IMPORTING HISTORICAL DATA INTO BMS

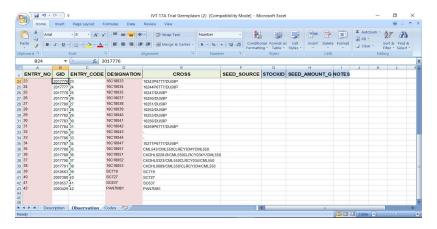
DATA IMPORT TOOL allows users to upload historical datasets generated from other software into Breeding Management System (BMS). This can be Nursery or Trial field data.

Contents

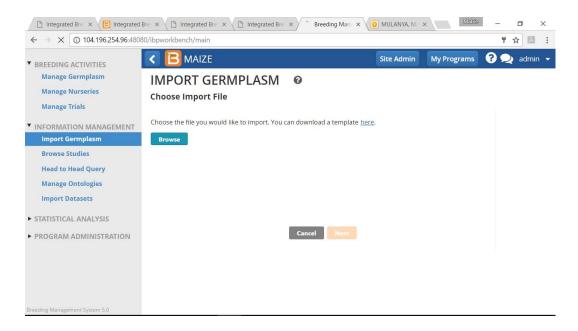
Obtaining GIDs to insert into the Data File	2
Import Germplasm List	2
Export germplasm list to retrieve GIDs	5
Obtaining Location_IDs to insert into the Data File	6
Obtaining Person_IDs to insert into the Data File	7
Format the Historical Data	8
Data import from an excel file using the excel wizard	8
Structure of an Excel File for data import	8
Add the data for factors and variates	8
Run the Data Import Application	9
Map data columns to ontology variables	10
Import the data values	11
Data import from a Fieldbook formatted file	13
Import the Germplasm List	13
Obtain a Fieldbook Template	13
Format the file for a single location	13
Import the File using the Fieldbook Wizard	15
Loading data for multiple locations in the same fieldbook file	17
Using the VLOOKUP function to copy GIDs into an excel file	19

Obtaining GIDs to insert into the Data File Import Germplasm List

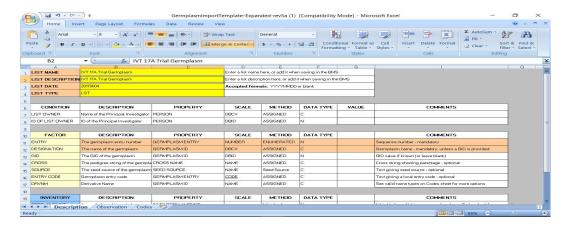
- Before importing any field data it's important to note that the **germplasm used in the trial or nursery must be allocated with GIDs** (Germplasm Identifiers) and these must appear in the data file. If the germplasm does not have GIDs, full BMS database integration requires that you first import the germplasm list into the database and assign GIDs to all germplasm and then enter those GIDs in the data file before importing data.
- Sort historical data by ENTRY NO per replication to identify germplasm used in the trial or nursery. **NOTE:** You only need germplasm entries of one replication to make a germplasm list to avoid duplicated GIDs. In this example we have 42 germplasm entries which have been replicated twice, therefore the total number of entries in the germplasm list will be 42 (Refer to IVT 17A Trial Germplasm list.xls).



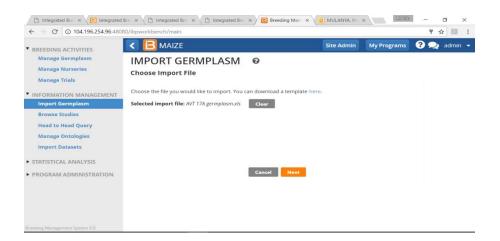
• You need a BMS germplasm template to import any germplasm. Download BMS template from IMPORT GERMPLASM under INFORMATION MANAGEMENT section.



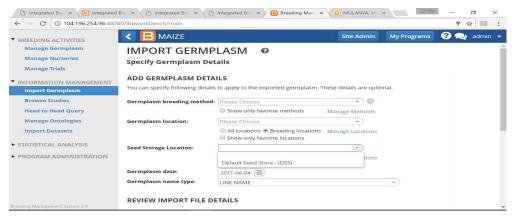
- The germplasm template (.xls) has three sheets:
- Description-Has meta data about the list
- **Observation** Has germplasm details where ENTRY NO and DESISGNATION (germplasm name) are mandatory
- **Codes** Provides optional column headers to customize germplasm details e.g the DVRNM appearing in the column header of observation sheet refer to derivative name in codes sheet.
- In Description sheet, fill the details of germplasm name, description and the date the list was created.



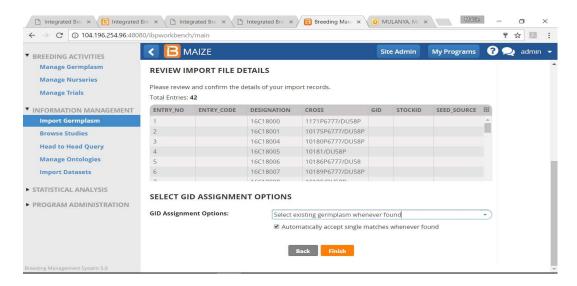
- When you have completed filling in details in the germplasm list, **SAVE** and return to BMS.
- Back to the BMS>Information management>Import germplasm. Browse to germplasm import file(.xls), upload the **GERMPLASM LIST** (Example file:*AVT 17A Germplasm.xls*) and select **NEXT**.



Fill in the appropriate details concerning the germplasm list using the dropdown arrows.

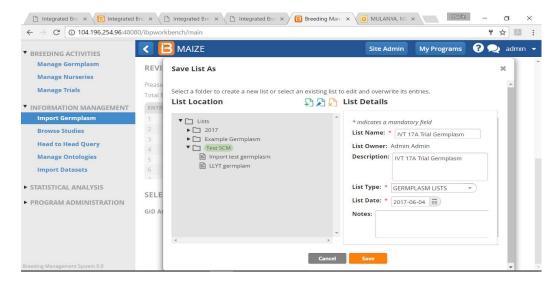


- There are two options when assigning GIDs:
- Assigning new GIDS-adding all entries to new records and assigning new GIDs
- Matching existing GIDs- Selecting existing germplasm whenever found to match Germplasm names in the import list to the ones existing in the database. It's advised to use this option to avoid duplication of GIDs.
- Select existing germplasm whenever found option and FINISH



• In case the germplasm in the import list matches with more than one in the database then you will have to select for the appropriate name from the existing list and select for the match to be used for other instances found in the import list.

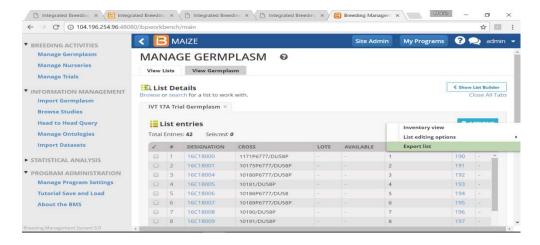
• Select continue, highlight the folder you will use and **SAVE** the list.



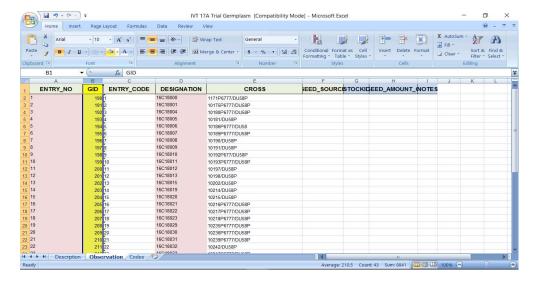
Export germplasm list to retrieve GIDs

The imported germplasm list must be exported from BMS to retrieve the GIDS.

- Select Manage lists from the menu. Browse to the imported germplasm list and open it. Close the selector window to view the list.
- From the actions menu select export list and You **MUST LOCK THE LIST** with the padlock icon appearing under the **ACTIONS** before you can export it.

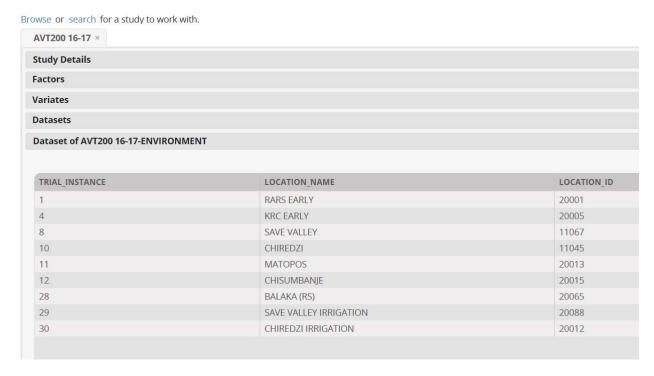


From the exported germplasm list, Select the observation sheet to view the GIDs.



Obtaining Location_IDs to insert into the Data File

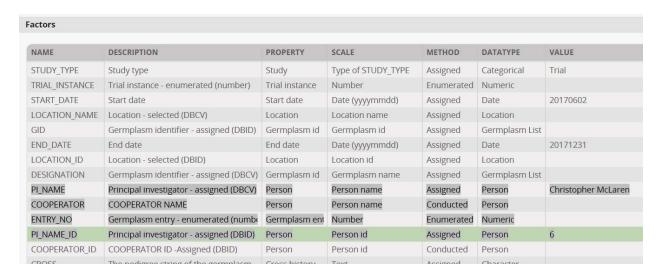
It is important to have LOCATION_NAMEs and IDs in the data import files if the data is to be correctly assigned to environments. However the current Data Import tool does no look up location IDs so you need to find the IDs by looking for the locations on the Locations tab of Manage Program Settings or by setting up a template trial with all the locations you intend to use (you may have to add some new ones) and then looking in the ENVIRONMENTS dataset of the template using the Browse Studies feature from the Information Management menu:



For frequent use it is desirable to maintain a file of Location Names and IDs.

Obtaining Person_IDs to insert into the Data File

You may want to specify the PI_NAME (if using Fieldbook Wizard) or the COOPERATOR for each location if using either import method. To load this data you need to enter the persons name exactly as in the BMS and also the persons ID (in variables PI_NAME_ID or COOPERATOR_ID). To find these IDs you need to create a template study (not using the name of the study you intend to import) and add the person you want to find the ID of as PI or COOPERATOR and use Browse Studies to look in the FACTORS section for the PI name and ID:



Or in the ENVIRONMENT dataset to fine the name and ID of COOPERATORS

TRIAL_INSTANCE	LOCATION_NAME	LOCATION_ID	COOPERATOR	COOPERATOR_ID	CROP_SEA
1	RARS EARLY	20001	Christopher McLaren	6	2017 B se
3	RARS LATE	20003	Christopher McLaren	6	2017 B se
4	KRC EARLY	20005	Christopher McLaren	6	2018 A se
8	ART MEDIUM	20010	Christopher McLaren	6	2017 B se
15	LUSAKA WEST	11530	Christopher McLaren	6	2017 B se
22	GART	20082	Christopher McLaren	6	2017 B se
23	KAFUE	20078	Christopher McLaren	6	2018 A se

It may be useful to retain a file of the persons and their IDs for easy reference.

Format the Historical Data

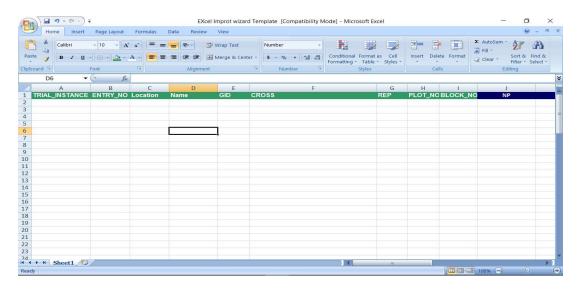
There are TWO options of file formats which can be used for historical data import into BMS:

- 1. Use of Excel import wizard format is the most flexible format for importing historical data. It requires one sheet containing factors and variates for the dataset to be imported. There are FOUR mandatory columns: TRIAL_INSTANCE, ENTRY_NO, PLOT_NO and GID to import phenotypic observations. However, it is desirable to import LOCATION_NAME and LOCATION_ID as well. You can also include additional descriptive data such as environment codes, season codes, entry names, and design factors. While this is the most flexible format it requires manual mapping of factors and traits.
- 2. Use of excel Field book format (Fieldbook format file.xls). This format applies to files which have been previously generated from BMS e.g. trials or nurseries and exported out. Such files are perfectly formatted for historical data import into BMS. This format does not require manual mapping and so if you have many similar trials it is a very efficient way to import the data.

Data import from an excel file using the excel wizard.

Structure of an Excel File for data import

• The Excel import wizard format for data import reads one file at a time. The column headers highlighted in green are FACTORS and those in blue are VARIATES. There are FOUR mandatory columns: TRIAL_INSTANCE, ENTRY_NO, PLOT_NO and GID. The colors are just for illustration, and any other column headers can be used for other labels and traits.



Add the data for factors and variates

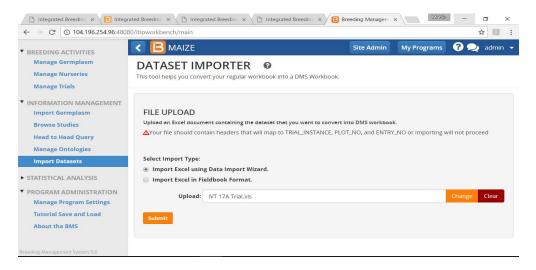
- Copy and paste the information in your trial into the **EXCEL IMPORT WIZARD TEMPLATE** as shown below. The TRIAL_INSTANCE column should contain distinct integers identifying distinct trial sites or environments. The optional column Location could contain location names (repeated for each plot at that location).
- The Name column in the example above could contain the germplasm name or DESIGNATION and mapping the columns will be easier if the column is in fact headed DESIGNATION. The CROSS

column could contain a cross string for each entry. Name, GID and CROSS can all be filled from the germplasm list file using the VLOOKUP function as described in the last section of this file. IF the CROSS column is present, no cells can be empty – just add "-", this is a known bug.

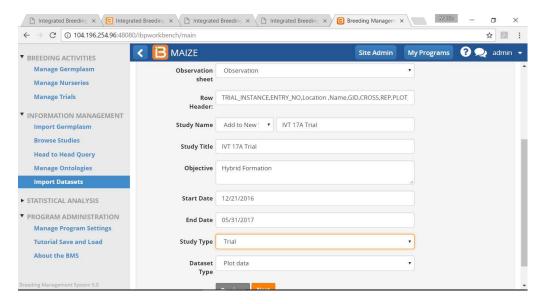
- In the column header named **GID**, copy the GIDs from the germplasm list that was exported from BMS. You can use the Excel VLOOKUP function as described in the last section of this file. Ensure that the **ENTRY NO** or **DESIGNATION** (name of germplasm) that appear in the Germplasm list match exactly those of the trial.
- PLOT_NO must just contain integers 1,2 ...n where n is the number of plots in each trial instance, for each trial instance. If the Trial design is different from the template format then you will need to edit the factors in the headers. If you have an incomplete block design be sure to use the header BLOCK_NO for the sub-block factor (and not BLK or BLK_NO) this is a known bug.
- If you intend to add **LOCATION_NAME** and **LOCATION_ID** (which you should), you must add those columns to the data sheet and fill them with the values from your locations file or from the template view (as shown above). Each plot for each location must have the same value of name and ID.

Run the Data Import Application

• Launch the **INFORMATION MANAGEMENT** from BMS then select **DATA IMPORT TOOL**. Select the import **EXCEL USING DATA IMPORT WIZARD** and select an excel file to import and **SUBMIT**.

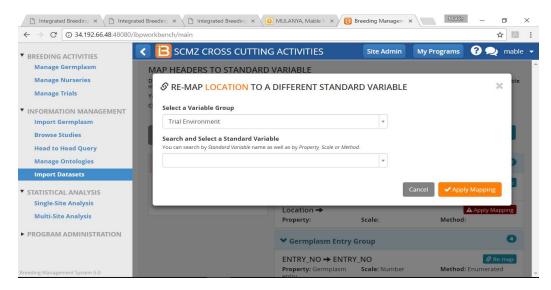


• Specify the observation sheet containing phenotypic data within the excel file. Select the appropriate column header and close. Enter study name, title and type.

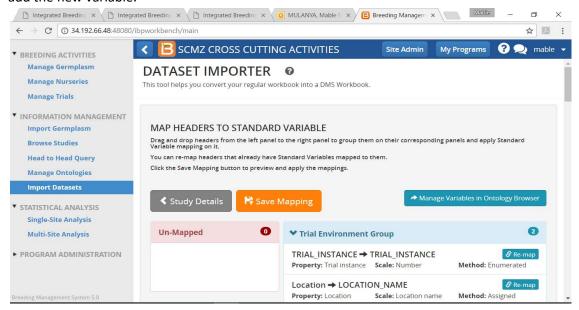


Map data columns to ontology variables

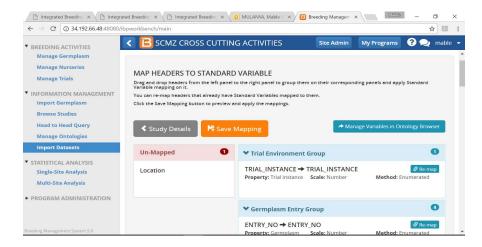
- The BMS will attempt to map the excel sheet column headers to corresponding database terms. Carefully review the mapped terms to ensure that the BMS correctly matched the truly meaning and measurement of the imported data.
- Review the details of the suggested matches by selecting re-map icon and Search for alternative matches by typing in the appropriate field (either TRIAL ENVIRONMENT, GERMPLASM,TRIAL DESIGN or VARIATE GROUP)
- In this example the term **LOCATION** is mapped under the **TRIAL ENVIRONMENT GROUP**. IF you have added LOCATION_NAME and LOCATION_ID instead they should go to the ENVIRONEMNT group. Similarly Season_Code or management codes should go into this group.



• For the UN-MAPPED TERMS, drag and drop them to their appropriate grouping either (TRIAL ENVIRONMENT, GERMPLASM,TRIAL DESIGN or VARIATE GROUP). If the alternative term is an actual match , select apply mapping. If a match is not found, you will have to go back to ontology manager to add the new variable.

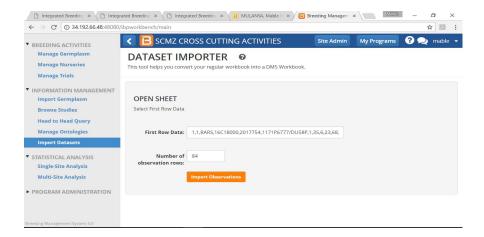


• When mapping is complete, save it and select confirm header mapping.

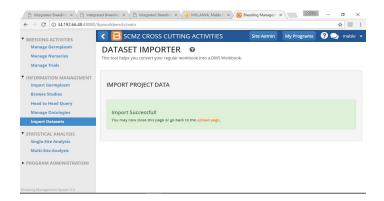


Import the data values

• Select the first row of the data and import the observations. The phenotypic observations for this trial will be saved in the database and you will receive a message for successful importation.



• You will receive a notification if import is successful.



You can retrieve the trial by selecting Manage Trials from the MENU and BROWSE for the Trial!

Data import from a Fieldbook formatted file

Import the Germplasm List

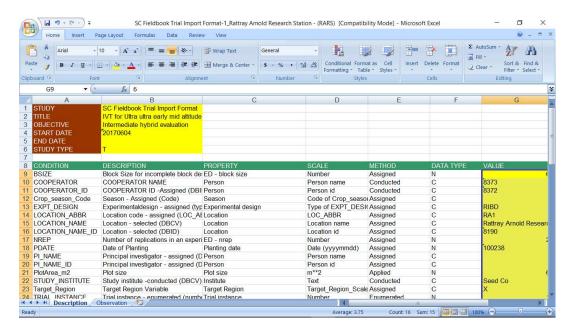
- You have to import a germplasm list into BMS to generate GIDs as indicated above.
- Export the list to get the GIDs to insert into the Fieldbook

Obtain a Fieldbook Template

• The fieldbook format option expects the data to be delivered in a file formatted according to the BMS fieldbook format like files exported from trials already in BMS. The easiest way to obtain such a file is to create a template trial which matches the trial you wish to import in terms of variables – factors and traits. This template should not have the same name as the trial you want to import. Then export the fieldbook from the template as though you were going to collect the data. The fieldbook format has two sheets: Description and Observation sheet.

Format the file for a single location

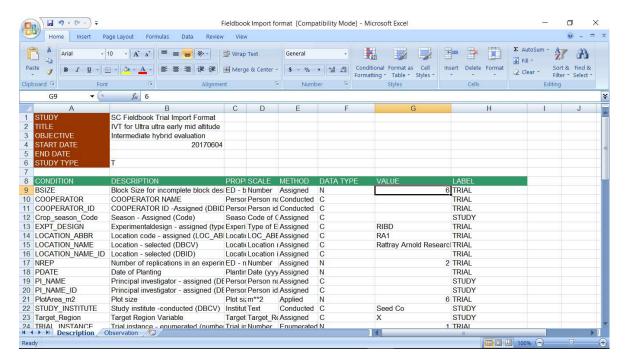
- Set the Study Name, Title, Objective, Start Date, End Date. The Study Type MUST be set to T for Trial. These values remain the same for subsequent files for each location. Edit the value fields in the CONDITION and CONSTANT sections on the **Description sheet** to reflect the information for the environment you are loading. In particular you can enter the LOCATION_NAME and LOCATION_ID in the CONSTANT section.
- You must delete the row describing PLOT_ID from the description sheet, and the corresponding column from the observation sheet. You can similarly delete rows and corresponding columns for any variables for which you do not have data.



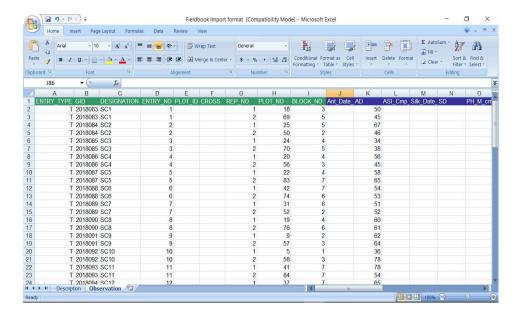
• Copy the GIDS using **VLOOKUP FUNCTION** from the germplasm list into a column header **GID** in the trial observation sheet as described in the last section of this file. You can also copy DESIGNATION and CROSS in this way if required.

• Copy the rest of the data from the historical data file into the observation sheet of the fieldbook. Save the file and review the trial to confirm that the mandatory column headers (TRIAL_INSTANCE, ENTRY_NO, PLOT_NO and GIDs) are present. A complete fieldbook formatted trial with the two sheets is as below

Description sheet

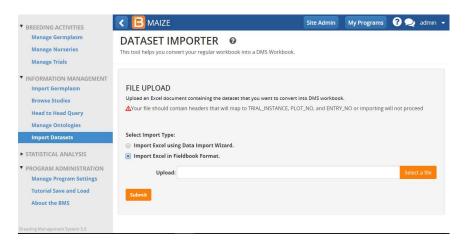


Observation Sheet

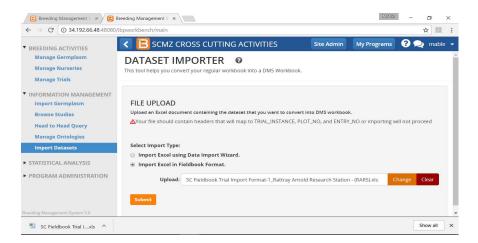


Import the File using the Fieldbook Wizard

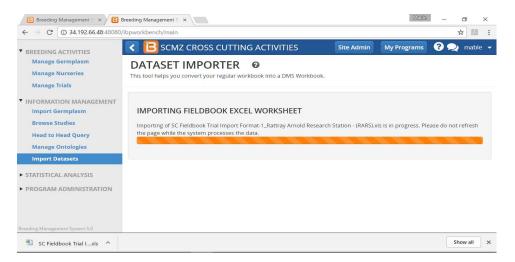
• In the BMS, select **IMPORT DATASETS** from the INFORMATION MANAGEMENT. Select, **IMPORT EXCEL IN FIELDBOOK FORMAT.**



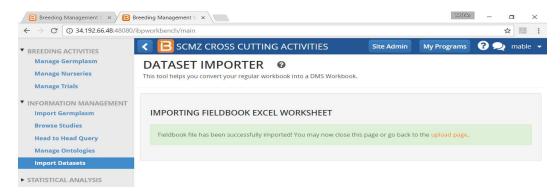
• Select the fieldbook formatted file and upload it.



• BMS will automatically map the terms in the fieldbook formatted trial with the database.

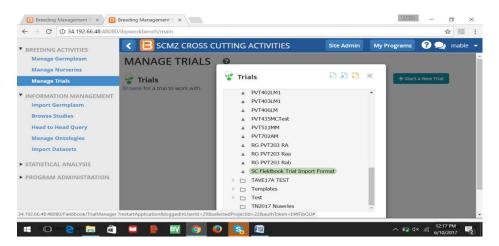


• When import is successful you will receive a notification.

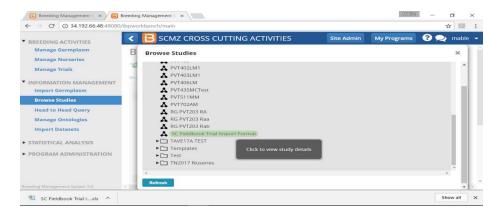


Retrieve dataset

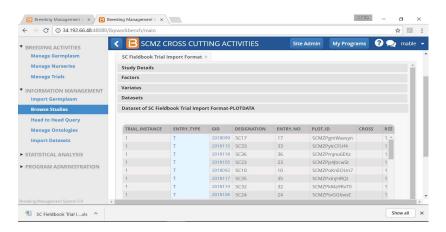
- You can add subsequent locations to the same trial by editing the CONDITION and CONSTANT sections of the Description sheet and pasting the GIDs and data for the next location in the observation sheet. Leave the Study name and other Study level variables unchanged.
- Retrieve the fieldbook formatted trial from the MANAGE TRIALS and BROWSE to select the file.



You can also retrieve the trial from the BROWSE STUDIES option under the INFORMATION
MANAGEMENT and select the trial.



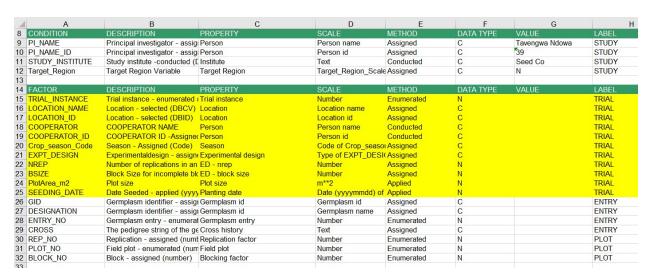
Close the Browse studies window and select DATASETS and then PLOT DATA to view the Trial.



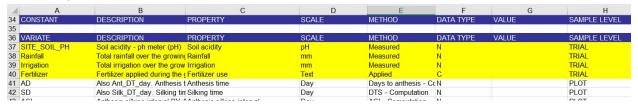
Loading data for multiple locations in the same fieldbook file

You can load data for multiple locations in the same fieldbook file but you need to change the structure of the template.

• Move rows describing TRIAL labels (as indicated in column H) from the CONDITION section of the Description sheet to the FACTOR section. (Leave only rows describing STUDY conditions)



• Move rows describing TRIAL variables (as indicated in column H) from the CONSTANT section of the Description sheet to the VARIATE section.

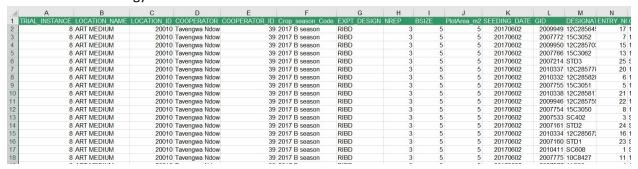


• Insert columns in the observation sheet and enter the names of the new variables you have just added to the FACTOR and VARIATE sections into the first row. The variable names must be exactly as

they appear in the description sheet and in the same order they appear in the CONDITION and VARIATE sections of that sheet.



- Append the data for the concatenated set of locations to the Observation sheet as for the Excel import mode described above. Take care that the location CONDITIONS and CONSTANTS contain the same value for every plot at that location.
- Note I had some trouble with Crop_Season_code where I had to enter the meaning of the code (2017 B season) rather than the code itself (17B). The same is true for Target Region which is a STUDY condition on the description sheet. You can look up the meanings of the valid values in the scales section of the Ontology Browser.

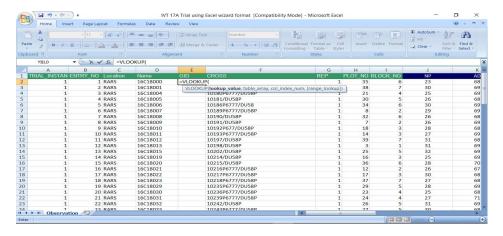


and

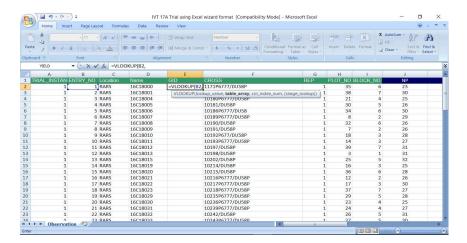
R	S	т	U	V	W	X	Υ	Z	AA	AB	AC
	SITE SOIL PH	Rainfall	Irrigation	Fertilizer	AD	SD	ASI	NP	PH	EH	EPR
1	2.66	108	100	40	69	70	1	47	2.35	1.35	0.574468
1	2.66	108	100	40	70	69	-1	46	2.25	1.1	0.488889
1	2.66	108	100	40	73	74	1	48	2.05	1.2	0.585366
1	2.66	108	100	40	68	69	1	48	2.45	1.35	0.55102
1	2.66	108	100	40	70	71	1	47	2.5	1.54	0.616
2	2.66	108	100	40	70	71	1	48	2.71	1.45	0.535055
2	2.66	108	100	40	71	73	2	46	2.6	1.4	0.538462
2	2.66	108	100	40	67	68	1	48	2.5	1.15	0.46
2	2.66	108	100	40	74	76	2	46	2.6	1.4	0.538462
2	2.66	108	100	40	69	70	1	48	2.4	1.3	0.541667
3	2.66	108	100	40	59	61	2	48	2.38	0.95	0.39916
3	2.66	108	100	40	57	58	1	46	2.3	1.1	0.478261
3	2.66	108	100	40	68	69	1	48	2.48	1.3	0.524194
3	2.66	108	100	40	73	72	-1	47	2.3	1.21	0.526087
3	2.66	108	100	40	70	71	1	46	2.54	1.32	0.519685
4	2.66	108	100	40	67	68	1	48	2.46	1.2	0.487805
4	2.66	108	100	40	70	71	1	48	2.63	1.3	0.494297
4	2.66	108	100	40	69	70	1	48	2.36	1.16	0.491525
4	2.66	108	100	40	69	70	1	48	2.48	1.29	0.520161
4	2 66	108	100	40	73	74	1	48	2 65	1.5	0.566038

Using the VLOOKUP function to copy GIDs into an excel file.

• Open the formatted Trial (either the simple excel file or a Fieldbook file) and insert the VLOOKUP formula into the first cell just below the GID column header on the sheet containing the observations. In our case this will be cell E2. To insert the formula, type =VLOOKUP(in cell GID column. In this example, the VLOOKUP formula will be in will be cell E2.



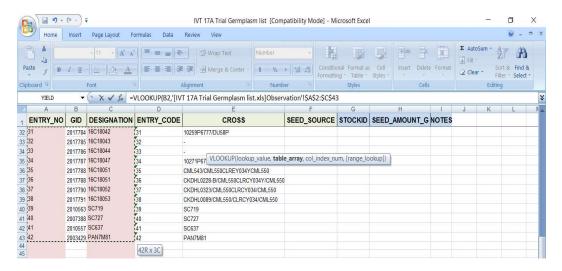
- The VLOOKUP formula has 4 elements;
 - a. Lookup value- the value you want to look for
 - b. **Table Array** range where your lookup value is located (the lookup value must be in the first column of this range)
 - c. **Col Index Num** The column number in the range containing the return value (relative to the first or look-up column)
 - d. **Range Lookup** matching choice you want (can be either TRUE meaning it's an approximate match or FALSE meaning it's an exact match)
- In this example, the lookup value is Entry NO.1 which is in cell B2. We will use Entry No. because is the common column header in the two sheets.



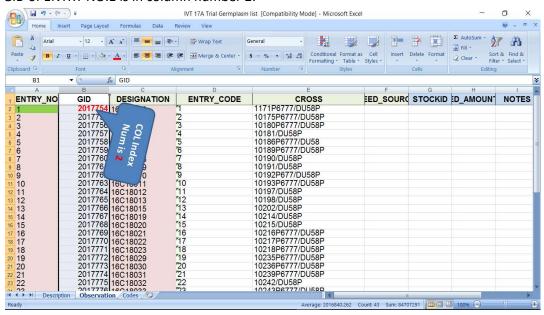
■ Enter the Table array [IVT 17A Trial Germplasm list.xls] Observation'!\$A\$2:\$C\$43] which is the range including our LOOKUP VALUE and the GID column we need to copy. In this case the Table Array is

to come from the Germplams List. (You can also copy the lookup array from the other file into another sheet or even the same sheet of the destination file before you set this range if you prefer.)

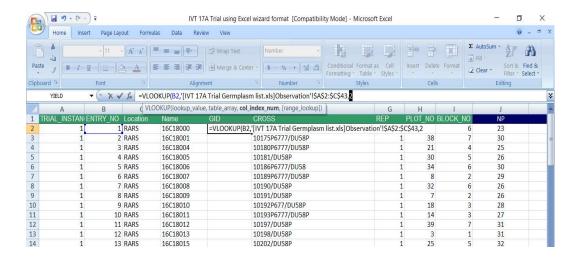
If your excel Version does not automatically generate dollar signs in the formula as (\$A\$2:\$C\$43) then you will have to add dollar signs (\$) to the table array address so that A2:C43 becomes \$A\$2:\$C\$43. Putting the dollar signs in the table array makes the cell references absolute so that when you copy the formula down the GID column, the VLOOKUP function will still be looking at the same table (Germplasm list) you referred to.



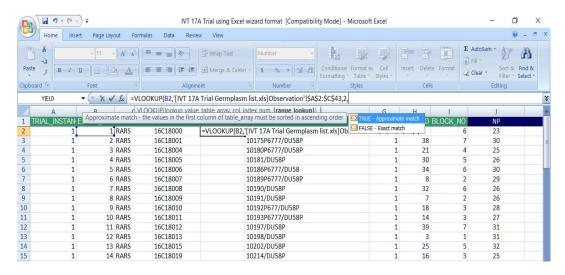
■ Enter the Col Index Num. This is the column number in the range (Table array of the germplasm list) containing the actual value you want to copy in relation to the Lookup value. In this case, ENTRY NO 1 is our lookup value and we want to copy GID for ENTRY NO 1. We check in our germplasm list and the GID of ENTRY NO.1 is in column number 2.



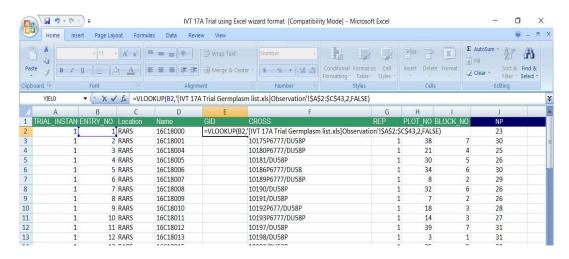
We then type 2 as our Col Index Num value.



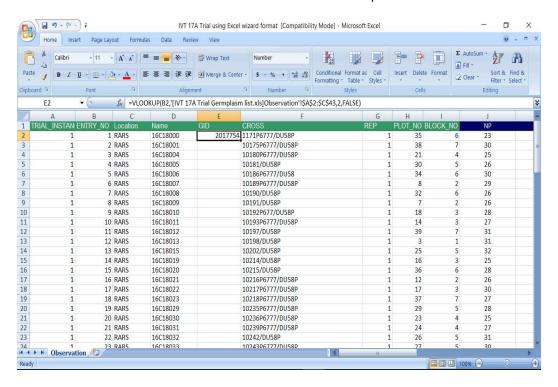
• Next, enter the RANGE LOOKUP which is the matching option. An approximate match is TRUE and Exact match is FALSE. We want an exact match so we type FALSE in the formula.



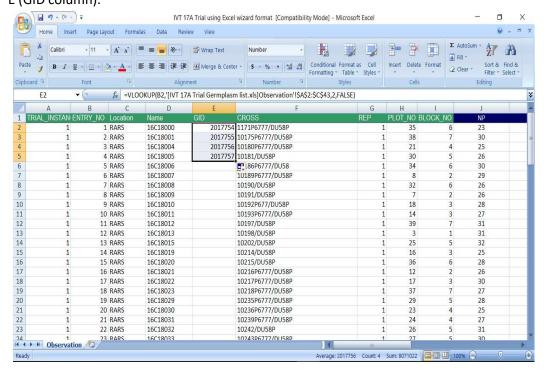
Finally type the CLOSING BRACKET and ENTER.



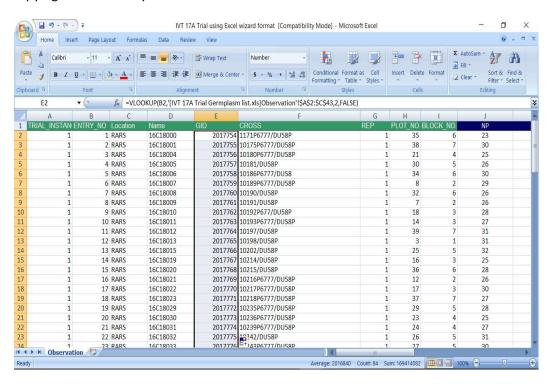
We notice that the GID for ENTRY NO.1 has been copied in the Trial.

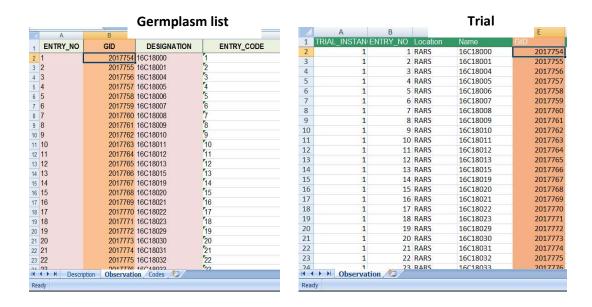


• To copy GIDs in the whole column, use the drag down tool in the bottom right corner of the cell E (GID column).

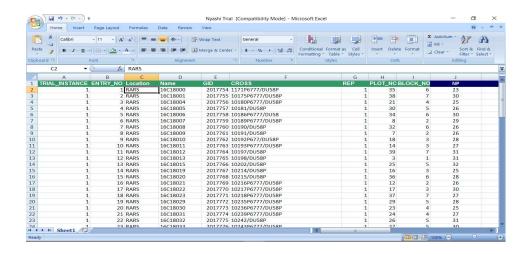


• The GIDs have now been copied and you need to cross check the two lists to confirm if the copying was correctly done.





- Once you have checked that the correct GIDs are going to the correct entry numbers you must copy the whole GID column in the template and paste it as values you cannot import a file with formulas in the cells. If you try you will ger a strange message that he Inport File must contain GIDs
- This is a Complete formatted Trial ready for import into BMS.



Learn more about BMS from the Tutorials at https://www.integratedbreeding.net