Appendix B: CSV Import File Formats for KDSmart version 2.0.45

B.Pearce 2016-04-21

Refer to http://www.kddart.org/kdsmart/ for the most recent documentation and other information.

Trial and Plots

During CSV import of a *Trial* you will be presented with the column headings found in the file and asked to assign an *Attribute Type* to each column heading.

The following tables list each *Attribute Type*. Pay particular attention to the ones marked in blue:

- *Trial Attribute*
- *Plot Attribute*
- *Trait*
- *Trait for Specimen*
- *Ignore*

The *Attribute Types* listed above may be applied to any heading in the CSV file.

The *Trial Attribute* and *Plot Attribute* types indicate that you wish *KDSmart* to retain the value as an attribute either for the *Trial* or for each *Plot* respectively. Applying *Trait* or *Trait for Specimen* to a column will cause a *Trait* of that name to be entered into the list of *Traits* for the *Trial*. Use *Ignore* to ignore columns in the CSV file.

Any columns with a name starting with "**Link:**", "**Date:**", "**Date_**", "**Specimen#:**", or "**Specimen#_**" will be automatically assigned as *Ignore* (but you can always change that if you wish).

Trial & Plot Import Options

When importing a heading classified as *Trait* or *Trait for Specimen* *KDSmart* will use hints in the heading to determine whether it is only a *Trait Name* or if the heading also identifies a *Trait Instance*.

This is designated as the *Trait Name Style*.

You may choose one of the following options:

Option	"Base" instance number	Format	Example / Description
No Trait Instance detection			All Trait headings will have a single Trait Instance created. Headings of AMT1, AMT:1, AMT__1 will correspond to three different Traits with exactly the names provided.
Ends in colon followed by digits	1	<traitName> : <instanceNumber>	AMT, AMT:2 results in Trait Instance numbers of 1 and 2.
Ends in two underscores then digits	1	<traitName> _ _ <instanceNumber>	AMT, AMT__2 results in Trait Instance numbers of 1 and 2.

The other options you need to select are how *KDSmart* determines *Specimens* from the CSV headings. This is described in another section below (**Specimen Count and Trait Instances**).

You may specify whether or not "sub-plot" data is to be collected here or while scoring by touching the *Plot* icon:

The following three tables list, in turn, the headings that are used for different parts of a Trial Import.

The first table lists the headings that pertain to *Trial*-specific data.

Attribute Type	Section	Headings Automatically Recognised	Description
Trial Planting Date OPT	Trial	PlantingDate, TrialPlantingDate, DatePlanted, Planted, DateSown, Sown, SownDate	This is the *base* value used for computing *ELAPSED_DAYS* values for Traits (e.g. *Days to Flowering*).
Trial Name OPT	Trial	TrialName	If present, this is used as the name of the trial. If not present, the name of the CSV file will be used.
Trial Alias OPT	Trial	Abbreviation, TrialAbbreviation, TrialAcronym, Acronym	A short name for the Trial that will be used on screens where there is limited space.
Trial Column Name OPT	Trial	ColumnName, NameForColumn	This is the word used for the X coordinate (see the attribute type *X-Column* in the *Plot headings* table below). Example: for "Range/Row" the value here would be "Range".
Trial Row Name OPT	Trial	RowName, NameForRow	This is the word used for the Y coordinate (see the attribute type *Y-Row* in the *Plot headings* table). Example: for "Range/Row" the value here would be "Range".
Trial Plot Name OPT	Trial	CellName, NameForCell, PlotName, NameForPlot	This is the word used to describe each "cell". For many plant-related field trials this is likely to be "Plot". For other trials/experiments it may be "Panel", "Pond", "Test-tube" etc.
Trial Database Id OPT	Trial	DatabaseTrialId, TrialId, TrialNumber	May be provided to cross-reference to the originating database.
Trial Attribute ATTR	Trial	May apply to any heading in the file. See **Trial Attribute Headings** below	Retain the value as an attribute of the *Trial*. If you apply this type to a column then every row must have the same value.

This continues the table above but lists the headings that are related to per-Plot data.

Attribute Type		Section	Headings Automatically Recognised	Description
Plot Id	ID	Plot	CellId, PlotId, EntryId, Plot	If present, uniquely identifies the Plot in the Trial.
Plot Column (X)	ID	Plot	PlotColumn, X, Column, Col, ColumnX, X-Column, Column_No	The X–coordinate of the Plot within the Trial. If present, the (X,Y) pair must uniquely identify the Plot in the Trial.
Plot Row (Y)	ID	Plot	PlotRow, Y, Row, RowY, Y-Row, Range_No, Range	The Y–coordinate of the Plot within the Trial. If present, the (X,Y) pair must uniquely identify the Plot in the Trial.
Plot Note	OPT	Plot	PlotNote, Note, TrialUnitComment	
Plot Type	OPT	Plot	PlotType	If present, this specifies the type of plot (e.g. Control, Check, etc.).
Plot Barcode	OPT	Plot	PlotBarcode, Barcode, TrialUnitBarcode,	See the User Guide section on Barcode Scanning for further details of how the data in this column is used.
Plot Tags	OPT	Plot	Tags, PlotComments, Comments	A list of pipe-separated (**|**) *Tag* labels.
Plot Database Id	OPT	Plot	DatabasePlotId, TrialUnitId, TrialUnitNumber	May be provided to cross-reference to the originating database.
Specimen Count	OPT	Plot	SpecimenCount, #Specimens	Please read the section below regarding ***Specimen Count***.
Plot Attribute	ATTR	Plot	May be applied to any heading in CSV. See **Plot Attribute Headings** below.	Retain the value as an attribute of the *Plot*. Each row may have a different value (including blanks). Choose any subset to display as *Plot Info* while scoring.

This table continues from the two above but lists the Attribute Types for Trait variations and for excluding a column from import.

Attribute Type	Section	Headings Automatically Recognised	Description
Trait OPT	Trait	May be applied to any heading in the CSV file.	Marks this column as representing a Trait that is scored **only for Plots**. The *Trait Name Style* governs if a *Trait Instance* is being referenced or not.
Trait for Specimen OPT	Trait	May be applied to any heading in the CSV file.	Marks this column as representing a Trait that is scored **only for Specimens.** The *Trait Name Style* governs if a *Trait Instance* is being referenced or not.
Ignore		May be applied to any heading in the CSV file.	Indicates that this column should **not** be imported.

A simple example follows on the next page.

Specimen Count and Trait Instances

If present and non-blank, the integer value in this column will create the given number of specimens in the Plot.

This will apply to all of the columns that have an *Attribute Type* of Trait or Trait for Specimen.

By using the Specimen Count column, you are indicating that you wish *KDSmart* to create that many specimens *in each plot.*

If you do NOT mark a column as Specimen Count then no specimens will be created during the import of the data file; in this case, if you wish to

add a specimen to a plot, use the popup menu that appears when you touch the Plot icon: (during the scoring activity) and choose the **Add Specimen** option. You will then be presented with a list of all the *Traits* that are defined as Trait for Specimen and may select which of these are to be scored for the new *Specimen*.

Plot Attribute Headings

The items below identify headings that are automatically classified as *Plot Attributes* by *KDSmart*.

Please note that this list will be replaced by the *CSV Import Profile* functionality at a later release.

- `Treatment`
- `SelectionHistory, Selection_History`
- `ReplicateNumber, Replicate, Rep, Repeticion, REP_NO`
- `GenotypeName, Genotype, GID`
- `Origin, 0rigin` (that's a "zero")
- `Designation`
- `Type`
- `CID, SID, CROSS, SOURCE, ENTRY_NO, SUB_BLOCK, PLOT_NO`
- `PedigreeName, Pedigree, BreedersPedigree`
- `Type`

Trial Attribute Headings

The following are automatically classified as *Plot Attributes* by *KDSmart*:

- `SiteName, Site, Location`
- `SiteYear, Year`
- `TrialType, TrialTypeName`
- `TrialStartDate, StartDate`

Values in CSV files

CSV files may contain data in the various data "columns".

For columns marked to be imported as *Traits*, the following rules apply:

1. Blank values will be treated as un-scored Traits as will the word `UNSCORED` (the latter is not required unless you want to make un-scored trait values very obvious in a CSV file).
2. The word `MISSING` will be imported as a special token denoting a missing value.
3. The word `NA` will be imported as a special token denoting "not-applicable" or "not-available" (whatever you take it to mean).

All other values will be checked using the validation rule for the Trait. Any failure to pass the validation check will cause the entire import to fail.

If you select a column to be used as the *Trial Name*, the value in the first data line may not be the same as any existing *Trial* in the database.

Similarly, if you have not chosen a column as the *Trial Name*, the name of the file (excluding anything from the last "." onwards) will be used to check for a pre-existing trial in the database.

These *Trial Name* checks are done in a case-insensitive manner.

A Simple Example

A file with three lines consisting of:

```
date sown,Row,Range,MZC,MZP,#Specimens,LLEN
2014-12-15,23,4,,,2,
2014-12-15,21,5,,,0,
```

Using the "After Specimen Count" described in the previous page, the following definition could be used in *KDSmart*:
(note that shaded row in this table is NOT part of the data file):

Attribute Type	PlantingDate	X-Column	Y-Row	Trait	Trait	SpecimenCount	Trait for Specimen
CSV Line 1	date sown	Row	Range	MZC	MZP	#Specimens	LLEN
CSV Line 2	2014-12-15	23	4			2	
CSV Line 3	2014-12-15	21	5			0	

This diminutive Trial has only two plots; they are at positions (23,4) and (21,5).

The *Traits* MZC and MZP are being scored for each *Plot*.

The first *Plot* has two specimens for which the trait LLEN (leaf length) will be scored.

The second *Plot* has no *Specimens* that require scoring.

If, however, while Scoring you were to add a *Specimen* to the latter *Plot*, you will then be asked to score the LLEN *Trait* for the new *Specimen*.

Additionally, because there were no columns designated as *X-Column-Name* nor *Y-Row-Name*, the default names of "Column" and "Row" will be used to describe the coordinates of the *Plots*.

If desired, you could edit the *Trial Details* and change the names to "Row" and "Range" respectively to make the nomenclature align more closely if that is your usual vocabulary.

Spaces, Case Sensitivity and Blank Lines

When reading and parsing the headers, *KDSmart* removes all spaces and performs a case-insensitive comparison to match CSV headings with the automatically recognised built-in headings.

It also skips over any blank lines and takes the first non-blank line as the headings line.

Trait

The following details the required and optional headings in a CSV file that contains details of *Traits* that will be used by *Trials*.

Heading	Alternates	Notes
TraitName	Name	This column heading **must** be present. All others are optional. The value should be kept short and, if you wish to use the name in *KDXplore* with CALC *Traits* no spaces are permitted and the name must begin with a letter. We recommend using the underscore (_) character or CamelCase to improve readability. See also the description of the import option regarding *Trait Name Style*.
TraitAlias OPT	Alias	If provided, should be a *shorter* form of the *TraitName* to use during scoring.
TraitBarcode OPT	Barcode	Used to identify the Trait when using a barcode scanner for scoring.
TraitUnit	Unit, UnitName, TraitUnitName	You can provide anything you like that will assist you in remembering what to enter while Scoring.
TraitDescription	Description, Desc	A longer description of the Trait.
TraitDatatype	Datatype	One of CATEGORICAL, DATE, ELAPSED_DAYS, INTEGER, DECIMAL or TEXT. See *Appendix A* for further details.
TraitValidation	TraitValRule, Validation, ValidationRule	See *Appendix A* for further details.
DatabaseTraitId	TraitId	If you provide this column, *KDSmart* will record the value and export it on request (a reference to the originating database). See *Protected Traits* below.

Note that you can also create a new *Trait* directly in *KDSmart*. If you do this and are also uploading or synchronising your data to *KDXplore* or *KDDart* you may have to reconcile the differences if there is already a *Trait* of the same name (compared in a case-insensitive manner).

Overwriting existing Traits

You may provide entries for existing *Traits* in a CSV file. The *Traits* will be matched by trait name (ignoring case) and incoming *Traits* with the same name as one already in the database will be checked for compatibility:

Trait in Database	Trait in CSV file	Action Taken
Data Type is TEXT	Any Data Type	The Trait Data Type in the CSV file will replace the data type in the database.
Any other data type	Data Type is TEXT	A warning is issued identifying the line and the import continues excluding the line from the CSV file.
Any other data type	Data Type is not TEXT	A warning is issued if the data types are not the same (and the import continues) Otherwise: • CHOICE: must have the choices or the old choices must be a subset of the new choices • INTEGER and DECIMAL: the limit "exclusions" must be identical, however the new *Trait's* limits may be "wider" than the current *Trait* in the database If any of incompatibilities above are identified, a warning is issues and the import continues without altering the current *Trait* in the database.
Description	Description	Descriptions are truncated to the current database limit then compared for equality. If they are the same (ignoring case), a warning will be issued but the new *Trait's* description will replace the current one in the database.

Protected Traits

If a *Trait* has been imported with a *DatabaseTraitId*, it is deemed to be "protected".

In this case you may only edit the *Trait Alias* and changes to the validation rule are not permitted except for INTEGER and DECIMAL Traits where you may change the range of acceptable values but only to make the range smaller.

For example, if an INTEGER *Trait* has been defined to have values from 0 to 100 you may only change the lower limit to be 1 or more and the upper limit to be 99 or less.

Calculated Traits

These have the *Trait Data Type* of CALC. No measurements are collected by *KDSmart* for these *Traits*.

The value may only be viewed in *KDXplore* during data curation.

For more details about the syntax of expressions and the mathematical functions supported, please read *KDSmart User Guide, Appendix A (Trait Data Types)*.

Tags

The following details the required and optional headings in a CSV file that contains details of *Tags* that will be used to annotate *Plots* and *Specimens* when Scoring a *Trial*.

Heading	Alternates	Mandatory?
Label	TagLabel, Comment	Yes
Description	TagDescription, Desc, CommentDescription	Yes
DatabaseLabelId	LabelId, Id	No – but useful for cross-reference to the originating database.