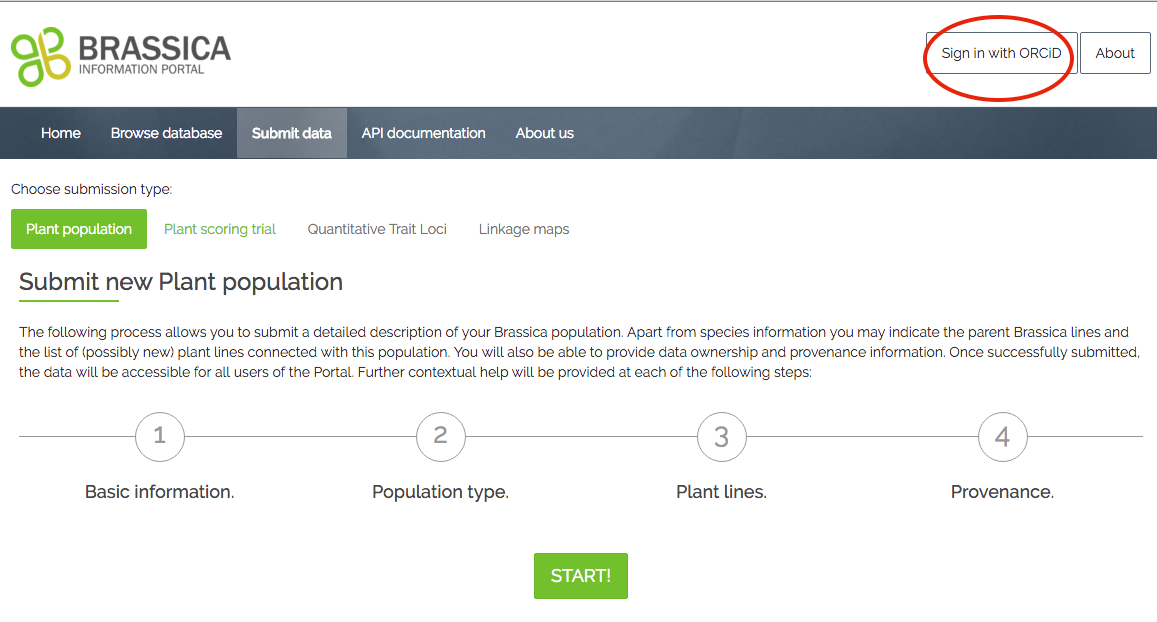
This document is a manual for the steps taken during the data upload workshop on the 07.09.2016.

It lists the steps necessary to use the ruby client to submit population data and the wizard to submit trial data to the Brassica Information Portal.

It further names all fields required for the submission of a new experimental plant population and plant trial.

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 register, go to [bip.earlham.ac.uk](https://bip.earlham.ac.uk/) and click on the ‘sign in with ORCiD’ button in the top right corner.

Please create an account prior to the workshop, as our time is limited.



Your submission is split into two different submissions which need to be executed in the following order:

1. Population Submission (ruby client)
   1. Population Submission extended
2. Trial Submission (web- interface based)
   1. Trial Submission extended

**Population Submission using the ruby client**

You submit your experimental plant population first. This is the collection of the genetically different plant material used for your trial. In your trial submission you are asked for the name of your experimental population, which is why it is necessary to submit population information first. The fields required and the population submission client are specific to a diversity foundation set, not a crossing population.

The population submission client is a ruby script parsing information (objects) from a .csv spreadsheet provided by the user to the mapped location in the BIP database via the BIP-API. In order to understand the underlying resource and attribute names used in the script, please see the respective tables in the [API-documentation](https://bip.earlham.ac.uk/api_documentation).

During your submission you will be using terminal. Please be familiar with the basics in navigating folders and moving and manipulating files. We will be working in pairs for this part so if you are not comfortable, I will pair you up with a more experienced person.

**Install software:**

-download ruby 2.3+

-for OS X e.g. use brew:

download brew following the instructions on the website:

<http://brew.sh/>

then type:

brew install ruby

-for Windows, use rubyinstaller:

Download the installer by following the instructions on the website:

<http://rubyinstaller.org/>

**Files required:**

In the workshop materials folder, files required for the population submission are files numbered with 1\_ to 2\_. 1b and 2b are used for an extended population submission with more complex script modification. We will start with the following files:

**1\_BnaDFFS.csv**

* The input .csv file which contains metadata about the experimental population

**2\_population\_submission.rb**

* The ruby client, which helps to transfer (“parse”) the content from 1\_BnaDFFS.csv into the Brassica information portal.

1\_BnaDFFS.csvreflects the content of the current commands in the population\_submission.rb client and can be altered to your requirements/ available data.

**File manipulation:**

The population\_submission.rb client currently enables the submission of all fields listed in table 1.

This section walks you through each step that needs to be taken in order to modify the client and input template for your own population submission.

**Population description:**

Open the file 2\_population\_submission.rb and scroll through the client-file. It may look daunting, but the sections in which information needs to be altered or added are well signposted using by the below patterns:

#++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

The actions to be taken for the file manipulation are also documented in the script. Here is an example:

# for workshop: add the column number from your CSV beside the correct variable.

For your own population submission, you will have to alter the following information directly in the client:

[ACTION] Beneath “ 1. Creating experimental plant\_population “, you will see information as displayed in figure 1.

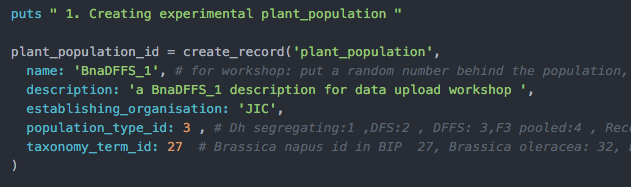


Figure example text describing the plant population in population\_submission.rb

Please enter your population information such as the population-

- name,

- description

- establishing\_organisation

- population\_type\_id

- taxonomy\_term\_id

directly in the script.

Nomenclature examples for the plant\_population.name are: BnaTNDH (Doubled Haploid mapping population derived from cross between Tapidor DH and Ningyou 7) or BnaDFS (a *Brassica napus* diversity foundation set).

**The .csv file**

|  |
| --- |
| **Table 1. Fields required\* for the submission of an experimental Population. They are presented below by <resource\_name>.<attribute\_name>.**   * Plant\_populations.name * Plant\_populations.population\_type * Plant\_populations.description * Plant\_population.establishing\_organization * Plant\_lines.plant\_line\_name * Taxonomy\_terms.id * Plant\_varieties.plant\_variety\_name * Plant\_accessions.plant\_accession * Plant\_accessions.year\_produced * Plant\_acessions.originating\_organisation   \*Note that more fields are available for submission of information to the BIP; to find the description of their names, please go to our [API-documentation](https://bip.earlham.ac.uk/api_documentation). |

With the help of the client 2\_population\_submission.rb you are able to read in the content of your own .csv input file with your population metadata to the Brassica Information Portal.

The template 1\_BnaDFFS.csv (Figure 2) reflects the current commands in the 2\_population\_submission.rb client to read in data from that .csv file. It can be altered to the information available on your experimental plant population.

The template files (Figure 1 and 2) contain the minimal requirements of information for a population submission to the Brassica Information Portal. More fields are available in the database. To find out which information can be added, please go to our [API-documentation.](https://bip.earlham.ac.uk/api_documentation)

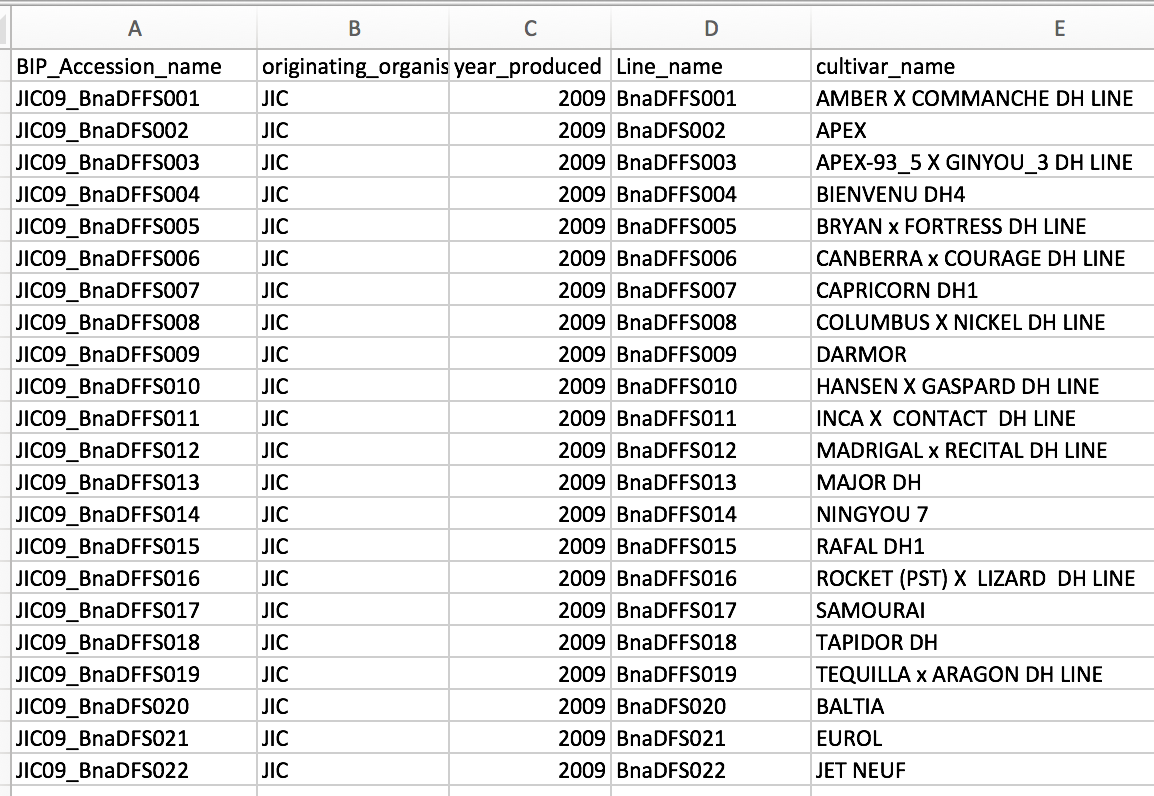


Figure 2 1\_BnaDFFS.csv is a mock- population template that corresponds to the population submission client in the workshop material..

**Defining the input to the script**

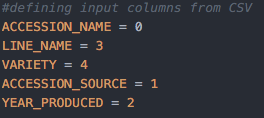


Figure the input variables must correspond to the location of the column headers in your .csv file.

In script section “# defining input columns from CSV in the script”

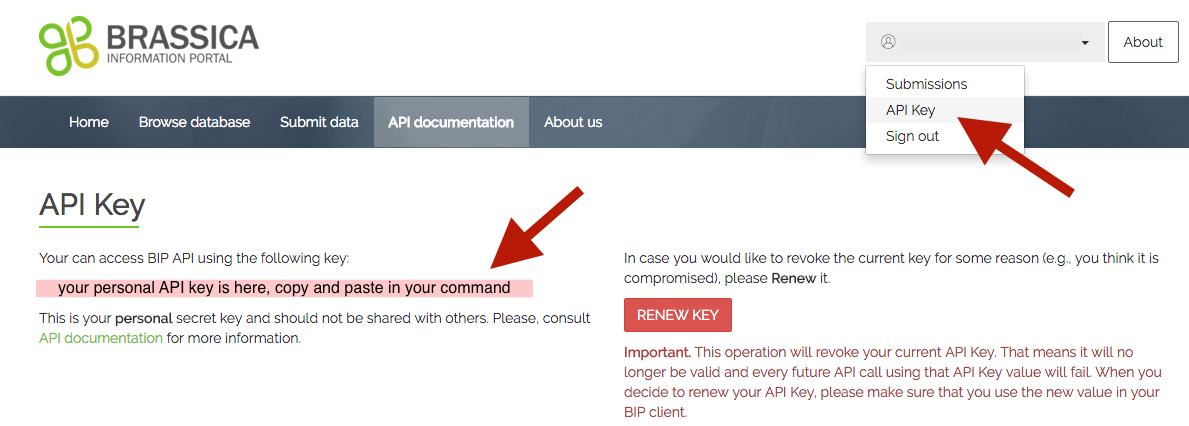
you define the client’s input parameters using the columns from your .csv. It is possible to add more input parameters. A list of required and additional parameters can be found in table 1 and the API documentation.

0 in the script is the first column in your .csv. 1 is the second column in your .csv. For example you can see that Accession\_name information is located in the first column in the .csv template (Figure 2), but needs to be defined as “0” in the client file ( Figure 3) and so on. [ACTION] open 1\_BnaDFFS.csv and 2\_population\_submission.rb, navigate to the appropriate section in the script and add the column number from your CSV beside the correct variable.

**API key:**

[ACTION] Sign in to BIP to retrieve your API key:

1. Click on API key in the popup bar under your user name.
2. Copy the API key and paste in your command to run the submission client (see below).

****

**Run submission client:**

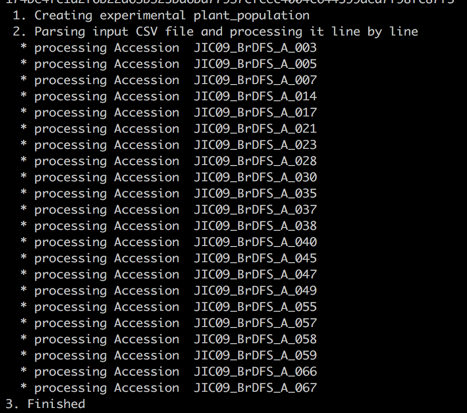
After double-checking your submission file and script, use your terminal or command prompt to navigate into the folder your workshop files are located in and run the command below. Make sure your input file and the submission script are located in the same folder or add the paths to their name:

ruby 2\_population\_submission.rb <your\_input\_file.csv> <your\_api\_key>

If you are new to command line, click on the links for basic information on how to navigate

your computer [using terminal (OSX)](https://www.youtube.com/watch?v=-Vl4rpZVA6I) or [Command Prompt (Windows)](https://www.youtube.com/watch?v=MNwErTxfkUA) .

[ACTION] Run the client with your input file and api key using the command above.



You should receive a message that looks very similar to this the one in Figure 4. Now you can open the [Brassica Information Portal](https://bip.earlham.ac.uk/) and find your population by typing its name into the main search bar.

Figure 4 Output after running the client successfully

**Population Submission extended**

**File manipulation level 2:**

**Manipulating a function**

Sometimes you have more information than the minimal requirements that always need to be submitted to the database. It is desirable to submit as much information about your experimental population as possible. Using the client makes it easy to do so. In that case you can simply add a few lines of code to expand the submission client to the additional information you can provide. For example, you have information about the genetic status of the founder line material you used to bulk up seeds for your experiment ( see workshop material file 1b\_BnaDFFS\_genstatus). The first step is to re-define the input of your script (see Fig. 3 in comparison to Fig. 5), where a new variable name, let us call it GENETIC\_STATUS, needs to be added and the input column number added accordingly. The variable can be called anything, as long as the variable name is consistently used throughout the client, it does not make a difference. We recommend using obvious names that you can recycle in future scripts. [ACTION] Add the Variable GENETIC\_STATUS to the script and assign it the corresponding column number from the .csv input file.

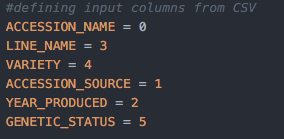


Figure The new variable GENETIC\_STATUS has been added and the location in the csv specified.

Further, [ACTION] add the genetic\_status attribute to the function that defines the submission of plant\_line associated information (as shown in Fig. 6).

Be aware that other information related to the population may not have to be submitted as plant\_line associated information, but for example as plant\_variety information. Then, you need to find the respective function and add attributes to that one instead.

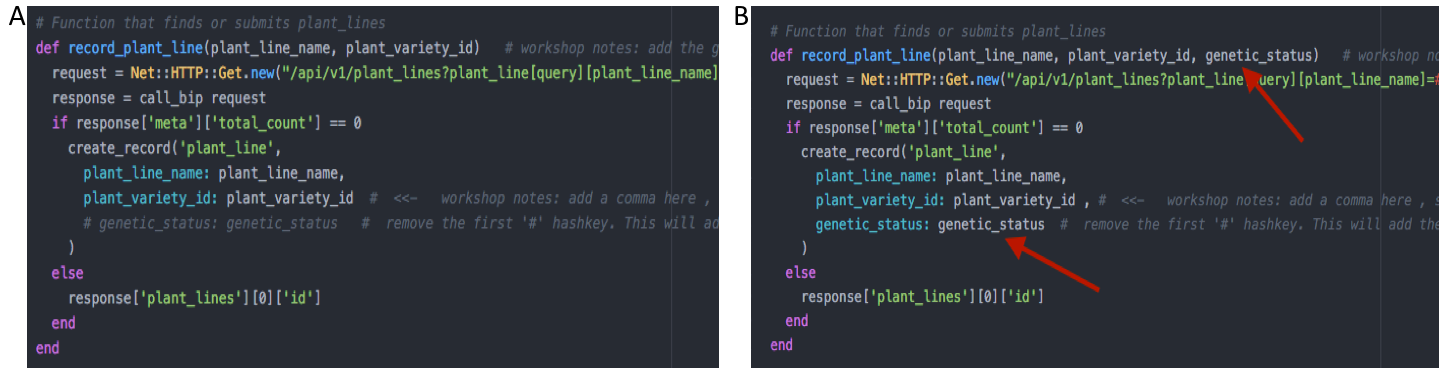


Figure 6 function that records plant\_line information. A: function without genetic\_status attribute. B the code that is added to the function to enable genetic\_status submission. A list of attributes can be found in the API-documentation.

Finally, [ACTION] find the line where the function gets called on the bottom of the script and add the input variable (previously defined as GENETIC\_STATUS) there, as also instructed in the client file (see Fig 6). Bear in mind that, should your attribute not belong to any plant\_line, but for example belong to an accession, you would have to add the variable to “record\_plant\_accession” instead.

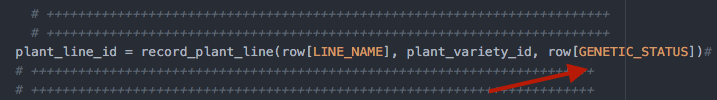


Figure 7 The calling of functions like “record\_plant\_line” from Fig.6 is located at the end of the script. The arrow points to the element added to enable GENETIC\_STATUS submission.

In the same fashion as the ability to submit genetic status information has been added to the script, one can delete objects that are not available for your submission. Bear in mind that minimal requirements need to be met ( see table 1). The API documentation lists all possible attributes that you could submit information to. In the case of a population submission, you can add information related to the Plant\_population, Plant\_lines, Plant\_accessions and Plant\_varieties.

**Double-checking the script**

-          Are all .csv headers correctly associated with the variables in the script?

-          Do variables need to be added to the script?

* If yes: make sure you add them in the appropriate create\_record function, with the wording of the fields exactly as shown in the API documentation
* Typos in the newly-added input parameters? -check with the API documentation

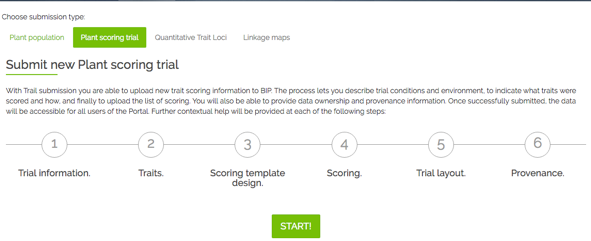
- Do variables need to be added /deleted from the script?

* If yes: make sure you delete them in section #defining input columns from CSV, in the create\_record function, and at the end of the script, where you call the function and define the input parameters.

Should you be having problems with in the altered script, look at file 2b\_population\_submission.rb , where all alterations have already been made for you and you can double check your changes with that file.

[ACTION] submit your altered script in the same way as described above.

**Trial Submission using the wizard**

After submitting your Population, 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](https://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. In the table below you see a list of all fields during the submission process. Those fields marked with \* are compulsory for the submission.

[ACTION] Open the 3\_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. In real life, things are more tricky. But to get your head around the actual submission process and initial practice, we work with easy data for now. In section Trial submission extended, we work with more complex data.

Each step in the submission corresponds to several [ACTION]s, so please follow the instructions below and the instructions on the website:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Step 1** asks you for general trial information. When coming up with a Plant trial name, try to follow the nomenclature in the table to the right.  An already existing Trial name is called 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.  **In step 2** you define the traits which you have investigated during your trial. For the workshop, you will submit traits that are already recorded in the database (the two first traits), and you will define one 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. Please see below for the description of your 3rd trait. The traits are made up and just act as an example of what you could fill in in case your traits are not yet recorded in the database.  **In step 3** you 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 –in the workshop, we submit raw data. Also, you specify whether you give information on the plant\_lines or plant\_varieties associated with your germplasm – in the workshop, we tick Plant varieties.  **In step 4** you will be able to download the template you use for the submission of your trait data. It has been created according to the choices you have made so far in your submission (e.g. traits to be submitted). If you see that the headers don’t 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 5** is optional, where you can submit an image of the trial layout, in case you submit raw data. Such an image would be helpful to interpret patterns in the raw data- in the workshop we don’t have such information.  **Step 6** asks you to fill in some information about the provenance of the data you just submitted. 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. | **Trial submission**   |  |  |  | | --- | --- | --- | | **Step 1 - Trial information** | | | | Plant trial name\*  <institute id>\_<YYYY>\_<trialSubject>\_<raw/processed> | | | | Project name\* | | | | Experimental plant population\*  BnaDFFS\_<group number> | | | | Trial year\* | | | | Trial description\*  mock trial submission during workshop | | | | 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 | | | | plant varieties | | | | specify technical replicate numbers (raw data) | | | | select design factors: block plot rep etc (raw data) | | | |  | | | | Step 4 - Scoring Template Submission | | | | download .csv file template | | | | 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\* | | | |

**Morphological traits to add 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 add 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.

[ACTION] Open the 3b\_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