# Data Management in Multi-disciplinary African RTB Crop Breeding Programs

**Afolabi Agbona, Prasad Peteti, Béla Teeken, Olamide Olaosebikan, Abolore Bello, Elizabeth Parkes, Ismail Rabbi, Lukas Mueller, Chiedozie Egesi, and Peter Kulakow**

**Abstract**  Quality phenotype and genotype data are important for the success of a breeding program. Like most programs, African breeding programs generate large multi-disciplinary phenotypic and genotypic datasets from several locations, that must be carefully managed through the use of an appropriate database management system (DBMS) in order to generate reliable and accurate information for breeding-decisions. A DBMS is essential in data collection, storage, retrieval, validation, curation and analysis in plant breeding programs to enhance the ultimate goal of increasing genetic gain. The International Institute of Tropical Agriculture (IITA), working on the roots, tubers and banana (RTB) crops like cassava, yam, banana and plantain has deployed a FAIR-compliant (Findable, Accessible, Interoperable, Reusable) database; BREEDBASE. The functionalities of this database in data management and analysis have been instrumental in achieving breeding goals. Standard Operating Procedures (SOP) for each breeding process have been developed to allow a cognitive walkthrough for users. This has further helped to increase the usage and enhance the acceptability of the system. The wide acceptability gained among breeders in global cassava research programs has resulted in improvements in the precision and quality of genotype and phenotype data, and subsequent improvement in achievement of breeding program goals. Several innovative gender responsive approaches and initiatives have identified users and their preferences which have informed improved customer and product profiles. A remaining bottleneck is the

A. Agbona (✉)
International Institute of Tropical Agriculture (IITA), Ibadan, Oyo State, Nigeria

Molecular & Environmental Plant Sciences, Texas A & M University, College Station, TX, USA
e-mail: a.afolabi@cgiar.org

P. Peteti · B. Teeken · O. Olaosebikan · A. Bello · E. Parkes · I. Rabbi · C. Egesi · P. Kulakow
International Institute of Tropical Agriculture (IITA), Ibadan, Oyo State, Nigeria

L. Mueller
Boyce Thompson Institute for Plant Research, Ithaca, NY, USA

85

effective linking of data on preferences and social information of crop users with technical breeding data to make this process more effective.

## 1 Introduction

An appropriate database management system (DBMS) is essential for any organization to generate reliable and accurate information that will guide decision-making (Meiryani, 2019). Similarly, African breeding programs need a well-structured database management system, to carefully manage the large multi-disciplinary phenotypic, genotypic, and social datasets generated from multiple breeding programs across diverse regions and agroecosystems comprising genetic resource collections, plant breeding trials, on-farm trials, processor evaluations, and consumer testing. A DBMS is a software designed to define, manipulate, retrieve and manage data in a database. It generally manipulates the data itself, the data format, field names, record structure and file structure (Jaekel, 2013). A DBMS is essential in data collection, storage, retrieval, validation, curation and analysis in plant breeding programs to enhance the ultimate goal of increasing genetic gain, and informing breeding efforts at each advancement stage.

The International Institute of Tropical Agriculture (IITA), working on roots, tubers and banana (RTB) crops like cassava, yam, banana and plantain have deployed the use of a Findable, Accessible, Interoperable, Reusable web-based database system; BREEDBASE (https://www.BREEDBASE.org) which was borne out of the NextGen Cassava Project (https://www.nextgencassava.org). The NextGen project seeks to modernize cassava breeding using cutting-edge tools for efficient delivery of improved cassava that satisfies end user needs to farmers in sub-Saharan Africa. The ultimate vision is to improve genetic gain and to deliver cassava varieties with increased yield and disease resistance and other highly preferred traits into the hands of these farmers.

BREEDBASE is an open source, open access, web-based, breeding software available to the scientific community. It can accommodate phenotypic, genotypic, and environmental data collection, storage and analysis tools. It also includes support of the PhenoApps and the breeding API (BrAPI; Selby et al., 2019)) allowing tool integration among the breeding community (Simoes et al., 2019). The functionalities of these databases in data management and analyses have been instrumental in achieving key breeding goals such as monitoring and improving genetic gain, as well as increasing adoption of varieties with end-user preferred traits based on improving data quality and data analysis.

BREEDBASE capabilities include employment of:

- a user-friendly ontology (https://www.cropontology.org): The ontology is the controlled vocabulary that describes the trait phenotyped in a crop. Traits are

grouped into classes (Morphological, Physiological, Quality, etc.) and each has its associated method and unit of measurement defined. In simple terms, it is a dictionary of traits.

- statistical analyses: Data analysis tools that can help breeders make inferences from their dataset. Some of the tools include mixed models for single trial analysis, Genome Wide Association Study, stability analysis, and simple descriptive statistics.
- interfaces with BrAPI (Breeding API, https://brapi.org/): BrAPI is a RESTful (Representational State Tranfer) web service Application Programming Interface (API) that helps to simplify integration and exchange of data across system and databases. Through its interface, exchange of phenotypic and genotypic data is possible.
- barcode-based data collection using the PhenoApps (http://phenoapps.org/): PhenoApps are a suite of barcode-enabled phenotyping tools. Tools include *Fieldbook* for phenotype data collection, *Coordinate* for genotype tissue sample collection and tracking, and *Inventory* for weighing samples without the need for data transcription.

BREEDBASE instances such as cassavabase.org can be launched for any crops using the Docker solution for ease of deployment.

## 2   Structure of BREEDBASE

BREEDBASE is developed using an open-source data schema (CHADO) and other software (PERL, JavaScript frameworks like JQuery, D3, Bootstrap). The CHADO database schema is a widely used database schema for model organism databases (Jung et al., 2011). It is a modular, ontology based, flexible design that can easily be implemented (Figs. 1 and 2).

All the source codes and database schemas used in the development of BREEDBASE are open source and available for download at https://github.com/solgenomics (Tecle et al., 2014) (Fig. 3).

Cassavabase was the first and is the most widely used instance of BREEDBASE and currently accommodates 1164 users from 22 breeding programs. It holds 459,000 accessions from 4070 phenotyping trials (of which 2642 are from the IITA cassava breeding program), 365 phenotyping traits collected for 34,000 genotypes, and over 15.3 million phenotypes representing 962,000 plots from 436 locations with approximately 19,000 images linked to plot information (Fig. 4).
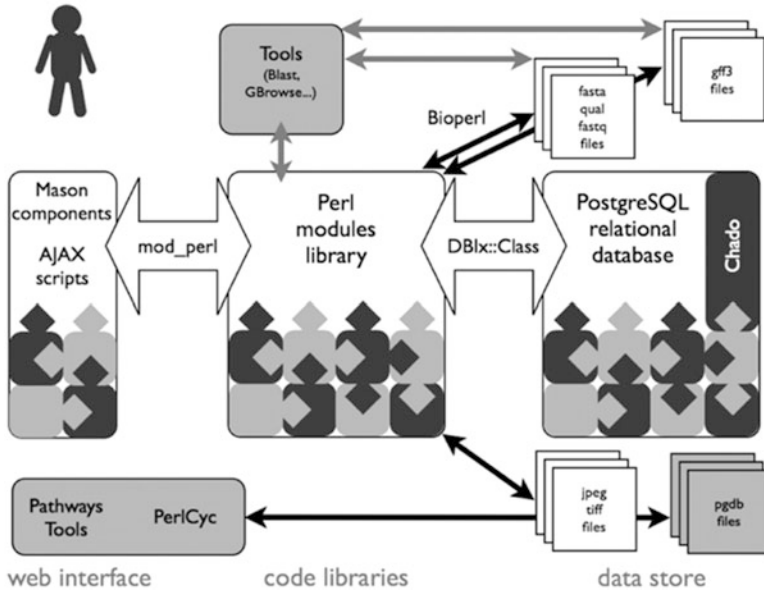
**Fig. 1** BREEDBASE database schema. (Photo credit: Mueller's Lab in BTI, Ithaca)
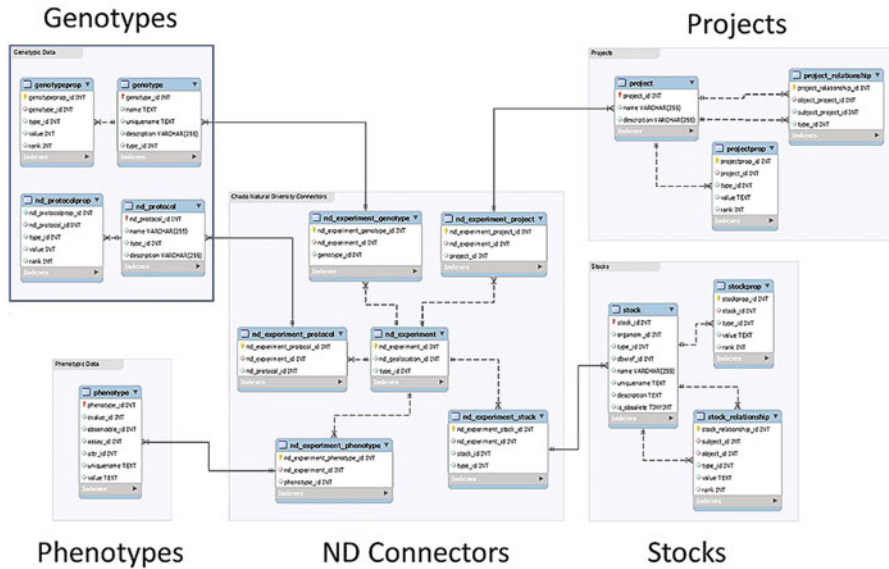


**Fig. 2** BREEDBASE entity relationship schema. (Photo credit: Mueller's Lab in BTI, Ithaca)
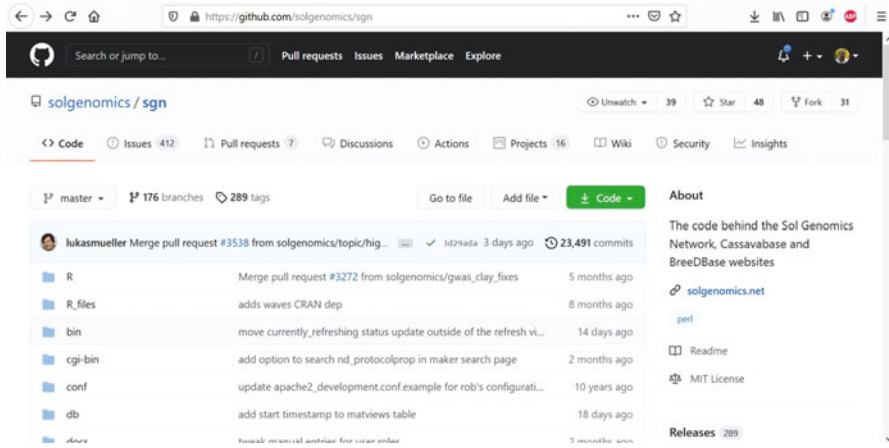
Fig. 3  BREEDBASE code available on solgenomics.net



Fig. 4  Distribution of data currently hosted in Cassavabase

## 2.1  Implementation of the Cassavabase Mirror Site for Data Sustainability

The Cassavabase mirror site https://iita-mirror.cassavabase.org has been hosted at IITA in Ibadan, Nigeria since 2016. A mirror site is a replica of a website or network node. The concept applies to network services, and have different URLs than the original site, but a mirror site has identical or near-identical content (Glushko, 2014) as the primary database. Similarly, the Cassavabase mirror site is currently used as the replica of the main site hosted at the Boyce Thomson Institute Ithaca, NY USA (https://cassavabase.org/). The software and databases are updated every week. The

mirror site provides a real-time backup of the original site, it reduces network traffic, improves access speed, and ensures availability of the original site for technical reasons. Mirror sites are particularly important in developing countries, where internet access may be slower or less reliable (Sekikawa et al., 2000). The Cassavabase mirror site was installed using the necessary dedicated hardware, software, and servers, with the help of Lukas Mueller of Boyce Thomson Institute. The Cassavabase mirror site was also established by the NextGen cassava project to build local capacity to host the primary production site to be maintained at IITA by the end of the project. For a project driven database, this is essential to ensure sustainability of the database should the funding and support system for it change.

## 2.2 BREEDBASE-Centered Data Management Workflow

IITA cassava breeding implements a data management workflow centered around Cassavabase. It uses the DBMS functionalities to plan and implement new trial designs, drawing from the wealth of knowledge provided from previous data. This creates a well-guided approach to modernize breeding. The database easily connects to PhenoApps for quality assured data collection using barcodes, and is then integrated into Dropbox (http://dropbox.com) for short-term data storage and also to enable access to the data prior to final curation and uploading to Cassavabase. The three different types of data generated are phenotypic, genotypic and social datasets.
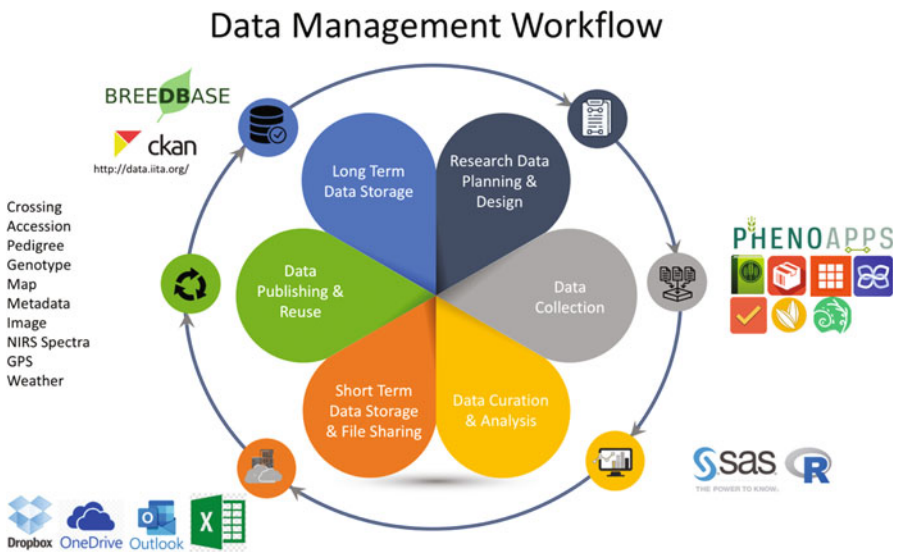


Fig. 5 Cassava breeding data management workflow

Data analysis is carried out using in-house scripts developed in R (R Core Team, 2018) purposefully for plant breeding trial data processing. Efforts are in place to enhance the phenotypic analytical capacity of the database to ensure maximal usage. However, there are often challenges uploading social data from surveys and farmers' trials into Cassavabase due to routine variations in variable terms used from one study to another oweing to the descriptions provided by the respondents who are mostly processors, marketers and farmers. Complimentary Knowledge Archive Network (CKAN), an open-source data portal (http://data.iita.org/) is adopted for storing these datasets. BrAPI calls can then be used to source data from Cassavabase into CKAN (Fig. 5). Data linkage with other crop breeding applications can also achieved using BrAPI. This reduces unnecessary duplication of tasks and more efficient use of resources.

## 3   Application of the Cassava Trait Ontology

The cassava trait ontology enhances the interoperability and effectiveness of data exchange between databases by providing standard concepts (including breeder, farmer and end-user terms) to describe the phenotypic information stored in those databases. The cassava ontology workspace within the database (https://cassavabase.org/search/traits) currently describes 365 variables terms and 206 post-composed terms representing important trait groups for several characteristics captured including traits from recent surveys with end-users for crop improvement (agronomic, biotic and abiotic stress, morphological, physiological and quality traits). The cassava ontology has been migrated from solgenomics: (https://github.com/solgenomics/cassava/tree/master/ontology) to the Planteome repository (https://github.com/Planteome/ibp-cassava-traits) and issues are being tracked via https://github.com/Planteome/variable-issue-submission/issues. Making changes to the cassava trait ontology file involves a sequence of validation process. When a new trait is to be added, a request can be made using a submission form available for the entire RTB community on https://submit.rtbbase.org/. The request is then posted as an issue on the Planteome repository (https://github.com/Planteome).

## 4   Product Profiles and Customer Profiles

Product profiles and customer profiles deal with an important question in breeding: for whom are we breeding (which users) and what are the breeding products needed to achieve the targeted outcomes of increasing genetic gain in farmers' fields and improving livelihoods and stimulating gender equity of cassava and empowerment of users along the value chain? To know for whom breeding is conducted it is

necessary to prioritize the segment of people targeted. Donors often stress the need to improve livelihoods of small-scale farmers and other value chain actors, so the first step is to clearly define the target groups to be able to include as many cassava users as possible in a socially inclusively manner. Tools to improve social inclusion in product profile development were developed under the gender and breeding initiative led by The CGIAR (Consultative Group on International Agricultural Research) Research Project on Roots Tubers and Bananas program in cooperation with IITA, the Alliance of Bioversity and CIAT (International Center for Tropical Agriculture) and CIP (International Potato Center) with support from the CGIAR Excellence in Breeding Platform. The result of this was the creation of the Gender Plus tools   (https://www.cgiar.org/innovations/g-tools-for-gender-responsive-breeding/) (Ashby & Polar, 2021a, b, c; Orr et al., 2021a, b; Polar et al., 2021). Information to inform customer and product profile needs sustained cooperation between socio-economists, anthropologists, gender specialists, and food scientists. Information on customer and product profiles is not static but subject to continual change as a result of socio-economic and variety preference dynamics. The customer and product profile information informs breeding on the number of pipelines needed, the number of preferred traits to monitor and the specific traits to prioritize and use throughout the selection stages. This demands a proper documentation and prioritization of traits of which the results and sources should be systematically documented and integrated within the DMBS system, independently from specific project funding to assure continuity and advancement. The Gender Plus (G+) tools partly addressed the structural integration of value chain actor information segregated by gender and other social factors, as well as information on the relative importance of the different food products made from cassava. Redesigning of the CGIAR customer and product profile tools is ongoing and cassava breeding has played an important exemplary role in this demand-led stage gate (Cooper, 1990; Kotch, 2018; Ragot et al., 2018) breeding effort, as cassava programs held much of the required information on user preferences accounting for the intersections of value chain actors, social segments such as gender groups, poverty, food security status and socio- cultural regions (Polar et al., 2021; Polar & Ashby, 2021; Teeken et al., 2021). Such information needs proper investigation and systematic alignment with the whole breeding process.

Currently cassava breeding programs in West Africa and East Africa have identified 4 provisional product profiles that are heavily Nigeria and Uganda focused because of the 10-year efforts funded through Nextgen cassava and complemented by RTBFoods project funding. The provisional product profiles are:

- *Cassava for food security*: Focus on cassava to be used for fermented food products gari-eba and fufu (major cassava food products in Nigeria) produced and processed by smallholder farmers and processors.
- *Cassava for the fresh market*: Cassava that can be boiled and eaten and/or pounded, important for food security among cassava farming households and specifically for the northern half of Nigeria where cassava is a secondary crop.

This cassava can also be dried into cassava chips that can be milled to create cassava flour which is most common in Northern Nigeria and Uganda.

- *Biofortied cassava for improved nutrition*: Cassava that is biofortified and aimed at increasing nutritional security especially for nutritionally insecure social segments that do not have access to sufficient other sources of vitamin A such as vegetables.
- *Cassava for industry*: Cassava that can be used as starch and ethanol sources for processed food products (such as composite foods) and non-food products.

These profiles are currently end product focused. In the course of the new initiatives on customer and product profiling these classifications can change through including profiles specifically focused on certain user segments that need extra attention from a social inclusion perspective.

The matrix of issues to consider includes prioritizing locations and customer (user) segments, inputs from value chain actors, ecological and cultural regionality, as well as the demand led stage gate breeding focus on variety replacement and commercialization of seed delivery, and finally the intersection of all these domains with gender. This complex matrix needs to be filled as completely as possible in order to informed the minimum number of product profiles needed and the composite of traits needed to be prioritized within each of them. This will assure that maximal impact can be cost effectively achieved with regards to the different development goals' impact areas.



**Fig. 6** Marie Angelique Laporte, Elizabeth Arnaud from Crop Ontology and Afolabi Agbona visited BTI to develop strategies for storing farmer/processor related traits in BREEDBASE in November 2019

## 4.1   The Need to Harmonize Datasets Generated Within the Breeding Program to Fully Optimize Adoption of Product Design and Development Strategy

Continuous interaction between the crop ontology, database management, and breeders will be an important step going forward. This used to happen through the crop ontology workshop but this may not have all stakeholders present. In 2019, an RTB (Roots, Tubers and Bananas) workshop (Fig. 6) was held at the Boyce Thompson Institute where the crop ontology group including the curators met with the database group to discuss a common approach for RTB ontology quality content for agronomic, quality, gender-sensitive and Participatory Varietal Selection (PVS) traits and variables.

## 5   Application of BREEDBASE for Quality Control

Plant phenotypic data comprises information that can be analyzed as datasets individually or combined with existing datasets and reanalyzed. The correct interpretation, comparability, replicability and interoperability of these data is only possible provided the collected data are equipped with an adequate set of useful metadata. The metadata contains information needed to understand and effectively use the data. Thus, metadata is receiving increasing attention across a broad spectrum, to help interpret phenotypic data to achieve the goals of the scientific community. The rows and columns of numeric and textual observations contained within a data set are frequently referred to as raw data. Raw data are usually considered valuable if they can be used within the scientific framework of the study that generated the data. Interpreting and using raw data to investigate a study's underlying theoretical or conceptual model(s) requires an understanding of the types of variables measured. The measurement units, the data quality, the conditions under which the variables were measured, and other relevant facts are all needed and are provided in the metadata. Information is then generated from the combination of raw data and metadata. BREEDBASE collects metadata such as study name, study description, study year, location, date of planting, date of harvesting, plot length, plot width, number of replications, number of blocks, plant stands per plot, field size, unit spacing etc. Metadata like plot length, plot width and field size are very important when estimating yield.

Other useful data collected by BREEDBASE and linked to the metadata include weather data, GPS, plot images, and crossing information. The collected weather data are available on https://weather.rtbbase.org/. Weather data collected include temperature, rainfall, light intensity and day length. Plot level GPS data has also been collected using the handheld Garmin 20x device, for recent trials from Ibadan and Ubiaja in Nigeria. This information is also available on Cassavabase. Plans are under consideration to install a Real Time Kinematic (RTK) positioning system,

which would help to collect more accurate GPS data. More than 19,000 plot images, linked to traits described in the ontology have been uploaded, and image analysis like root necrosis, whitefly counts, etc. can be performed on these images. The crossing information, which includes cross name, female parent, male parent, cross type, number of flowers, number of fruits, and number of seeds are properly managed in Cassavabase. Crossing block activity can include more than 10,000 crosses, carried out over a 3-month time period, by a large team of specialized field technicians. We have implemented a new PhenoApps tool; Intercross, which makes it easier to track and collect the crossing information using barcodes that increases the quality of the crossing data. This information can be linked to the seedling nursery trial to provide proper pedigree linkages. Information on task or gender specific benefits of some traits also informed trait selection in breeding complementing phenotype and genotype information. Breeders are informed of the social or gender implications of some traits when selected or prioritized.

## 6 Integrating Feedback from Social Data for Enhancing Decision-Making in the Breeding Pipeline

The choice of parents fed back into the breeding pipeline is guided by selections from the advanced and late testing stages as well as the feedback from demand creation trials (DCTs), farmers' trial evaluations, surveys, participatory processing, and other methods that provide direct feedback from end users. Demand creation trials are mostly used for variety promotion. It enables processors to choose the most suitable variety for their needs. Variety demand by diverse end users in differentiated cassava markets is determined by communication with the end-users, demand creation trials and inferences from cassava variety adoption studies. Cassavabase contains most of the DCT data from recent years across different locations such as Ago-Owu, Ikenne, Ilorin, Abuja, etc. This provides a decision-making tool for processors to access production and processing value of varieties to generate demand.



Fig. 7  Social engagement with farmers and processors. (Photo courtesy of Béla Teeken)

In order to complement quality data collection efforts to inform breeding or crop improvement initiatives, the survey team constantly or periodically engages farmers, processors and other food chain actors/end-users in the evaluation of new and existing (crop-cassava) varieties alongside local and commonly grown popular varieties, to determine their trait preferences at the production, processing and utilization stages, market valuation of traits and benefits or gains accrued utilizing these improved crop varieties (Fig. 7). Using approaches such as mixed methods to collect quantitative and qualitative data on gender and social aspects (Teeken et al., 2018), informative datasets and inferences have been generated over the years using the Tricot triadic comparison citizen science technology in large scale participatory variety selection and consumer testing (van Etten et al., 2019, 2020; Moyo et al., 2021). The online tablet-based 1000Minds (www.1000minds.com) survey, that is a pairwise comparison between options of equal monetary value using the PAPRIKA method (Potentially All Pairwise Rankings of all possible Alternatives), Mother-Baby trials (Teeken et al., 2021), RTBfoods project (https://rtbfoods.cirad.fr/) methodologies, participatory processing and consumer testing activities (Ndjouenkeu et al., 2021; Forsythe et al., 2021; Teeken et al., 2021; Amah et al., 2021), have all been managed in the DBMS. Although scalable protocols and ways to systematically and more effectively connect these data to to breeding and food science data to inform breeding need further development which is fortunately a mandate of the new Market Intelligence and Product Profiling initiative of the One CGIAR (CGIAR, 2021). Survey results have already significantly informed further improvement to the cassava ontology in order to integrate farmer's and other users' descriptions as traits. The results have also informed gender responsive product profiles for food (especially related to processibllity and food product quality), industry and biofortification. Solutions that have been adopted by some farmers through training platforms like Tricot include standardized cassava spacing and the slant planting which farmers have observed to give more and better or increased yield per plot.

It is imperative to understand the way farmers and processors describe a trait as well as the value placed on such a trait. This will inform the ontology of such trait. Scalable approaches such 1000minds and Tricot are promising in providing such user information because they focus on centralized data management (e.g. www.ClimMob.net for the Tricot approach) and scaling and aim to standardize procedures to generate user feedback in such a way that it becomes an integrated part of the breeding data allowing optimal data integrationLinking of ClimMob and Breedbase is one of the main objectives of the developers of Tricot and ClimMob (van Etten et al., 2020). There is a need for controlled and unified trait description before the social data collected during field surveys or participatory varietal selection can be useful to breeders. The crop ontology will provide a controlled vocabulary set for economically important traits (Shrestha et al., 2010) described by farmers, processors and other value chain stakeholders. Such traits will not be useful or meaningful to breeders until an interdisciplinary study is conducted by breeders, social scientist and food scientists, in which food scientists translate verbal trait descriptions collected by social scientists into measurable traits for breeders' use.

Language barriers, the translation/interpretation of verbal/raw field data collected by social scientists, the adoption of a unified concept for inquiry during surveys as well as the design of an appropriate template that can accommodate social science and food science data are part of the bottle-neck limiting integration of feedback of social data for decision making in breeding programs.

This calls for adoption or incorporation of gender responsive studies to identify regional and cultural differences/similarities in description of traits preferred by men and women (Olaosebikan et al., 2018, 2019; Teeken et al., 2018). This was one of the objectives of the GREAT (Gender Responsive Researcher Equipped with Agricultural Transformation) program organized in sub-Saharan Africa to train researchers that can transform African Agriculture through conducting innovative socially inclusive and end-user-oriented studies. The program trained multidisciplinary teams to structurally integrate gender into the technical and bio-physical sciences.

Trained experts (data curators, application developers, data analysts) or scientists (in breeding, social science, food science) will work together to validate trait descriptions and process traits to a form that can easily be comprehended by breeders before such traits can be used in breeding programs. With regards to food product quality traits this is currently happening within the RTB foods project (https://rtbfoods.cirad.fr/) where the social scientist presents the crop characteristics preferred by users, ranked in order of importance based on survey, participatory processing and consumer testing, to food scientists and breeders for a further translation into operational traits and to determine the first two additional traits to focus on. E.g. the product profile for the gari and eba cassava food products currently addresses colour and food product texture, with Standard Operation Procedures (SOP) being developed to measure these informed by evaluating good and less good varieties as processed and evaluated by users. This multidisciplinary process has happened but will be formalized into a RTB foods product profile document. The focus crops here are Cassava, Cooking banana, Sweet potato, Potato and Yam. Product profiles are formulated per country and for the different major food products made from the crops. With regards to cassava a current MSc research, using ground penetrating radar technology, is looking at how the current Nextgen varieties of cassava perform with regards to early maturity, which will reveal if special selection for early maturity will be necessary.

Decision making in breeding cannot be reached without synthesis of social and food scientists' data collected at different points of contact with farmers and consumers. Social data collected on the emotional, hedonic and organoleptic descriptions of trait preferences of raw food crops as well as culinary food product characteristics will be tested in the laboratory by food scientists to confirm social data. Food science transforms social data into measurable data that can be quantified and presented to the breeder for incorporation into breeding objectives and development of market driven or demand-led breeding to meet the need of stakeholders in crop value chains, as championed by the Excellence in Breeding Platform.

More so, it is necessary to develop Product Profiles for crops which will assist breeders in shaping breeding objectives that will be beneficial to end-users. This is

achievable through concerted efforts of social, physical and bio-physical scientists by engaging in an interdisciplinary project that will lead to development of SOPs which can be adopted across the CGIAR's international research centers as well as national partners in sub-Saharan Africa. This will enhance the development of agriculture as well as the reliability of data collected since the centers and partner institutes will be using the same approaches and methods of operation from the field through the laboratory to data storing methods. This will also foster relationships within and between the institutes.

Near Infrared Spectroscopy (NIRS) offers possibilities to link root and food characteristics (Alamu et al., 2021). This requires following rigorous SOPs for NIRS scans on fresh roots from breeding trials, on the intermediate food products (if applicable) as well as on the final food products as prepared in food science labs following processing and preparation steps that have been externally validated by participatory processing of contrasting clones (new clones and processor preferred clones) in the working environment of the users (village cottage processing, or larger scale using mechanized processing depending on the product profile). The clones to be evaluated with the users in their own environment are preferably grown in the same trail and close to the production source of the users, which often implies an on-farm trial or breeding location close to the communities of the processors. This concert of activities can result in finding relationships between food product quality traits and physiochemical characteristics of fresh roots which would allow for earlier selection for food product quality and processability (the amount of drudgery involved in processing) of different varieties. All these activities are currently put in place through the cooperation between the Nextgen cassava and RTB foods projects. Within cassava breeding two proof of concepts are currently identified: the investigation of the discoloration during processing of the root resulting in pale or brownish food products, which is hypothesized to be related to dextrinization, the browning related to caramelization of sugars (non-enzymatic browning) in relation with the presence of polyphenols, (enzymatic browning) in the fresh roots. Another proof of concept is related to the final textural properties, hardness, smoothness, mouldability, adhesiveness and strechability of the dough like food products which are hypothesized to be related to amylase and pectin contents.

## 7    Promoting BREEDBASE Functionality for Increased Usage

To ensure the adoption of technical solutions in BREEDBASE and the proper integration of social solutions to inform decision-making, Quality Champions were appointed from different breeding programs. They are the "go to" people for quality control (QC) and BREEDBASE-related topics. Their roles include:

- Creating awareness and access to best practices/state-of-the-art techniques in quality management.

**Fig. 8** Capacity development across partner stations

- Developing and implementing SOPs, making use of key performance indicators (KPIs) and quality metrics to ensure the correctness of data and other practices involved in breeding.
- Improving efficiency of breeding data collection, curation and storage
- Effecting an increase in the usage of Cassavabase in daily breeding activities
- Training users on QC and data management.

The Quality Champions are also actively involved in promoting digitization of practices; promoting the use of electronic data capture, procuring digital inputs like barcode labels for both phenotype and genotype stocks to improve data quality, among others.

Knowledge sharing is key for the successful implementation of our technical and social solutions. To this effect, regular training of users on the usage of BREEDBASE and PhenoApps tools is essential (Fig. 8).

Every year at IITA Headquarter, we organize biweekly trainings for technicians, supervisors and students for a 2 to 3 months period, usually in the first quarter of each year. During these trainings, we focus on functionalities of Cassavabase, which are mainly useful for the technicians to do their day-to-day breeding activities, such as creating lists, searching for and downloading phenotypic data and layouts, designing barcodes, etc. We also focus on PhenoApps tools like Fieldbook, Coordinate, Intercross, etc. We have also extended this training beyond Nigeria, Uganda and Tanzania to include all IITA regional locations in Southern Africa, East Africa and Central Africa along with national agricultural research systems (NARS) partners. A dedicated community of practice partnership (COPP) has been set up to engage additional NARS breeding staff from 9 countries including Rwanda, DR Congo, Kenya, Cote d'Ivoire, Sierra Leone, Ghana, Zambia, Malawi, and Mozambique. The COPP mainly focuses on expanded use of digital tools, germplasm exchange, support in field management, trial design, data management, use of genotyping for markers and variety identifications and peer visits for knowledge exchange among breeding programs.

## 8 Fostering Continuous Improvement

To maintain continuity of practice and to further improve the system, we are exploring collaborations with different cassava initiatives ranging from the fundamentals of breeding program optimization, development of improved product profiles and efficiently delivering new products to farmers; to addition of new features, increased training of technical staff and linking traits across the value chain using the cassava trait ontology.

In conclusion, the wide applicability of BREEDBASE has encouraged wide acceptability among breeders in the global cassava research programs. The user-friendliness of the DBMS and availability of SOPs for most breeding processes allows a smooth cognitive walk through for the users. All these solutions have resulted in improvements in precision and quality of phenotypic and genotypic data, thus resulting in the overall improvement of breeding program goals.

A bottleneck identified is the full integration of social and anthropological data related to gendered, regional and socially inclusive trait preferences and other relevant social information that informs the customers and product profiles to focus on a more demand led breeding approach, but also the way breeding is organized as it also should determine stakeholders that are to be represented in product advancement and variety release procedures. The mentioned One Cgiar Market Intelligence and Product Profiling innitiatve (CGIAR, 2021) will be important to allow to tackle this bottleneck. The cassava breeding unit in Ibadan has played a role in setting an example for acquiring such social information and is continuing to contribute to redesigning customer and product profiling procedures that will also be informed by market intelligence tools. This is especially important because public breeding is specifically tasked with creating social impact in the form of poverty alleviation among smallholder value chain actors. The Tricot citizen science scaled participatory variety selection approach offers a platform for learning and variety dissemination and building partnerships with users but is also important to test the external validity of the multilocation breeding trials as it systematically evaluates variety performance under farmer conditions. This is important as the main objective of breeding is to increase genetic gain in farmers' fields. Current initiatives are ongoing to connect Tricot data efficiently to BrAPI and Breedbase through the ClimMob online platform (www.climmob.net) (van Etten et al., 2020; Manners et al., 2022, forthcoming) also including almost real time climate data through 'climatrends' and 'chirps'(https://CRAN.Rproject.org/package=climatrends, and https://CRAN.Rproject.org/package=chirps) (de Sousa & van Etten, 2021; de Sousa et al., 2020). Although public breeding shares demand driven approaches with the private sector, it is the explicit social inclusiveness and focus on sustainable development goals (https://www.un.org/sustainabledevelopment/sustainable-development-goals/) that sets public breeding apart. We are confident that during the years to come this part will be well integrated with the modernized DBMS systems.

# References

Alamu, E. O., Nuwamanya, E., Cornet, D., Meghar, K., Adesokan, M., Tran, T., Belalcazar, J., Desfontaines, L., & Davrieux, F. (2021). Near-infrared spectroscopy applications for high-throughput phenotyping for cassava and yam: A review. *International Journal of Food Science and Technology, 56*, 1491–1501. https://doi.org/10.1111/ijfs.14773

Amah, D., Stuart, E., Mignouna, D., Swennen, R., & Teeken, B. (2021). End-user preferences for plantain food products in Nigeria and implications for genetic improvement. *International Journal of Food Science and Technology, 56*, 1148–1159. https://doi.org/10.1111/ijfs.14780

Ashby, J. A., & Polar, V. (2021a). User guide to the G+ product profile query tool (G+PP). In *CGIAR Research Program on Roots, Tubers and Bananas, user guide 2021–2*. Lima, Peru. www.rtb.cgiar.org/gbi. https://hdl.handle.net/10568/113167

Ashby, J. A., & Polar, V. (2021b). User guide to the standard operating procedure for G+ tools (G+SoP). In *CGIAR Research Program on Roots, Tubers and Bananas, User Guide. 2021–3*. International Potato Center. www.rtb.cgiar.org/gbi. https://hdl.handle.net/10568/113166

Ashby, J. A., & Polar, V. (2021c). Description sheet to the gender plus product profile query tool (G+PP). In *CGIAR Research Program on Roots, Tubers and Bananas*. International Potato Center. www.rtb.cgiar.org/gbi. https://hdl.handle.net/10568/113191

CGIAR. (2021). *Market intelligence and product profiling*. Retrieved from: https://storage.googleapis.com/cgiarorg/2021/10/INIT05-Market-Intelligence-and-Product-Profiling.pdf

Cooper, R. G. (1990). Stage-gate systems – A new tool for managing new products. *Business Horizons, 33*, 3.

de Sousa, K., & van Etten, J. (2021). *ClimMobTools: API client for the 'ClimMob' platform* (R package version 0.3.9). https://CRAN.R-project.org/package=ClimMobTools

de Sousa, K., van Etten, J., & Solberg, S. Ø. (2020). *Climatrends: Climate variability indices for ecological modelling* (R package version 0.1.6). https://CRAN.R-project.org/package=climatrends

Forsythe, L., Tufan, H., Bouniol, A., Kleih, U., & Fliedel, G. (2021). An interdisciplinary and participatory methodology to improve user acceptability of root, tuber and banana varieties. *International Journal of Food Science and Technology; Special Issue: Consumers Have Their Say: Assessing Preferred Quality Traits of Roots, Tubers and Cooking Bananas, and Implications for Breeding, 56*, 1115–1123.

Glushko, R. J. (2014). *The discipline of organizing: Core concepts edition*. O'Reilly Media.

Jaekel, T. (2013). The role concept for relational database management systems.. Hong Kong

Jung, S., Menda, N., Redmond, S., Buels, R. M., Friesen, M., Bendana, Y., Sanderson, L. A., Lapp, H., Lee, T., MacCallum, B., Bett, K. E., Cain, S., Clements, D., Mueller, L. A., & Main, D. (2011). The chado natural diversity module: A new generic database schema for large-scale phenotyping and genotyping data. *Database (Oxford), 2011*, bar051.

Kotch, G. P. (2018). *Applying stage-gates to better manage public breeding programs | Excellenceinbreeding* [www document]. https://excellenceinbreeding.org/blog/applying-stage-gates-better-manage-public-breeding-programs. Accessed 12.8.19.

Manners, R., de Sousa, K., Teeken, B., et al. (2022). An agile framework for decentralized on-farm testing supported by citizen science. A review (under review with *Agronomy for Sustainable Development*) (forthcoming).

Meiryani, A. S. (2019). Database management system. *International Journal of Scientific & Technology Research, 8*(06), 309–312.

Moyo, M., Ssali, R., Namanda, S., Nakitto, M., Dery, E. K., Akansake, D., Adjebeng-Danquah, J., van Etten, J., de Sousa, K., Lindqvist-Kreuze, H., Carey, E., & Muzhingi, T. (2021). Consumer preference testing of boiled sweetpotato using crowdsourced citizen science in Ghana and Uganda. *Frontiers in Sustainable Food Systems, 5*, 620363. https://doi.org/10.3389/fsufs.2021.620363

Ndjouenkeu, R., Ngoualem Kegah, F., Teeken, B., Okoye, B., Madu, T., Olaosebikan, O. D., Chijioke, U., Bello, A., Oluwaseun Osunbade, A., Owoade, D., Takam-Tchuente, N. H., Biaton Njeufa, E., Nguiadem Chomdom, I. L., Forsythe, L., Maziya-Dixon, B., & Fliedel, G. (2021).

From cassava to gari: Mapping of quality characteristics and end-user preferences in Cameroon and Nigeria. *International Journal of Food Science and Technology, 56*, 1223–1238. https://doi.org/10.1111/ijfs.14790

Nextgen Cassava. (n.d.). https://www.nextgencassava.org accessed Friday 16 July 2021 & Grant proposal narrarive Phase II: https://www.nextgencassava.org/wp-content/uploads/documents/NextGenRenewal_Proposal%20Narrative_2017%20copy.pdf

Olaosebikan, O., Kulakow, P., Tufan, H., Madu, T., Egesi, C., & Teeken, B. (2018). A case study of cassava trait preferences of men and women farmers in Nigeria: Implications for gender-responsive cassava variety development. In H. A. Tufan, S. Grando, & C. Meola (Eds.), *CGIAR gender and breeding initiative* (pp. 35–43). CIP.

Olaosebikan, O., Abdulrazaq, B., Owoade, D., Ogunade, A., Aina, O., Ilona, P., Muheebwa, A., Teeken, B., Iluebbey, P., Kulakow, P., Bakare, M., & Parkes, E. (2019). Gender-based constraints affecting biofortified cassava production, processing and marketing among men and women adopters in Oyo and Benue States, Nigeria. *Physiological and Molecular Plant Pathology, 105*, 17–27. https://doi.org/10.1016/j.pmpp.2018.11.007

Orr, A., Polar, V., & Ashby, J. A. (2021a). User guide to the G+ customer profile tool (G+ CP). In *CGIAR Research Program on Roots, Tubers and Bananas, user guide 2021–1*. Lima, Peru. www.rtb.cgiar.org/gbi. https://hdl.handle.net/10568/113168

Orr, A., Polar, V., & Ashby, J. A. (2021b). Description sheet to the gender plus customer profile tool (G+CP). In *CGIAR Research Program on Roots, Tubers and Bananas*. International Potato Center. www.rtb.cgiar.org/gbi. https://hdl.handle.net/10568/113190

Polar, V., & Ashby, J. A. (2021). Gender report and template forms for G+ tools. In *CGIAR Research Program on Roots, Tubers and Bananas, report template 2021–1*. Lima, Peru. www.rtb.cgiar.org/gbi. https://hdl.handle.net/10568/113189

Polar, V., Teeken, B., Mwende, J., Marimo, P., Tufan, H. A., Ashby, J. A., Cole, S., Mayanja, S., Okello, J. J., Kulakow, P., & Thiele, G. (2021). Chapter 16: Building demand-led and gender-responsive breeding programs. In G. Thiele (Ed.), *Root, tuber and banana food system innovations*. Springer. https://link.springer.com/content/pdf/10.1007/978-3-030-92022-7.pdf

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ragot, M., Bonierbale, M., & Weltzien, E. (2018). From market demand to breeding decisions: A framework. In *CGIAR gender and breeding initiative working paper 2*. CGIAR Gender and Breeding Initiative.

RTB Foods. (2018). *Project description*. https://rtbfoods.cirad.fr/project/rtbfoods-description. Accessed 1 Apr 2022.

Sekikawa, A., et al. (2000). Internet mirror sites. *The Lancet, 355*(9219), 2000.

Selby, P., et al. (2019). BrAPI – An application programming interface for plant breeding applications. *Bioinformatics, 35*(20), 4147–4155.

Shrestha, R., Arnaud, E., Mauleon, R., Senger, M., Davenport, G. F., Hancock, D., Morrison, N., Bruskiewich, R., & McLaren, G. (2010). Multifunctional crop trait ontology for breeders' data: Field book, annotation, data discovery and semantic enrichment of the literature. *AoB Plants, 2010*, 1–11. https://doi.org/10.1093/aobpla/plq008

Simoes, C., et al. (2019). *Breedbase: A digital ecosystem for plant breeders*. African Plant Breeders Association (APBA) Conference.

Tecle, I. Y., Edwards, J. D., Menda, N., Egesi, C., Rabbi, I. Y., Kulakow, P., Kawuki, R., Jannink, J. L., & Mueller, L. A. (2014). solGS: A web-based tool for genomic selection. *BMC Bioinformatics, 15*, 1–9. https://doi.org/10.1186/s12859-014-0398-7

Teeken, B., Olaosebikan, O., Haleegoah, J., Oladejo, E., Madu, T., Bello, A., Parkes, E., Egesi, C., Kulakow, P., Kirscht, H., & Tufan, H. (2018). Cassava trait preferences of men and women farmers in Nigeria: Implications for breeding. *Economic Botany, 72*, 263–277. https://doi.org/10.1007/s12231-018-9421-7

Teeken, B., Agbona, A., Bello, A., Olaosebikan, O., Alamu, E., Adesokan, M., Awoyale, W., Madu, T., Okoye, B., Chijioke, U., Owoade, D., Okoro, M., Bouniol, A., Dufour, D., Hershey, C., Rabbi, I., Maziya-Dixon, B., Egesi, C., Tufan, H., & Kulakow, P. (2021). Understanding cassava varietal preferences through pairwise ranking of gari-eba and fufu prepared by local farmer–processors. *International Journal of Food Science and Technology, 56*, 1258–1277. https://doi.org/10.1111/ijfs.14862

United Nations. (2021). *The sustainable development goals*. https://www.un.org/sustainabledevelopment/sustainable-development-goals/. Accessed 16 July 2021.

Van Etten, J., Beza, E., Calderer, L., Van Duijvendijk, K., Fadda, C., Fantahun, B., et al. (2019). First experiences with a novel farmer citizen science approach: Crowdsourcing participatory variety selection through on-farm triadic comparisons of technologies (tricot). *Experimental Agriculture, 55*(S1), 275–296. https://doi.org/10.1017/S0014479716000739

Van Etten, J., Abidin, E., Arnaud, D., Brown, E., Carey, E., Laporte, M.-L., López-Noriega, I., Madriz, B., Manners, R., Ortiz-Crespo, B., Quirós, C., de Sousa, K., Teeken, B., Tufan, H. A., Ulzen, J., & Valle-Soto, J. (2020). In www.rtb.cgiar.org (Ed.), *The tricot citizen science approach applied to on-farm variety evaluation: Methodological progress and perspectives* (RTB Working Paper. No. 2021-2). CGIAR Research Program on Roots, Tubers and Bananas (RTB).