IITA Transforming African Agriculture CGIAR	**Crop:** Maize **Function:** Data Management	**SOP #**	IITA-MZ-SOP03
		Revision #	IITA-MZ-SOP03-01
		Implementation Date	15/07/2022
Page #	1 of 11	**Last Reviewed/Update Date**	30/06/2022
SOP Owner	Data Manager, Breeding (Simon Imoro)	**Approval Date**	15/07/2022

Standard Operating Procedure (SOP) for Maize Data Management

Authors & Contributors: Simon Imoro(S.Imoro@cgiar.org), Wende Mengesha(W.Mengesha@cgiar.org), Silvestro Meseka(S.Meseka@cgiar.org), Bunmi Bossey(B.Bossey@cgiar.org), Abebe Menkir (A.Menkir@cgiar.org)

1. Introduction

The role of data management cannot be underestimated in the modernization efforts of breeding operations in accelerating improvement in genetic gains. Current breeding programs generate chunks of data which they rely on for decision making at all stages of the breeding cycle. Lack of standard procedures in data collection and management is likely to introduce errors into the data unknowingly. These errors can negatively impact genetic gain and the uptake of improved maize varieties by smallholder farmers. This document describes standard operating procedures (SOPs) for implementing breeding data workflows to ensure that all necessary breeding data for maize is recorded appropriately and made readily accessible. This is a living document subject to update to accommodate current learning and workflows in the breeding cycle.

2. Purpose

The purpose of this document is to outline the roles, responsibilities, and procedures to be followed in conducting data collection and management in maize breeding activities for all phenotyping trials using enterprise breeding system (EBS) and digital tools. This SOP also intends to guide the Field Coordinator or the Field Technicians in recording data using the available breeding data management systems. It also provides data validation checks to ensure the accuracy and integrity of the data that will be stored in EBS.

3. Scope

This document contains the data management procedure required for maize breeding activities. It covers creating of new trials in EBS, preparing data collection tools and files, data collection, data export, data validation and curation, uploading collected data to EBS and data analysis.

4. Definition of terms

The Enterprise Breeding System is an open-source cloud-based breeding informatics software being developed for crop breeding programs serving resource-poor farmers in Africa, Asia, and Latin America. It provides applications for core breeding and data management activities, enabling a high-quality experience for decision-making processes. It's a relational database that can manage different types of breeding data; plan, create, and manage breeding trials.

Field Book App is an Android-based mobile application for plant phenotyping. The app allows users to import a trial layout file in CSV format, and a trait list text file to setup data collection.

CKAN is the world's leading open-source data portal platform for open-access data storage. CKAN makes it easy to publish, share and work with data including cataloging, storing and accessing datasets with a rich front-end, full API.

Android handheld device/Tablet: A tablet or smartphone that runs the Android operating system and support almost all the key features found in a regular tablet PC

Entry list: Group of germplasm materials for an experiment

Suppressed data: Data suppression means data will still be stored in EBS for audit purposes but cannot be used when carrying out the data analysis task.

Data validation: This involves checking whether the observation recorded coincides with the accepted data type or list of accepted values. It also automates formula computation to the traits as necessary.

Data Curation: Removing inconsistent and redundant data from a trial data and ensuring the accuracy and integrity of data in stored in the EBS.

Trait list: Group of traits to be evaluated in an experiment.

Experiment (Trial) is a top-level activity in the EBS. Experiments are designed to test or generate germplasm for specific qualities or to observe the performance of existing germplasm. Experiment and Trial are used interchangeably in this document.

Occurrence: Each experiment can be composed of one or more occurrences. Each occurrence is an instance of the experiment that includes some or all the entries included in the original experiment entry list.

Zebra ZM/ZT series Printer, Toshiba SX5 printer, RENA Package printer, Epson FX-2190 printer, Laser printer

5. Roles and Responsibilities

All staff involved in implementing breeding activities in the maize improvement program at IITA must use the breeding data management SOP. No alteration should be made to the procedures unless approved exceptionally by the program leaders. The list of individuals responsible for each section of the data management SOP in the breeding data cycle are listed below.

Crop Lead (CL) Responsible for the overall management of the trials and for delegating team responsibilities. The CL is the lead breeder and coordinator of Maize Improvement Program at IITA.

Breeder (B) Coordinate the field layout of experiments, planting and checks on the implementation of defined protocols on the different experimental sites. Ensures all trials are established in the on-station and out-stations respectively. This includes trial management and data collection.

Trial Manager (TM) Oversees trial preparations and management protocols, land acquisitions, oversees planting in the outstations. Also supervises planning of inputs and other planting logistics for the various stations.

Research Supervisor (RS) Coordinates the activities of the Research Technician to ensure that assigned tasks are carried out correctly. The RS involves in planting, field management and post-trial management practices as well as coordinates fertilizer application in on station and outstation experimental fields. S/he involves in the Nursery and seed increase protocols as delegated by the CL and B respectively.

Research Technician (RT) The Research Technician performs field tasks as defined in the trial protocols such as field data collection or field management practices. RT's responsibility is to perform assigned tasks including the use of digital tools defined in the protocol for capturing, storing, transmitting, and ensuring quality of data within defined time periods.

Research Administrative Manager (RM) Responsible for experiment creation in EBS in order to generate the randomization plan for the specified experimental design by the CL for the trials to be established. Generates Seed labels and Field-Tags for the trials. Carries out required analyses using the relevant statistical methods. Generates trial field layout for data capture using the digital tools. Receives transmitted data captured from the field and ensures that the data quality is uncompromised**.**

Data Curator (DC) is responsible for cleaning data and ensuring that meta data is complete for all trials in EBS.

Data Manager (DM) Provides end-user support and training on the usage of EBS and digital tools. The DM reinforces the implementation of data collection and data management protocols.

6. Procedure

A trial must be created in EBS before carrying out the tasks of data collection. The Data Manager must ensure that all the agreed agronomic traits are defined in the trait ontology created for Maize trials. The Breeders and Field Technicians shall use the Maize trait ontology as their reference in evaluating and scoring the phenotypic data. Manual recording using pen and paper is NOT allowed. Data collection must be done using digital tools such as the Field Book App and Tablets. The collected data should be downloaded from the Tablet/Handheld device same day after the data collection to a computer or online storage preferably to avoid data lost. The data must then be uploaded as soon as possible to the EBS for permanent storage.

A typical phenotyping workflow is shown below in figure 1, it provides an overview of the entire process using EBS and the Field Book App.

The SOPs described below focus on the data management aspects and give an update about how to create an experiment in EBS. General information and procedures for maize field trials design/field prepration and design can be found in the Procedures for conducting maize breeding trials SOP.

Create and Manage New Experiments in EBS

Step1: Log into EBS

Step2.1 Create a new experiment. Detailed step-by-step instructions on how to create an Experiment in EBS are available in the EBS user guide documentation.“Add a new

experiment link." *It's the responsbility of the Breeders and the Data Manager to create trials in the EBS.*

Step2.2 Create a trait file from the list of traits to download for Fieldbook App for data collection. Breeders shall coordinate with the data manager to ensure the right traits are downloaded for each trial. This link clearly outline how to prepare and download your trait file.

Step2.3 Export the triat file and trial field layout for data collection

Once the trait list is succesfully created, download the data collection files to be loaded into the Fieldbook App for phenotyping. The steps to achieve this are document in the EBS user guide document.

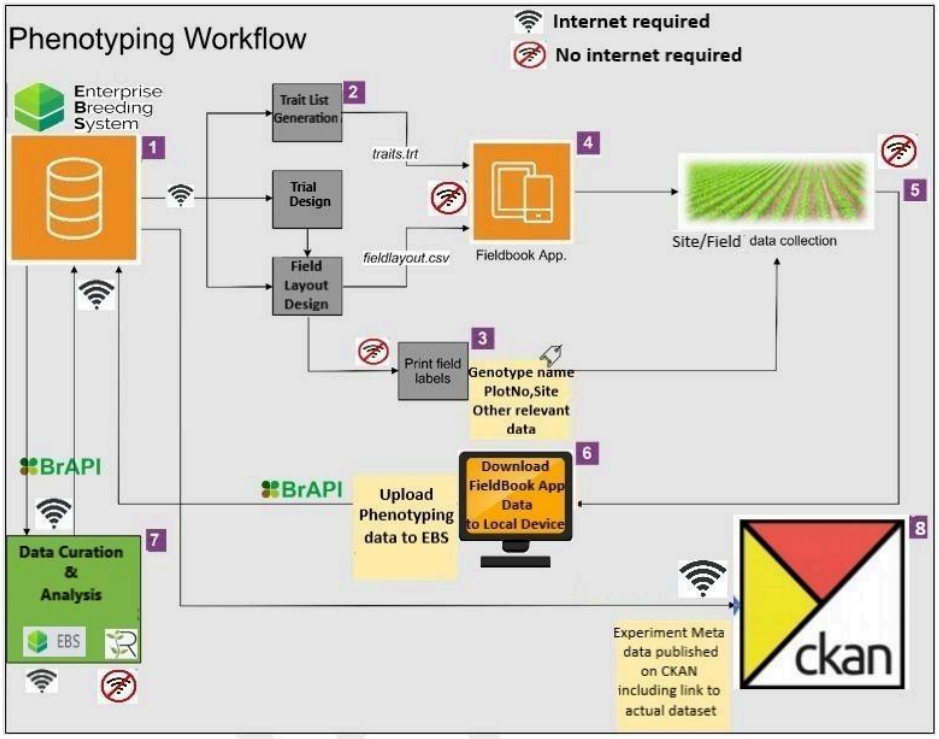

Figure 1: phenotyping data management workflow

Step3. Print field labels for the trial occurences

The exported file will contain amongst other fields the plot ID, entry name, plot number, block number, replication number, row number and column number. Each label should include at least the plot number as a 2D barcode, the entry name and replication number. Other additional information could be planting date and location. The recommended labels for field labeling are Self-tie labels (polyplas material, 150 microns, 4 across). Printing your field labels using your Laser printer or Zebra ZM / ZT series printer.It's the responsibility of the Trial Manager to print labels for every trial.

Step4. Prepare Fieldbook App and android handheld devices for data collection

4.1 Preparation

The Field Technicians shall ensure that the devices to be used are fully charged a day before the actual data collection takes place. The manager shall sure the Field Book App is properly installed on the handheld device from the Google Play Store and the preferred settings enabled to facilitate the task efficiently. The Field Book App is available on Google Play for free. To download the app, you will need to be connected to the internet. Ensure that Google Drive app is also installed on the device to be used in storing and sharing of files. The Trial Manager shall create a trait file that is composed of the traits to be observed during the data collection period. The Trial Coordinator shall download the trait and experiment layout files to the handheld device.

4.2 Generate a field layout design and trait file from your experiment created in EBS. Detail steps in EBS user guide document.

4.3 Import the field design file and trait file downloaded from the database to your android handheld device/tablet.

After downloading the field layout and trait files from EBS to your computer. The next step is to copy these to your tablet. See step by step instructional guide in this link.

Step5. Collect phenotypes using Fieldbook App and Android Tablets

An internet connection is not necessary during data collection. The App allows you to disable moving to the next diserable field during data collection if no data has been collected. This can be enabled in the advanced settings after setting up. We strongly encourage you to collect location information besides the phenotypes. The data manager shall ensure data technicians are up-to-date on new traits and features of the app and tablets for enhanced data quality. An important aspect when collecting phenotypic data is dealing with missing values. Ideally, rules for such missing values must be included in the data collection device, like Field Book App, but if not the case, consider the following guidelines.

Missing values occur when:

- All the plants died in a plot due to external reasons (not related with the quality of the genotype).
- Because for some reason it was not possible to record a data value in a plot.
- There was no data to evaluate for a trait. For example, Ears are needed to evaluate Ear Rot, Field Weight, so if there are no Ears, Ear Rot, and Field Weight a missing value is recorded for both traits.
- A frequent error is to write zero for a missing value or the other way around. A zero and a missing value are different because in the first case a value is assumed

for the trait, while in the second case it is indicated that the value is unknown. If for some reasons the results of a plot were not observed (like all the plants dying because of some external reason), a missing value must be assigned to the number of harvested plants and all other traits that would be observed at harvest. For post-harvest traits like the quality traits (Grain Weight, Grain Moisture, etc.) that need some Ears for evaluation, they should be recorded as missing value when there are no Ears available to get the grain for evaluations. Link to data collection using fieldbook app.

Step6.1 Export phenotypes from Field Book App to your computer/cloud storage

After the collection of phenotypic data with Field Book App, the phenotype data files can be downloaded to your computer. Typically, this will be done partially on different occasions since data recording in the field is often done by several users on several days. We recommend taking regular back-ups on the computers, and even better on a cloud server such as Google Drive, Dropbox, or OneDrive. Ideally, the data on every tablet should be exported on a daily basis. Step-by-step graphical instructions on how to achieve the above are available here.

Step6.2 Upload collected data to EBS.

It's required to **upload all field experiment data to EBS immediately after data collection to avoid losing the data on the tablets/computers**. The data can be in the raw state; yet to undergo quality checks by suppressing suspicious plot data in EBS. The state can then be updated to validated/curated after the quality checks is done. To upload data to EBS, follow these simple steps. The data manager shall ensure all trials in EBS has

their associated trait data including meta data such as trial description, site/field info, planting and harvest dates.

Step7. Curate data to ensure quality and integrity

The phenotype data files can be uploaded also partially (raw state) to EBS but it is important to not forget to curate the data by checking it for inconsistencies and correcting it if necessary. In the future, full data curation will be made possible within EBS and phenotypic data of an experiment will be signaled to be curated and complete, or not. The Data Curator shall be ensure all uploaded data is properly curated in EBS.

7.1 Data Cleaning

Data cleaning involves removing irrelevant or duplicated data, ensuring the accuracy of the data stored in EBS by updating the incorrect or data in question already reflected from the existing trials. Data correction can be done by updating the existing data values with the correct values. The Field Technicians together with Trial Coordinator are responsible for updating the existing data with the correct values, if necessary.

7.2 Data Validation

Quick data quality checks can be performed on experiment data in EBS. Basic descriptive statistics such us the MIN, Max, COUNT, AVG, can be view on a trait or selected traits. Suspicious values can be suppressed for verification and correction if necessary. This can later be unsuppressed after the verification/correction. Here are step-by-step graphical instructions on how to conduct data quality checks in EBS. Guidelines and detailed R based tools for maize data curation will soon be out and annexed to this document.

Data Curator is in charge of all data curation for EBS and **should be notified whenever phenotypic data is to be uploaded** to EBS for data quality checks.

7.3 Data Analysis

Experiment data can be analyzed by downloading the data from EBS into statistical software package like R. Soon some standard analytical models will be made available in EBS for data analysis. The lead Biometrician is responsible to assist breeders analyze trial data for selection decision making and for genetic gain evaluations.

7.3 Storage

All curated or analyzed data is uploaded back to EBS for permanent storage. The data curator shall ensure all curated data is uploaded to EBS.

Step8.Data Sharing/Open Access

After the complete dataset of an experiment has passed quality checks, the meta data is uploaded into CKAN with link to the actual data. It's the responsibility of the data manager to ensure that all completed trial data is shared on CKAN.

7. Appendix

7.1 Contacts for support

For Technical problems with digital tools (Zebra printers, ToshibaSX5 printer, Fieldbook App, EBS), printing field labels and uploading data to EBS, please contact: Simon Imoro (s.imoro@cgiar.org) and Bunmi Bossey (b.bossey@cgiar.org)

Data curation: Abduljelil Olalekan (O.Abduljelil@cgiar.org)

Experimental design and data analysis: Ibnou Dieng (i.dieng@cgiar.org) and Bunmi Bossey (b.bossey@cgiar.org).

8. References

Rife, T. W., & Poland, J. A. (2014). Field book: an open-source application for field data collection on android. Crop Science, 54(4), 1624-1627.

International Potato Center. (2020). Standard operating procedures for sweetpotato breeding data management. COP Breeding Data Management SweetGAINS. International Potato Center: Lima, Peru.