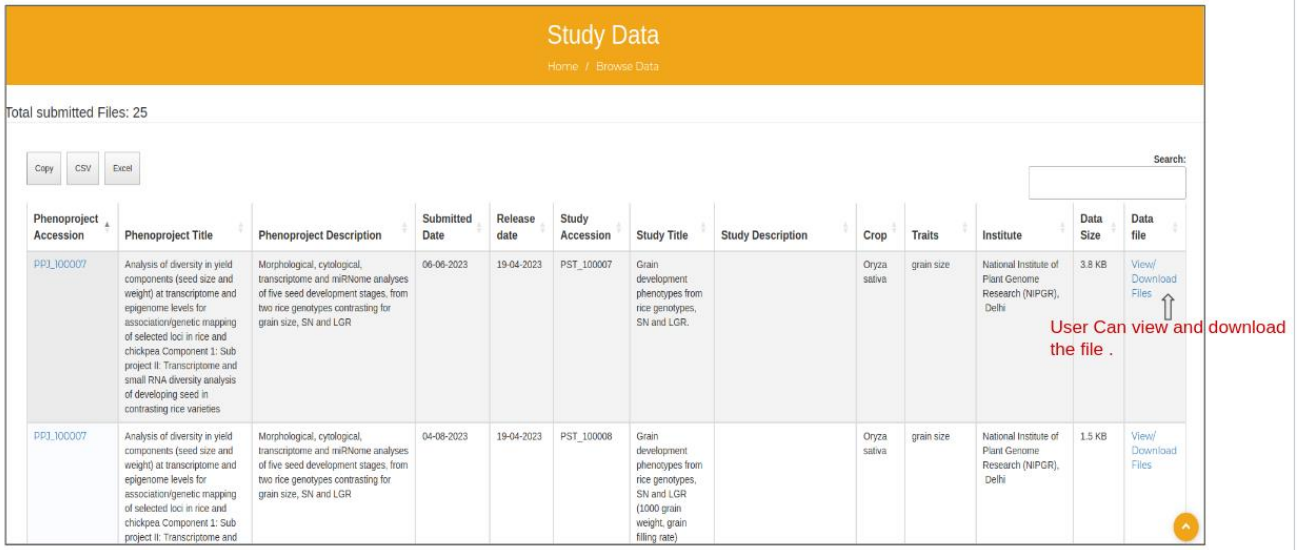


Figure 14: Snapshot of Browse data page

Browse and Create Ontology

As ontology based model of ICPD data submission is essential for documenting data to ICPD, it's possible that the user study specific terms might not be annotated here in the database. So users are advised to search the terms in the 'Browse Ontology tool' before creating a new ontology term. The process of creating ontology is very easy and the ICPD ontology database is instantaneously updated as soon as the term created and user can proceed with the submission without any delay.

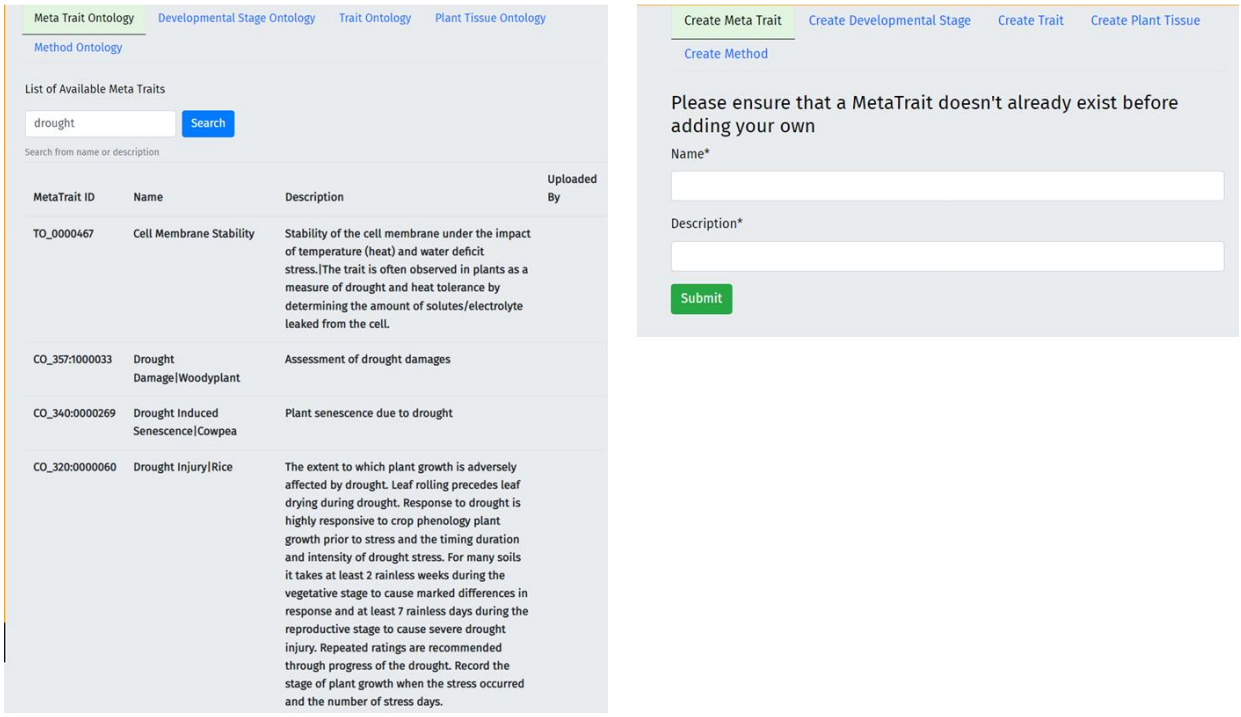


Figure 15: Snapshot of Browse ontology and create ontology page

User Manual and Sample forms

To guide users through the process of phenotyping data submission, the detailed standard operating procedures are available in the ICPD home page and user dashboard. Sample filled forms are given in the appendix section of this manual.

User Support

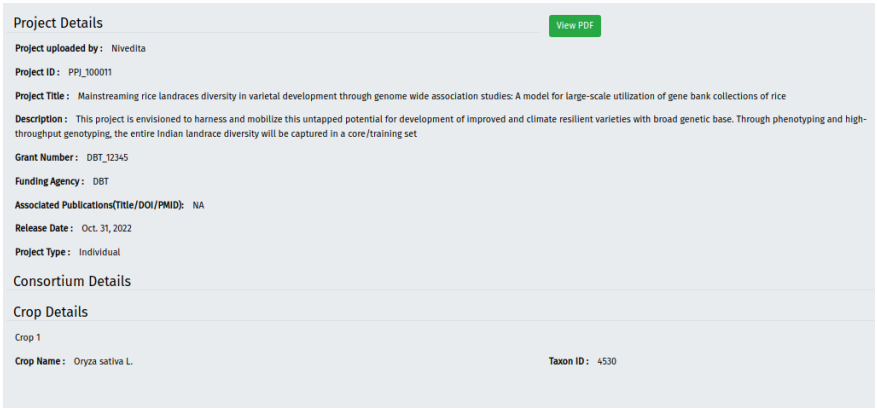
For any query, suggestion and support, kindly write to us @ICPD support: icpdicpdsupport@ibdc.rcb.res.in . Users can also watch the following video on youtube to know more about ICPD:

1. ICPD Webinar: https://youtu.be/Grx973vAelM?si=fz2pPwSKn-s3T1tG
2. Short video on ICPD: https://youtu.be/B-WjgpD57ag?si=CrJkT-u2bAIXW8-H

Appendix

Sample Forms

1. Phenoproject



Project Details

View PDF

Project uploaded by : Nivedita

Project ID : PPJ_100011

Project Title : Mainstreaming rice landraces diversity in varietal development through genome wide association studies: A model for large-scale utilization of gene bank collections of rice

Description : This project is envisioned to harness and mobilize this untapped potential for development of improved and climate resilient varieties with broad genetic base. Through phenotyping and high-throughput genotyping, the entire Indian landrace diversity will be captured in a core/training set

Grant Number : DBT_12345

Funding Agency : DBT

Associated Publications(Title/DOI/PMID): NA

Release Date : Oct. 31, 2022

Project Type : Individual

Consortium Details

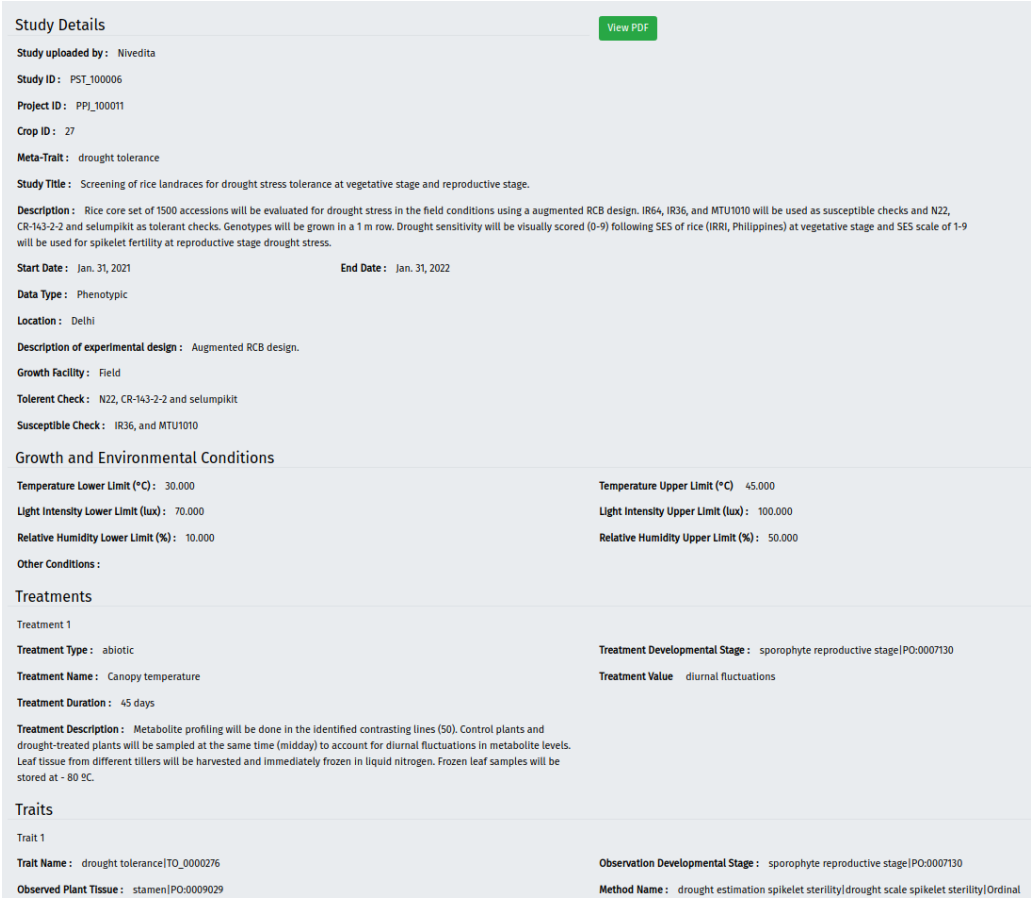
Crop Details

Crop 1

Crop Name : Oryza sativa L.

Taxon ID : 4530

2. Study



Study Details

View PDF

Study uploaded by : Nivedita

Study ID : PST_100006

Project ID : PPJ_100011

Crop ID : 27

Meta-Trait : drought tolerance

Study Title : Screening of rice landraces for drought stress tolerance at vegetative stage and reproductive stage.

Description : Rice core set of 1500 accessions will be evaluated for drought stress in the field conditions using a augmented RCB design. IR64, IR36, and MTU1010 will be used as susceptible checks and N22, CR-143-2-2 and selumpikit as tolerant checks. Genotypes will be grown in a 1 m row. Drought sensitivity will be visually scored (0-9) following SES of rice (IRRI, Philippines) at vegetative stage and SES scale of 1-9 will be used for spikelet fertility at reproductive stage drought stress.

Start Date : Jan. 31, 2021 **End Date :** Jan. 31, 2022

Data Type : Phenotypic

Location : Delhi

Description of experimental design : Augmented RCB design.

Growth Facility : Field

Tolerent Check : N22, CR-143-2-2 and selumpikit

Susceptible Check : IR36, and MTU1010

Growth and Environmental Conditions

Temperature Lower Limit (°C) : 30.000 **Temperature Upper Limit (°C)** 45.000

Light Intensity Lower Limit (lux) : 70.000 **Light Intensity Upper Limit (lux) :** 100.000

Relative Humidity Lower Limit (%) : 10.000 **Relative Humidity Upper Limit (%) :** 50.000

Other Conditions :

Treatments

Treatment 1

Treatment Type : abiotic **Treatment Developmental Stage :** sporophyte reproductive stage|PO:0007130

Treatment Name : Canopy temperature **Treatment Value** diurnal fluctuations

Treatment Duration : 45 days

Treatment Description : Metabolite profiling will be done in the identified contrasting lines (50). Control plants and drought-treated plants will be sampled at the same time (midday) to account for diurnal fluctuations in metabolite levels. Leaf tissue from different tillers will be harvested and immediately frozen in liquid nitrogen. Frozen leaf samples will be stored at - 80 ºC.

Traits

Trait 1

Trait Name : drought tolerance|TO_0000276 **Observation Developmental Stage :** sporophyte reproductive stage|PO:0007130

Observed Plant Tissue : stamen|PO:0009029 **Method Name :** drought estimation spikelet sterility|drought scale spikelet sterility|Ordinal

3. Data file

sr_no	study_id	project_id	meta_trait_name	treatment_agent	treatment_qualifier	trait_name	tissue	accession	b1_t1	b1_t2	b1_t3	b2_t1	b2_t2	b2_t3	b3_t1	b3_t2	b3_t3
1	PST_100006	PPJ_100011	drought tolerance	Canopy temperature	diurnal fluctuations	drought tolerance	stamen	A1	12	45	6	77	88	99	7	56	43
2	PST_100006	PPJ_100011	drought tolerance	Canopy temperature	diurnal fluctuations	drought tolerance	stamen	A2	3	9	6	5	12	11	21	23	4
3	PST_100006	PPJ_100011	drought tolerance	Canopy temperature	diurnal fluctuations	drought tolerance	stamen	A3	2	89	43	55	45	12	65	NA	54
4	PST_100006	PPJ_100011	drought tolerance	Canopy temperature	diurnal fluctuations	drought tolerance	stamen	A4	NA	9	56	78	65	45	87	54	43
5	PST_100006	PPJ_100011	drought tolerance	Canopy temperature	diurnal fluctuations	drought tolerance	stamen	A5	6	76	7	98	87	65	12	11	32
6	PST_100006	PPJ_100011	drought tolerance	Canopy temperature	diurnal fluctuations	drought tolerance	stamen	A6	6	6	6	23	98	78	45	34	1