Introduction

This document gives an overview on the order of steps necessary to submit information on your trial data to the Brassica Information Portal. This can be done using the wizard and the ruby client.

Data submission to the Brassica Information Portal is split into two different submissions and need to be executed in the following order:

1) Population Submission
2) Trial Submission

BIP user account

In order to submit content to the Brassica Information Portal, you need to sign in with your ORCiD account. If you don't have an ORCiD account yet, you will be referred to their services from our Sign in:

To log in or register, go to bip.earlham.ac.uk and click on the 'sign in with ORCiD' button in the top right corner.
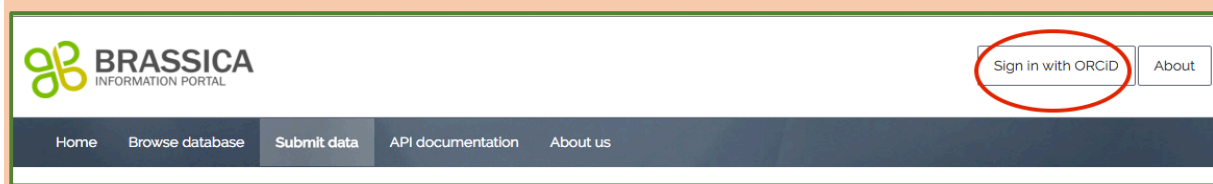
Trial submission using the wizard

Ones your experimental plant population is deposited in BIP, you can submit the trait scoring data (your measured traits) of your trial. This is a 6-step process, during which you also submit metadata that describes your trial.

Go to bip.earlham.ac.uk and click on submit data. Choose "Plant scoring trial" and click "START!". The wizard walks you through all the steps, offering compulsory and optional fields to be filled out.
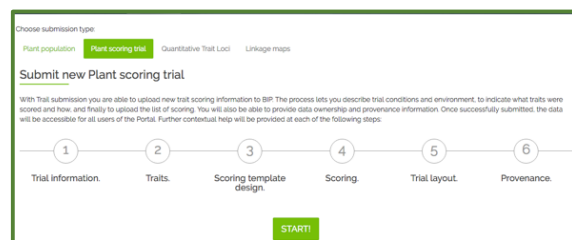

Figure 1 Trial data submission wizard

In the table below you see a list of all fields present during the submission process. Those fields marked with * are compulsory for the submission.

Open the 2_Trial_data – file you will be using for the upload and become familiar with the data. The file contains all columns necessary for the submission of this trial raw data. Each step in the submission corresponds to several actions, so please follow the instructions below and on the website.

Step 1: Provide general trial information. When coming up with a Plant trial name, you may apply following scheme:
<institute_id>_<YYYY>_<trialSubject>_<raw/processed>
An example of existing Trial name is:
U.Nottm_2016_RIPRleafminerals_REMLmeans
In this step you also link your previously submitted plant_population to your plant_trial, by selecting your experimental plant_population from a drop-down menu.

Step 2: Define traits of interest. If you have measured traits with the same method as are already registered in the database, you start typing the trait's name in the box beneath "trait descriptor list". If the trait does not yet exist, you can define a new trait by clicking "add new trait descriptor". When adding new traits, some fields are compulsory, following the Crop Ontology model of <Trait><Method><Scale>. Other information can be added in case it is available.

Step 3: Specify what information you will upload together with your trait scores. This can be the number of technical replicates or the trial design in case you want to submit raw data (as in a tutorial example).
You may also add information on the plant_lines or plant_varieties associated with your germplasm.

Step 4: Download template spreadsheet you will later use for the submission of your trait data. The same template could be used in practical work of data collection throughout your project. It has been created according to the choices you have made so far in your submission process. If you realise that the headers do not correspond to the data you want to submit, navigate back to previous submission steps and amend them accordingly. An altered template can then be downloaded at this step.
Please **be careful when pasting your trait scores** beneath the correct headers, as they may not appear in the order they are recorded in your source spreadsheet.

Step 1 - Trial information

Plant trial name*

Project name*
Experimental plant population*

<div style="text-align:right">Your previously submitted plant population</div>

Trial year*
Trial description*
select: data status: raw vs. processed (analysed) data
Institute*
Terrain
Soil type

Statistical factors

Country*

Place name*

Trial location site name

Latitude

Longitude

Altitude

Step 2 -Traits

Select trait descriptors*

When adding new trait descriptor:

Descriptor name*	Materials	Precautions
Trait category*	Instrumentation required	Scoring method
Units of measurements*	Calibrated against	Possible interactions
Score type	Likely ambiguities	Additional annotations
Where to score	Controls	

Step 3 - Scoring template design

Select genetic material origin

plant lines or plant varieties

select design factors:block,plot,rep etc (raw data)

specify technical replicate numbers (raw data)

Step 4 - Scoring Template Submission

---

download .csv file template

Step 5: This step is optional, and allow attaching an image of the trial layout. Such an image would be helpful to interpret patterns in the raw data.

Step 6: Provide information about the provenance, e.g. list personnel involved in generating the data you are about to submit. You can choose to put an embargo on the data and wait with the submission until for example your paper is ready to be published. Whilst under embargo, the data is only visible to you (when logged in).

add trait scores to corresponding header

add plant scoring unit (=sample_id)*, plant accession*, originating organisation*,

Plant line or variety*

upload .csv file*

Step 5 Submit Trial Layout

Submit image of trial layout

Step 6- Provenance

Data owned by

Data provenance

Comments

Visibility-public/private*

We start with a simple example. In the reality, extend of meta-data required to properly describe an experiment is larger and one often deals with incomplete data. In the following section 'Trial submission extended', we work with more complex case.

A)

Plant scoring unit name	polytunnel	rep	sub-block	pot	Plant accession	Originating organisati	cultivar_nam	a-toc 1st rep	a-toc 2nd rep	y-toc 1st rep	y-toc 2nd rep
p-0000001	1	1	1	1	U.Nottm_BnASSYST-001	Nottingham	Alesi	89.057061	268.55549	162.27601	
p-0000002	1	1	1	2	U.Nottm_BnASSYST-002	Nottingham	Remy	172.73416	99.85896		
p-0000003	1	1	6	3	U.Nottm_BnASSYST-003	Nottingham	Robust		71.130024	109.731	
p-0000004	1	1	6	4	U.Nottm_BnASSYST-004	Nottingham	Alaska		131.91514	172.82017	
p-0000005	1	1	6	5	U.Nottm_BnASSYST-005	Nottingham	Pirola		109.731	242.14636	109.731
p-0000006	1	1	6	6	U.Nottm_BnASSYST-006	Nottingham	Milena	99.85896	172.82017	213.6103	172.82017
p-0000007	1	1	6	7	U.Nottm_BnASSYST-007	Nottingham	Allure	71.130024	242.14636	109.731	242.14636
p-0000008	1	1	6	8	U.Nottm_BnASSYST-008	Nottingham	Agalon	131.91514	154.53718	172.82017	161.8433

B)

Plant scoring unit name	Plant accessi	Originating o	Plant variety	Leaf Caesium	Leaf Silver concentration	Leaf Aluminium conce	Leaf Arsenic	Leaf Boron c	Leaf Calcium	Leaf Cadmiu	Leaf Chromium concentration
a-0000001	U.Nottm_Bn	Nottingham	Alesi	0.17	0.01	9.12	0.02	12.09	9353	0.15	0.66
a-0000002	U.Nottm_Bn	Nottingham	Remy	0.13	0.01	8.55	0.02	11.46	9979	0.18	0.68
a-0000003	U.Nottm_Bn	Nottingham	Robust	0.18	0.01	12.57	0.02	10.64	8866	0.12	0.68
a-0000004	U.Nottm_Bn	Nottingham	Alaska	0.15	0.02	9.9	0.02	12.89	9282	0.13	0.63
a-0000005	U.Nottm_Bn	Nottingham	Pirola	0.18	0.01	8.92	0.02	14.14	10882	0.17	0.67
a-0000007	U.Nottm_Bn	Nottingham	Milena	0.16	0.01	8.99	0.02	12.55	10297	0.18	0.67
a-0000008	U.Nottm_Bn	Nottingham	Allure	0.2	0.01	9.79	0.02	14.1	11111	0.18	0.68
…	…	…	…	…	…	…	…	…	…	…	…

*Figure 2 Readily filled out submission templates, which would be uploaded to the BIP in Trial submission step 4. The user determines the final content of the trial submission template depending on the type of data submitted. **A)** A submission template enabling raw data submission was filled out with information on the trial design (columns B-E) as well as technical replicates (columns J, L) **B)** A submission template generated for processed data, where the trait measurements (columns E-L) are derived from statistical analysis.*

Morphological traits to be added manually:

Trait: plant dry weight
Unit of measurements: kg
Scoring method: Remove soil from roots, let dry in oven for 24 h, weigh on a scale.
Materials:
Plant part: whole plant

Trait: basal root thickness
Unit: cm
Method: generate cross-section in the middle of the basal root (tip to beginning) and measure thickness using a ruler
Plant part: basal root

Trait: root dry weight
Unit: mg
Method: remove root from aboveground biomass, clean with water, careful not to remove roots during that process. Dry in a bag a drying chamber for 24h, measure weight of roots within the bag by putting on a scale. A generic dried bag –weight was subtracted from all samples.

Trait: stem colour
Unit: colour
Method: pictures were taken at the same light intensity with a black square as reference in all images and analysed with Image J.

Trait: Lower stem length
Unit: cm
Method: use a tape measure to measure length of lower stem beginning at lowest leaf to the ground.

Biochemical traits to be added manually:

Trait: Leaf nitrogen content
Unit of measurements: ng
Scoring method: Kjeldal, see Publication et al 1967
Plant part: leaf

Trait: Seed copper concentration
Unit of measurements: ng/mg (dry weight)
Scoring method: 3 seeds per tube and occasionally four for very small seeds from multiple replicate plants per accession were left to soak in nitric acid over night and then digested in block heaters prior to analysis by inductively coupled plasma-mass spectrometry (ICP-MS). This data was then weight normalised so that data from each element was in the units mg/kg
Plant part: seed

Trait: Pod and stem Aldehyde functional group content
Unit: µg
Scoring method: * 3 pods and 3 sections of stem (10 cm) were collected from each plant, and inserted into Pyrex glass tubes (122 mm (L) x 19 mm (D), roughly 20 ml vol.).
* Wax was extracted using chloroform. Chloroform was spiked with a C22:0 alkane and C17:0 fatty acid as internal standards. Internal standard concentrations were 10 µM in chloroform (giving 100 pmol/µl for GC

injection). Wax was extracted for 30 seconds using 10 ml of the chloroform/IS mixture. Sample tubes were inverted continuously during the extraction phase. PTFE cap liners were used to ensure minimal extraction of plasticisers.

* Pod and stem material are removed from the tubes, and samples dried under vacuum in a centrifugal evaporator (Genevac, EZ-2) at ? 42°C using program low BP. Samples can also be dried under a stream of nitrogen at ? 30°C.

* Wax components were derivatized with 200 µl of BSTFA + TMCS (99:1) for 1 hour at 85°C.

* After cooling, 800 µl of heptane was added (total vol. 1 ml), and 1 µl was injected onto the GC-FID/MS.

* GC program:

* Initial temp: 50°C, to 200°C at 50°C/min, to 280°C at 2°C min, to 325°C at 10°C/min, hold for 6.5 minutes.

* Inlet temp: 280°C, split-less injection.

* Column: HP-1MS UI, 30 m x 0.25 mm x 0.25 µm, flow = 1 ml/ min.

* FID temp: 325°C

-a beautifully long and detailed description of the method- good for trait reproducibility.

Trait: Canopy leaf Kjeldal N
Unit: ng/mg (dry weight)
Method: 10 canopy leaves (the youngest leaf that is almost fully expanding, not cauline), from 10 plants representing the plot were sampled into a C5 pre-labelled brown envelope. Labelled control envelopes were placed at the end of each 13 winter plots and every 10 spring plots for use in subsequent biomass calculations. The 10 leaves were then weighed on a balance in the envelope that they were collected in. Balance was set to auto collect weights.
Plant part:leaf

Trait: Early leaf Kjeldal N
Unit:ng/mg (dry weight
Method: 10 early leaves (leaf that is still expanding, usually second most youngest leaf) were sampled from 10 individual representative plants per plot. They were picked and put into an envelope that was pre-labelled. The samples were weighed on a balance in the envelope that they were collected in. Balance was set to auto collect weights.
Plant part:leaf

Trait: Chlorophyll-a content
Plant part: Leaf
Unit: %
Method: the percentage of chlorophyll a over total Chlorophyll content was generated using a spectrometer, following Wellburn et al 1977.

Trial Submission extended

This section uses the Trial submission wizard again, but draws from a more complex data sheet to give a more "real world" example.

Open the 2b_Trial_data – file you will be using for the upload and become familiar with the data. This file is a bit more complex: The format follows one data model from the RIPR spreadsheets, but the content is made up. You will notice that some columns may not need to be submitted to the database as they contain the same information and would lead to unnecessary duplication of information. For example, the header sample_id and sample are very similar. Follow the same steps

as before. After downloading the submission template, you have to decide what information from this spreadsheet can be omitted from the submission.

Answer: sampls, line_number, Collaborator1_name, Internal_accession_name, JIC_internal_accession_id, ug/g

Glossary

Experimental Plant population: A set of genetically diverse germplasm used for phenotyping experiments.
Population type: genetic background of the experimental plant population.
Accession: seed or plants of a plant line arising from a single generation.
Originating Organisation: The Institution at which the Accession is generated.
Sequence: SRA Sequence identifier, which will be automatically cross-linked to the SRA entry.
Year: the year the seed lot has been produced.

Population type descriptions

Population Type	Description
BC1	a first generation backcross population: the result of crossing the F1 of a cross between two fully homozygous diploid parents to one of the parents; the software detects from the genotype
F1	First generation of two distinctly different inbred lines. This is often not considered for phenotyping unless studying heterosis. In agronomy, the term "F1 hybrid" is usually reserved for agricultural cultivars derived from two parent cultivars. If using two parent cultivars, mention F1
F2	an F2 population: the result of selfing the F1 of a cross between two fully homozygous diploid parents. (https://www.kyazma.nl/docs/MQ6Manual.pdf, p 47)
F3	an F3 population: the result of selfing the F2 of a cross between two fully homozygous diploid parents. (partially adapted from definition of an F2 population: https://www.kyazma.nl/docs/MQ6Manual.pdf, p 47)
RIx	a population of recombinant inbred lines in the x-th generation: the result of selfing an F2 with single seed descent; x must be specified: 2<=x<=99, RI2 is equivalent to an F2. (https://www.kyazma.nl/docs/MQ6Manual.pdf, p 47)
DH	a doubled haploid population: the result of doubling the gametes of a single heterozygous diploid individual, linkage phases originally (possibly) unknown. (https://www.kyazma.nl/docs/MQ6Manual.pdf, p 47)
DH1	a doubled haploid population produced from the gametes of the F1 of a cross between two homozygous diploid parents. (https://www.kyazma.nl/docs/MQ6Manual.pdf, p 47)
DH2	a doubled haploid population: the result of doubling the gametes of an F2 population, one doubled gamete from one F2 plant. (https://www.kyazma.nl/docs/MQ6Manual.pdf, p 47)

HAP	a haploid population: the gametes (or derived individuals) of a single heterozygous diploid individual, linkage phases originally (possibly) unknown. (https://www.kyazma.nl/docs/MQ6Manual.pdf, p 47)
HAP1	a haploid population derived from the F1 of a cross between two fully homozygous diploid parents. (https://www.kyazma.nl/docs/MQ6Manual.pdf, p 47)
Crossing Population	a population resulting from a cross between two heterogeneously heterozygous and homozygous diploid parents, linkage phases originally (possibly) unknown. (https://www.kyazma.nl/docs/MQ6Manual.pdf, p 47)
BCpxFy	advanced backcross inbred line family: starting from the BC1 repeatedly backcrossing to the same parent (as used for the BC1) of each individual resulting in a single offspring per individual, followed by selfing with single seed descent; the backcross parent p and the generations x and y must be specified: p = A or B, x is the number of backcrosses including the one for creating the BC1: 1<=x<=99,y is the number of selfings: 0<=y<=99, BCa1F0 is equivalent to BC1. (https://www.kyazma.nl/docs/MQ6Manual.pdf, p 47)
IMxFy	random intermating (preferably chain crossing) the individuals resulting in a single offspring per individual, followed by selfing with single seed descent; the generations x and y must be specified: x is the number of intermatings including the two for creating the F2: 2 <= x <= 99, y is the number of selfings: 0<=y<=99,IM2F0 is equivalent to F2, IM2Fy is equivalent to RIx with x=2+y. (https://www.kyazma.nl/docs/MQ6Manual.pdf, p 47)
Experimental collection	A collection of often wild individuals from across the globe, grouped together as population.
Diversity core collection	Selected varieties and lines from germplasm from multiple countries that display diverse field performance and growth types.