data
POOL

Codebook Completion Guide

for DataPool alpha version

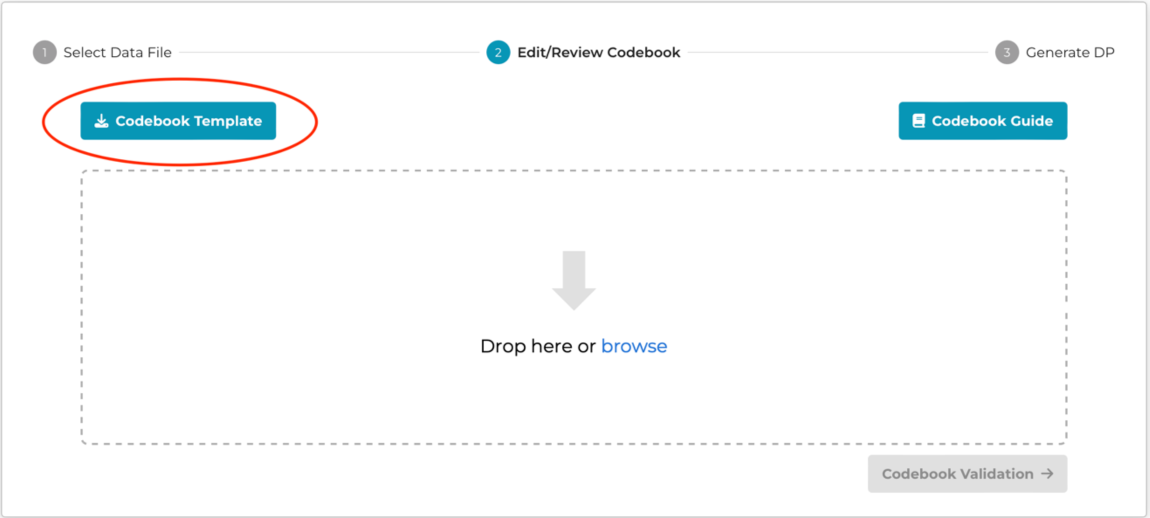
EXCELLENCE IN AGRONOMY

powered by SCiO

What is a DataPool codebook?

The DataPool codebook is a Microsoft Excel™ spreadsheet file, where the contents of a data pool are defined and contextualized. It guides the system to understand the variables included in a datafile and properly manage and align the respective datapoints in the broader DataPool ecosystem.

DataPool users are called to fill in the required information on a codebook template, provided to them when they add a new version of a data pool.

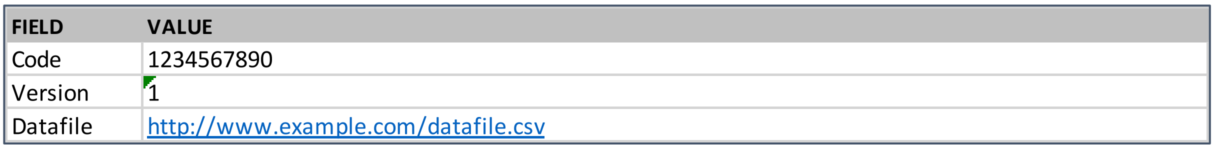


The organization of the codebook template and the actions expected by the users to complete the codebook are described in the following section.

How is a codebook organized and edited?

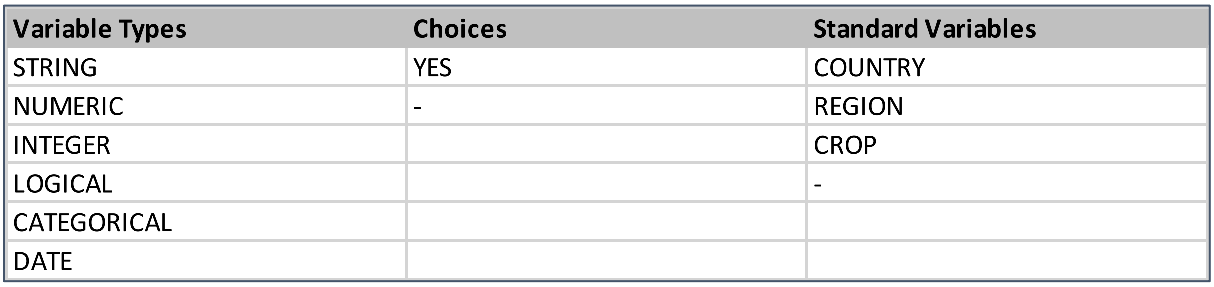
Every DataPool codebook comprises three sheets.

The **metadata** sheet includes basic metadata information for the data pool associated with the data file covered by the codebook.



FIELD	VALUE
Code	1234567890
Version	1
Datafile	http://www.example.com/datafile.csv

The **lists** sheet defines all controlled input lists used in the main part of the codebook, the variables sheet.

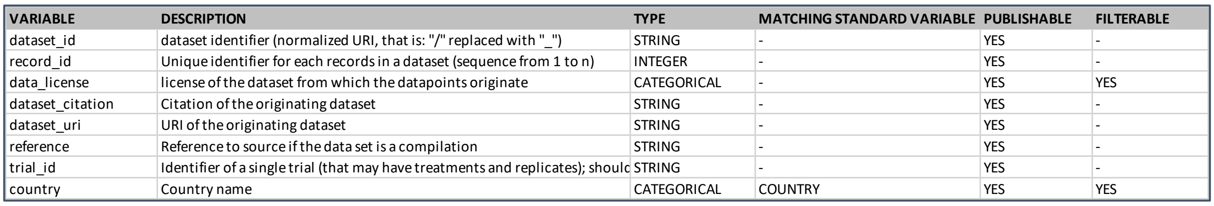


Variable Types	Choices	Standard Variables
STRING	YES	COUNTRY
NUMERIC	-	REGION
INTEGER		CROP
LOGICAL		-
CATEGORICAL		
DATE		

The variables sheet is where the magic starts to happen. The sheet includes seven (7) columns.

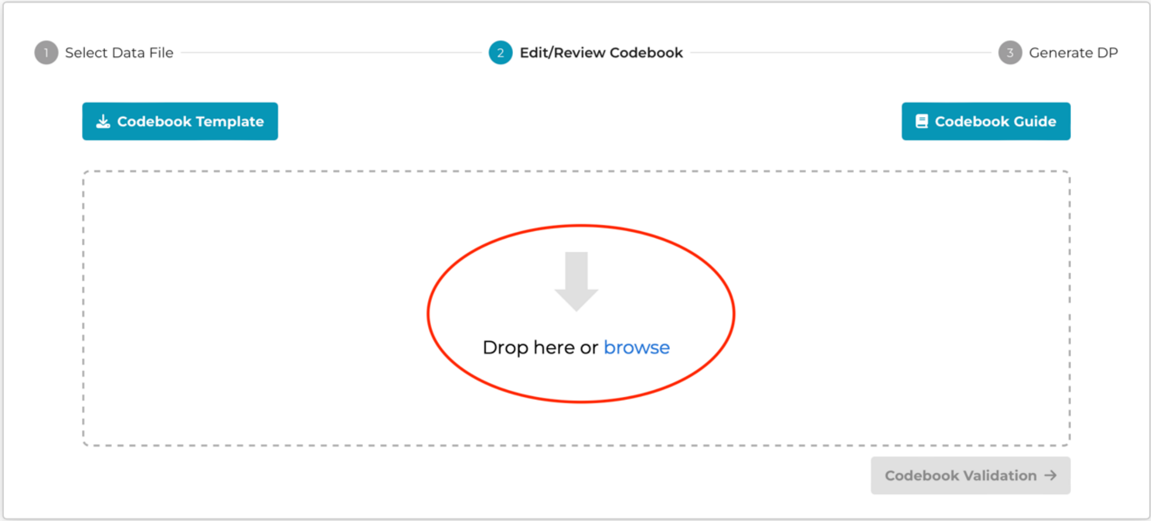
a. **VARIABLE**: This is automatically filled-in by analyzing the CSV datafile newly associated with the data pool. It contains all variable names defined in the datafile, in the order they appear in the datafile.
b. **DESCRIPTION**: The column is filled-in by the data pool user; it provides details and explanations for the relevant variable.
c. **TYPE**: This defines the type of the relevant variable under the typology followed by DataPool and included in the lists sheet. Supported variable types are: STRING, NUMERIC, INTEGER, LOGICAL, CATEGORICAL, DATE. The users editing the codebook must define the type of each variable, selecting one of the available options.
d. **MATCHING STANDARD VARIABLE**: This is used to declare the mapping of a datafile variable with one the variables in the DataPool Standard Schema. In the alpha version, the COUNTRY, REGION, CROP are supported. Users are called to select the standard variables matching their own variable, if such a mapping exists; otherwise, they can leave the relevant cell as is.
e. **PUBLISHABLE**: This is used to declare that the variable should be actually included in the data pool. By default, all variables defined in the datafile are declared as publishable. If the users wish to omit a variable, they select the '-' option from the dropdown box in the relevant cell.

f. **FILTERABLE**: This is used to declare that the variable should be included as a filter option for the data pool's advanced search. By default, all variables defined in the datafile are declared as non-filterable. If the users wish to include a variable as a filter, they select the 'YES' option from the dropdown box in the relevant cell.



VARIABLE	DESCRIPTION	TYPE	MATCHING STANDARD VARIABLE	PUBLISHABLE	FILTERABLE
dataset_id	dataset identifier (normalized URI, that is: "/" replaced with "_")	STRING	-	YES	-
record_id	Unique identifier for each records in a dataset (sequence from 1 to n)	INTEGER	-	YES	-
data_license	license of the dataset from which the datapoints originate	CATEGORICAL	-	YES	YES
dataset_citation	Citation of the originating dataset	STRING	-	YES	-
dataset_uri	URI of the originating dataset	STRING	-	YES	-
reference	Reference to source if the data set is a compilation	STRING	-	YES	-
trial_id	Identifier of a single trial (that may have treatments and replicates); shoulc	STRING	-	YES	-
country	Country name	CATEGORICAL	COUNTRY	YES	YES

When the codebook is completed, users can upload it to the data pool via the same screen they obtained the template, and continue with the validation and data pool version creation steps.


1 Select Data File
2 Edit/Review Codebook
3 Generate DP
Codebook Template
Codebook Guide
Drop here or browse
Codebook Validation →