**The FDACS Tool for Quality Assurance of BMP Datasets**

**Script:** FDACS\_QA\_single\_BMP

**Date:** Aug. 12, 2024

**Introduction**

A major issue in preparing any dataset for subsequent use is whether the data meet the expectations of users in terms of accuracy, completeness, documentation, and other criteria. The FDACS tool for quality assurance (QA) conducts a series of tests on a dataset. These tests may usefully be run multiple times while completing a dataset to avoid problems becoming harder to resolve as more data are entered. This feature of assessing quality during the creation process distinguishes “quality assurance” from the more common term “quality control”, which refers to assessing the quality of a finished product.

The R script FDACS\_QA\_single\_BMP reviews FDACS BMP datasets according to what we term the "four C's", whereby a dataset is:

1. **Correct**: The values are accurate within the expected range of measurement error. We emphasize that the main error-checking should be done as a part of the normal data management pipeline prior to loading into the FDACS BMP template.
2. **Complete**: The dataset is complete enough to enable further analyses without researchers having to seek guidance on how the crop was grown, weather conditions, etc.
3. **Coherent**: Identifiers (keys) that link data across sheets are used consistently.
4. **Compatible**: By linking the FDACS terminology to the ICASA standards, we expect that datasets can be used with a wide range of tools including artificial intelligence, machine learning and either simulation or statistical models.

The script produces a report as a PDF file. The output alternates between explanatory text and blocks of output from R. The script does not modify the target FDACS BMP dataset.

**Procedure for Running the Script**

The basic procedure for using the script is outlined below. We assume that the user is familiar with basic operation of the R language, use of RStudio, and understands how FDACS BMP datasets are organized.

1. Download the script ‘FDACS\_QA\_single\_BMP’ to a folder of your choice.
2. Copy the FDACS BMP dataset file to the same folder.
3. Open the R script ‘FDACS\_QA\_single\_BMP’ in RStudio.
4. If not done previously, install the packages that RStudio detects are required. These might include: openxlsx2, knitr, ggplot2, maps, mapdata, and reshape2.
5. In the script, edit the variable ‘data\_set\_name’ to contain the name of the FDACS BMP dataset to be assessed. This should be the only modification to the R script.
6. Run the script from RStudio by entering ‘alt-ctrl-r’ or navigating through the menu bar to Code -> Run Region -> Run All.
7. The script should run, outputting a few lines of information to assist possible de-bugging.
8. Upon completion, execute the ‘Knit’ command, which should appear as a button on the Source pane (upper left) of RStudio.
9. The ‘Render” tab should become active and will display a few processing steps as the document is converted to a PDF titled ‘FDACS\_QA\_single\_BMP.pdf’.
10. View the document in the same folder as the script.

This completes the process for generating a QA report for an FDACS BMP dataset.

**Troubleshooting**

One possible source of problems is if individual worksheets are completely empty. We have attempted to limit the errors that lack of data will generate, but it is difficult to know if all cases are covered.

The most likely source of small errors is mismatches among worksheet names, individual variable names, and values for identifiers.