**Steps in the submission of data to T3**

1. Submit all the line names and properties using the Line Submission Form.
2. Submit trial information using the Trial Submission Form (Experiment Annotation Form).
3. Submit phenotypic data using the Phenotype Submission Form. Step 3 can be iterated several times without going back to Step 2.

Notes on what information is needed on each form are given below. See data templates for more information.

**Line Submission Form**

Line Names are unique identifiers for a line within a crop. If a line with that name exists already within T3, the new submission will not be accepted but its breeding program, aliases, pedigree and properties will be displayed for confirmation (or not) that it is the same as the new submission.

Fields with an asterisk are required.

*\*Line Names*

In order to be able to search the database for specific lines, we have to use consistent formats. Use the following naming conventions for Line Names:

• Use all capital letters: BARONESSE

• Dashes within names can be maintained: 6B97-2245

• Spaces cannot be used in a name.  
Spaces will be replaced with an understroke “\_” if they occur between two letters (e.g., AC METCALFE –> AC\_METCALFE) or if they occur between two numbers (e.g., IA95 506 –> IA95\_506) but will be deleted if they occur between a number and a letter (e.g., SHORT 2 –> SHORT2 or IA95506 X10 –> IA95506X10). For any line that contains spaces or understrokes, an alias will automatically be created with both characters removed so that that line will turn up in a global search for the name.

*Aliases*

There can be several aliases, separated by commas. When a search is done on line names, lines with matching aliases will come up too.

*GRIN Accession*

This is the unique number used in the Germplasm Resource Information Network of the National Plant Germplasm System (NPGS). The number often starts with PI (for plant introduction). If you have this number please give it as that will facilitate linking T3 to GRIN.

*Pedigree*

Breeders can submit pedigrees in text format:

Mentor/Minerva//Vada\_mutant/4/Carlsberg/Union//Opavsky/Salle/3/Ricardo/5/Oriol/6153P40

We are using standard Purdy notation (Purdy et al., Crop Sci 8: 405-406).

Pedigrees in this format will be stored in a text field for reference by breeders. It is not necessary for the text version of pedigrees to be all caps with no spaces. Use the text format that is easiest for you to read.

Where possible, provide names of intermediate parents in the pedigrees. If a parent is already in the database, then it is not necessary to expand the pedigree to include earlier generations.

*\*Filial Generation*

This is the filial generation (number of generations of selfing since the hybrid + 1) of the single plant which is the common ancestor of all seeds that form this line. Possible values for Filial Gen. are single digits: integer from 1 to 8 corresponds to seeds from F1 to F8 derived plants ; anything greater than 8, use 9; if the seeds were from a doubled haploid (DH) plant, use 0.

*Values for descriptive variables:*

Variables with an asterisk are required.

Barley

\*Row Type: 2 or 6

Growth Habit: winter, spring, or facultative

Primary End Use: malt, feed, food, fuel, or other. Indicate the intended use. At the F4-F5 stage, the quality of a line for malting may not be known yet.

Hull: hulless, hulled, or empty (assumed to be hulled)

Wheat

\*Species: aestivum or durum

Grain Hardness: hard or soft, if unknown, NA

Color: Red or White, if unknown, NA

Growth Habit: Spring, Winter or Facultative

Awned or awnless: A or N (No awns)

Chaff Color: bronze or bright

Qualitative Height: tall or short

Comments: general text field to say something about the line (cannot be used as a query to sort lines)

**Trial Submission Form**

The following information can be entered in the form. Cells with a blue background are required, others are optional.

**Crop.** Wheat or barley

**Breeding Program.** Required. Program responsible for data collection (Three letter code for wheat, two letter code for barley)

**Trial Code**. Required. Experiment\_YYYY\_Location, where Experiment is short but descriptive, YYYY=Year Harvested, and Location is self-explanatory. Trial Codes should be unique across T3 for a crop.

**Experiment Code.** Optional. The experiment is one hierarchical level above the trial. The trial is carried out at one location in one year. The experiment may have several trials with similar (may be identical) entry lists, performed at different locations in different years. The Experiment Code should be short, descriptive, and unique across T3 for a crop.

**Location**. Required.

**Lat/Long of field.** Required. DO NOT use a degree symbol! Just put a number. We have found four coordinate standards:

|  |  |  |
| --- | --- | --- |
|  | Latitude | Longitude |
| Decimal Degrees (WGS84) | 42.4375 | -76.504167 |
| Degrees, Minutes & Seconds | N42 26 15 | W76 30 15 |
| GPS | N 42 26.250 | W 76 30.250 |
| UTM (X and Y) | 18N 376284 | 4699449 |

For simplicity, we would like to use the Decimal Degrees standard. There is a converter between these standards at <http://boulter.com/gps/>.

**Collaborator**. Required. Name of scientist.

**Experiment Description**. Optional. Agreed upon for TCAP-funded experiments, self-assigned otherwise. 300 characters maximum.

**Planting date.** Required.

**Harvest date.** Required. This is a required information. Therefore you can only submit this form AFTER you have harvested the experiment. If the experiment is not destined to be harvested, put the date of the last data collection event.

**Begin weather date.** Optional. If T3 should store weather data starting at some point *before* planting (e.g., to figure out soil moisture status), then put that date here. If blank, weather will be collected starting at planting date.

**Greenhouse trial?** (yes or no). Required.

**Seeding rate** (seeds/m2). Required. This is the TARGET density for the trial, not the actual rate for each line in the trial. Depending on how seed is packaged, rates may vary. Just give the target rate.

**Experimental design.** Required. A description of the design. We do not have a strict format for this description at this point.

**Number of entries.** Required. Include the checks in this number.

**Number of replications.** Required. Refers to the number of individual reps on which a particular trait was measured for normal entries (not checks: if there are repeated checks, checks may have more replications than normal entries).

**Plot size** (m2). Required.

**Harvested area** (m2). Required.

**Irrigation** (yes or no). Required.

**Other remarks.** Optional. Adjustments to means in data analysis or other particularities of statistical analysis. Other notes that may help in the interpretation of the results, for example, that harvest was delayed due to weather, etc.

**Phenotype Submission Form**

Phenotypes come from a single trial. Information about the trial must have been previously submitted to T3 using the Trial Submission Form and is identified by the Trial Code**.** The Trial Code row is the only row that has no information about the phenotypes, but has just one value, the Trial Code. The Trial Code given on this sheet is just an example.

\*All overall trial characterizations are required information.

**\*Trial Mean**: should be obtained from the lines being submitted to T3 lines and the checks included in the spreadsheet. If there were additional lines in the trial that are not being submitted to T3, they should be left out of the mean calculation and omitted from the submission form.

**\*Std Error**: For simplicity, we will assume a common error variance for all lines in an experiment (whether they are submitted to T3 or not), so the standard error of a mean can be calculated from the overall Mean Square Error (MSE) for the experiment. The formula for the Std Error = sqrt(MSE/r), where r is the number of replications.

**\*Replications**: refers to the number of individual reps on which a particular trait was measured for normal entries (not checks: if there are repeated checks, they may have more replications than normal entries). For most agronomic traits, this will be the same as the number of replications in the trial indicated in the Annotations file. For some quality evaluations, breeding programs may submit a single bulk sample, in which case the number of Replications for such traits would be = 1. Unless there are repeated checks in a trial, it may not be possible to provide estimates of standard errors (Std Error) and F statistics for such data. The standard deviation among a set of lines is not a useful measure of precision because it reflects the genetic variation among lines as well as errors in estimation. Breeders are encouraged to review their quality data carefully because they are in the best position to judge if values are reasonable for a given genotype.

Fields with an asterisk are required information for each line.

**\*Line Name**: T3 should also have been "made aware" of Line Names via the Line submission form.

**\*Check**: The value in the Check column indicates whether that line of data is the average for a line (0), for a standard check (1), or for a summary statistic, i.e., Trial Mean, Std. Error, Std. Error Diff., Prob>F, Coef. Var., or Replications at the top of the sheet (2). We have no rules on what is designated as a check and what is not. Information on whether a line is a check can be used by future analysts. In the meanwhile, breeders should use their informal understanding of what is a check in designating lines as checks. See below for information on the summary statisitics.

**\*Filial Gen**.: This is the filial generation of the single plant from which plant tissue was sampled for DNA extraction for genotyping. Possible values for Filial Gen. are single digits: integer from 1 to 8 corresponds to F1 to F8 derived plant tissue; anything greater than 8, use 9; if the tissue was from a doubled haploid (DH) plant, use 0.

**Seed stock year**: The year in which the seed that gave the DNA was harvested.

**Seed stock experiment**: The name of the experiment where the seed was harvested. This need not be a TCAP trial. The name is to ensure the same seed stock can be retrieved from the breeding program if needed.

**Seed stock ID**: The entry number of plot number within the experiment where the seed stock came from.

Values in the green columns should be means for each line calculated across the experiment, ideally considering lines as fixed effects (I think: this is something to be discussed with the statistics group).

**\*Pheno1**: Replace "Pheno1", "Pheno2" etc. with the names of phenotypes given in the T3 phenotype ontology page. Phenotypes without a matching name will not be accepted. If there is no data for a phenotype for a particular line, leave that cell EMPTY. We will develop a protocol for adding new phenotypes as allowable in T3. All phenotypes should have measurement protocols that TCAP scientists agree to adhere to.