

## Linking data across the value chain of underutilised crops – a multidisciplinary approach

Nur Marahaini, M.N., Jahanshiri, E., Salama, A. and Tengku Adhwa Syaherah, T.M.S.

*CropBASE Programme, Crops for the Future, The University of Nottingham Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor, Malaysia.*

### Article history:

Received: 5 June 2018

Received in revised form: 7 September 2018

Accepted: 10 October 2018

Available Online: 6 November 2018

### Keywords:

Data linkages,

Data model,

Data structure,

Multi-disciplinary data,

Value chain.

### DOI:

[https://doi.org/10.26656/fr.2017.3\(2\).104](https://doi.org/10.26656/fr.2017.3(2).104)

### Abstract

Limited research data and lack of data linkage across the value chain of underutilised crops are one of the major barriers to wider adoption of these species in the global food system. Observational and experimental data in agricultural sciences are often recorded in silos. For example, data on nutritional properties are recorded with limited linkage to genomic background of crop varieties and climatic conditions where the crops were grown. This can severely restrict the research impact. In this paper, a practical and bottom-up approach for describing value chain data with emphasis on data linkage across related disciplines is described. The objective was to demonstrate traceability of nutrient composition data from seeds to food product. Related data and their formats were identified from previous experiments and discussed with experts in related disciplines. Existing data standards and established databases were collected and studied. A storyline describing the journey from germplasm to food product was created to identify the interconnectivity between data points. The data structure design was then used to design a relational database and online data manager for the users to interact, view and edit their data. The storyline was implemented in a dynamic web interface that is connected to the relational database to present the details of journey in images, tables, analytical charts and maps. The initial response from the users shows potential and drawbacks of such interconnected data structure and web-based interface that are designed to collate all related data to the agricultural practices.

## 1. Introduction

To cope with aggressive changes in climates, 'business as usual' in agriculture will threaten the future food security (Moore *et al.*, 2017). It will also alter nutrient cycles, cause significant changes in soil moisture regimes, shift pest and disease outbreaks and therefore greatly influence the food production and food security (Fuhrer, 2003; Jones and Thornton, 2003). Resilience in food systems can be achieved through diversification with underutilised crops (Lin, 2011, Massawe *et al.*, 2015). Growing other crops besides major crops has the potential to improve food and nutritional security in the face of climate change (Mayes *et al.*, 2012; Massawe *et al.*, 2015). It is estimated that over 7,000 species are known to be used as food yet only around 30 species provide 95% of the world's food energy (Williams and Haq, 2000). Underutilised crops include a large array of the species and are vastly underexploited (Williams and Haq, 2000). Underutilised crops are also believed to be potential crops to supply high nutritional food to the

globe compared to major crops which are shown to see a decline in nutrients due to impact of climate change (Mabhaudhi *et al.*, 2017). However, limited knowledge of the crops makes it difficult to establish value chains for the food production and marketing.

Crops For the Future (CFF) as the only centre dedicated to research on underutilised crop is taking a holistic approach to research and aims to establish proofs-of-concept and prototypes to demonstrate the value chain of underutilised crops. By demonstrating the journey from field to table, CFF can also analyse gaps in the value chains and achieve specific objectives such as examining nutritional changes from production in the field to harvesting and product development of exemplar underutilised crops. To achieve this, CFF introduced CONNECT, a multi-disciplinary project that records data and follows the nutritional changes in exemplar crops of Bambara groundnut (*Vigna subterranea*) and moringa (*Moringa oleifera*) from production in the field, through harvesting, drying and processing to inclusion in the

\*Corresponding author.

Email: [marahaini.nizar@cffresearch.org](mailto:marahaini.nizar@cffresearch.org)

final food products. In order to achieve the project objective, records of data are needed to be collected to provide evidence in value chain establishment and identification of gaps in value chain. In order to organise data in an accessible form, a database is required to store data in proper format so that it can be used to visualise the processes in the value chain and preserve the data for future uses.

Database design is done through studying properties and interrelationships among data items to provide actual representations of the items in basic data structures of database. Different data models have different data structure to represent data in databases. In relational databases, fundamental structures to represent data is called relational data model (Codd, 1970; O'Neil and O'Neil, 2009). Database design is performed by assigning data items to columns of tables in a manner that maintains desirable properties. The logical design must reflect interrelationships among objects in the real world through tables and attributes/columns (O'Neil and O'Neil, 2009). Relational database model is one of the most commonly used data models in designing data structures. The model has its core with whole series of rules governing keys, relationships, joins, dependencies and modification anomalies. The presentation of the model is made by tables with table names, primary and secondary keys, field names and relationships (Homan and Kovacs, 2009). USDA National Nutrient Database for Standard Reference, Legacy (2018) is an example of database which uses this data model (United States Department of Agriculture, 2018).

This data model fits the purpose of designing database to represent interrelationships among objects in crops value chain and the traceability of nutritional changes of food through data visualisation despite having some limitations. Even though recording every single data point along the value chain is almost impossible given the allocated human resources to the research project, recording initial conditions in an intuitive way that is represented by relational data format is not impossible. The choice of relational database model was recommended because of the availability of various software tools that can be used for developing the database system and its user-friendly interfaces compared to other data models like Object-Oriented-Model (Leavitt, 2000).

The objective of this research is to build a data structure for value chain of crops that can demonstrate the journey through data, trace nutrient contents, and develop a practical web interface for the ease of data collection and management. In this article, a data structure that was developed for storing value chain data from seeds to food product of exemplary crops, Bambara

groundnut and moringa are discussed.

## 2. Methodology

### 2.1 Data flow

The crops have its own genetics and germplasm details stored in the respective tables (Figure 1). To create a connection, the accession name of the sack of seed or cutting is linked to the field plot. Plot stage data is connected to soil sample and field assessment information. Plot stage is also connected to harvested yield. In this study, the harvested yield from the field would either be stored or processed into powder form. The powder form is especially useful for food development departments to make innovative food products. These powders are stored in packs. The yield packages are related to products and samples. Samples are linked to parcels which are sent for nutrient analysis. All existing tables and fields are designed in relational format.

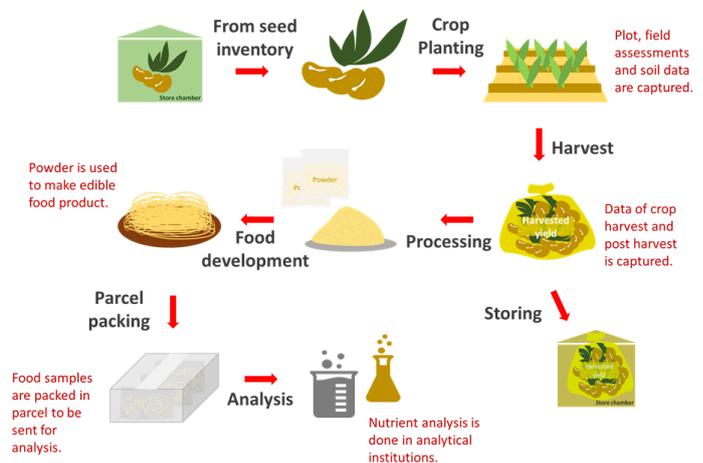


Figure 1. Visualisation of storyline journey from seeds to product for data collection.

### 2.2 Database logical design

The logical design involves entities transforming to tables, attributes transforming to table fields, and relationships established through primary and foreign keys among tables (Codd, 1972; Date, 1999, Napolitano et al., 2018). To avoid confusion, the entities are written with double quotation marks (e.g. "Plot"). There are 10 entities in the current database structure design; "Germplasm and seed inventory", "Crops", "Plot", "Harvested yield", "Yield package", "Soil", "Weather", "Products", "Samples" and "Parcel". The logical model which is composed of tables and fields with primary and foreign keys linkages is illustrated in Figure 2.

The tables related to an entity called "Germplasm and seed inventory" stores information of crops identity and source. The table stores common names and taxon names (e.g. common name, Bambara groundnut; scientific name, *Vigna subterranea* L.). This table is like

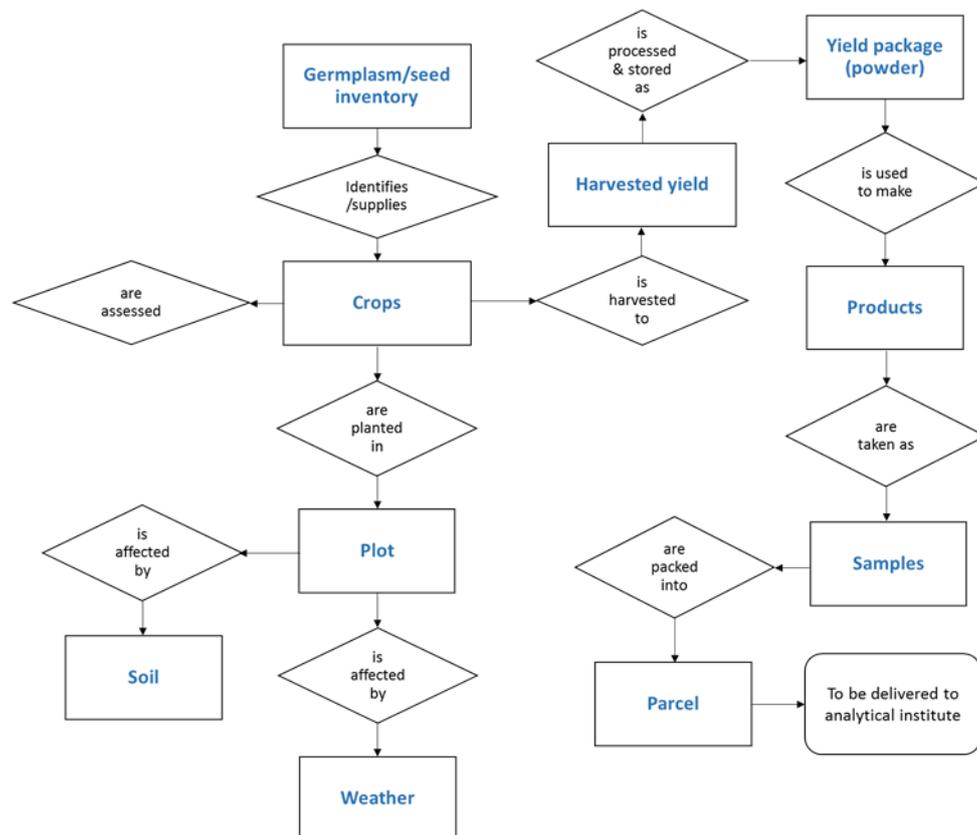


Figure 2. Illustration of entity-relationship diagram in logical database design.

the main table which stores the identity of the crops. The germplasm information of the crop is stored in a table which includes details such as accession number, accession biological status, collection source etc. The information on the seeds availability in CFF is also being recorded.

“Crops” related tables store information such as crop accession name, quantity of seeds or propagation materials, planting distance etc. This entity consists of tables related to field measurements and observations of the crops. The tables record the growth measurement data such as plant height, leaf length, leaf width etc. and yield measurement data such as fresh and dry weight of yield, moisture loss and harvest index. The tables also record the observation of pests and diseases occurrence which comprises details such as the name of possible diseases and pests, number of infected plants and severity index. Simple planting cost is also recorded in some designated fields.

The “Plot” entity includes tables to describe the details of plot usage such as the plot’s name, the date when the plot is started to be used, area of the plot etc. This entity also records activities being done on the plot such as sowing, fertiliser application, pesticide application, irrigation etc. Extra details such as the chemicals and fertilisers names and dosage can be recorded in the “Plot”-related tables.

“Harvested yield” stores information such as harvest date and the amount of harvested crops. It also includes

the logistics data of the harvested materials from point A to point B to trace the journey. The important details such as time of harvest, time of dispatch from point A, time received at point B and person in charge of the logistics are recorded. The post-harvest activities such as washing, drying, packaging of harvested materials are also recorded in fields related to this entity.

“Yield package” is the entity consisting of fields to describe the packages of products derived from processed harvested materials from the previous entity. The important details such as the pack name/ID, amount of the processed product, packaging date, storage start and end dates, storage settings i.e. temperature, humidity etc.

“Soil” is the entity which consists of tables and fields describing the soil details at the plot. The important details such as soil sample profile which includes sampling coordinate and layer details. This information is representing the identity of the soil sample which is used for analysing the chemical and physical properties of the soil (e.g. soil pH, organic carbon, phosphorus and nitrogen content, soil texture etc.). “Weather” is the entity which stores weather data taken from weather station which is installed in Field Research Centre (FRC) in CFF. The information such as air temperature, total rain, solar radiation etc. are stored in the “Weather” table.

“Products” is the entity to describe the product made from a crop or the processed harvested materials (e.g.

Bambara groundnut powder) which is also considered as a product. The tables store important data such as product batch code, product name, product description, ingredients, production methods and product sensory analysis.

The “Samples” entity has tables to describe samples (a portion of amount) of the product which is used for analysing nutrients content. It contains tables to store important data such as sample batch code, date sampled, sample category, food group, sample physical state, the name of the project which requires the analysis etc. It also stores sample management data such as processes of handling, storage and disposal. There is one important table in this entity storing the values from nutrients analysis, analytical method and other details such as tested date and contributor which contributes the data value into the table.

“Parcel delivery” records the details of samples which are being sent for analysis. The tables related to this entity stores information such as parcel code, packaging method, number of samples being sent and logistics data such as method of transportation of the parcel to analytical institution (institution where the nutrients analysis being done), date and time of parcel dispatch and reception, and location of analytical institution.

### 2.3 Database physical design

Data involved in the value chain were identified by observing the activities from seed material preparation to food product development for Bambara groundnut and moringa at various departments. Parameters recorded from the observations were mainly to satisfy two objectives; 1) to trace data from seeds to food product, and 2) to store research data from measurements and analyses made along the value chain. Previous experimental and management records were utilised to construct dataset parameters for data storage. Existing international standards related to food and nutrition databases were referred to add more descriptive parameters and to validate the parameter names and formats. To ensure completeness of data structure, discussions with experts and data producers and owners were continuously held to determine traceability parameters. A storyline was created to identify and refine data stages and captured data for the purposes of building data structure.

FAO/Bioversity Multi-Crop Passport Descriptors V.2.1 (Alercia *et al.*, 2015) was used as a reference to build the data structure of germplasm tables. To simplify the data collection, only major descriptors from this standard were taken. In this article “descriptor” and

“parameter” are used interchangeably. The parameter refers to variables or columns in a table and is the term used in computer science whereas descriptor or trait is a standard naming in scientific and breeding community.

ISCRIC-WISE Harmonised Global Soil Profile Dataset Ver. 3.1 (Batjes, 2008) developed by World Soil Information (ISCRIC) was used as a reference to design data structure for soil section. Soil analysis in the CFF Field Research Centre (FRC) contained variables that are missing in the list of ISCRIC-WISE data standard. Therefore, a new structure which is flexible to store soil data even when a new variable is introduced was designed.

International Network of Food Data Systems (INFOODS) Guidelines for Describing Foods (Truswell *et al.*, 1991) by Food and Agriculture Organization (FAO) was used to decide important data to be recorded as a food product. Important dataset in the guidelines was taken to be included in the structure. The designed structure was also influenced by the presentation of data in AUSNUT 2011-13 food nutrient database (Food Standards Australia New Zealand, 2016).

This Relational Database Model happened in a Moringa plot (Codd, 1970; Paredaens *et al.*, 1989; Lehnert *et al.*, 2000) was used to design the data structure. To avoid confusion, tables and fields names are written with capital letters and single quotation marks. Each table is assigned a table ‘ID’ column which stores the unique number generated when a record is made. The unique number is called the primary key and has automatic incremental value for every record. A table is connected to another through a foreign key as in Figure 3. The figure shows data is stored in records with unique ‘ID’. In this example, ‘ID’ from ‘PLOT’ table is stored in ‘PLOT\_ID’ field in table ‘EVENT\_TIMELINE’ as a foreign key. The foreign key is the unique number identifier of a record from another table. “Planting” and “Applying fertiliser” are the events happened in “Bambara groundnut plot A”. “Harvesting” is the event happened in Moringa plot.

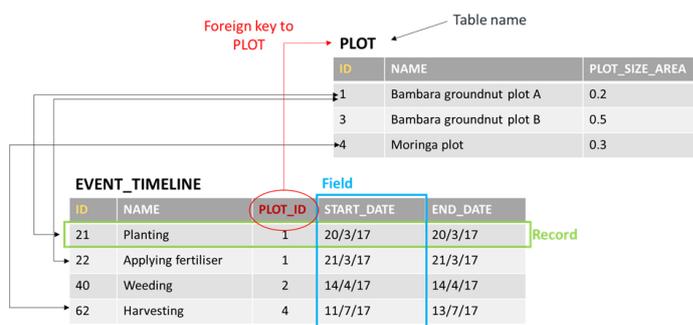


Figure 3. Connection of tables in relational database model.

Every set of tables are grouped to represent a single node in the value chain. The tables are interconnected to

each other through foreign keys. Data was structured in a way that related data from each discipline could be stored and whole data from the journey of seeds to food product could be shown and traced back and forth.

Each table records data in row and column. The row is called record and column is called field as shown in Figure 3. Every record in each table is representing a single entity. Every entity can be linked in a relational format to other entities in other tables. Based on the structure and the logic behind it, the relationship between entities can be one to one, one to many or many to many. Those relations are stored in the database scheme of MySQL database. The relations are designed to be in hierarchy and to follow the story timeline. The physical design which comprises of tables and fields with primary and foreign keys linkages is illustrated in Figure 4. However, due to limited space in the page, only main fields of the tables are shown.

#### 2.4 Development of online web graphical user interface

The database was administrated by MySQL Workbench. MySQL PHP Generator by SQLMaestro was used to connect to the database and automatically generate online web interface “Data Manager Interface” powered by PHP. Another web interface was developed for public access to visually present the data in a user-friendly website.

QR code system was implemented to reflect directly to the web interface. QR code printer was used to print QR code sticker from URL of each data point. The QR code sticker was pasted over the entity that represented the data point. For example, the row that represents a product in a table is displayed in the web graphical interface. The URL is taken to generate QR code which is printed and pasted over this particular product. So, the user can simply scan the QR code using any generic QR code scanner on any phone to be redirected to the URL that represents this product. If the user has sufficient privilege, the user can view/edit/delete the data based on pre-set privileges. The purpose of QR code implementation is to activate the traceability between online data system and physical products, plots, seed and others. Every entity in the project has a QR code sticker and a data point that reflect the information about this entity.

### 3. Results and Discussion

The data model that was developed for this research shows a faithful representation of value chain data from seeds to food product. The model connects ten identified entities from the value chain together through foreign keys in tables. Unlike the hierarchical model which

connects data from one-to-many or from parent to child (Silberschatz *et al.*, 2010), the relational model allows interconnectivity through one-to-one, one-to-many, many-to-one and many-to-many. The connection of entities is quite “linear” in the data model (Figure 2). It represents the significant objects and processes across the value chain. Though in reality, there might be more objects in the value chain as the value chain can conceptually be extended to include many layers and data points. For example, manufacturing of food product might come after food product development. This involves new datasets to be recorded. However, the relational data model presented in this article will be able to cope with new changes. However, there is a limitation in this model where some item descriptions may not be suitable for the object. For example, in this database structure, different extensive tables had to be designed to describe the yield assessment of each crop i.e. Bambara groundnut and moringa. This is because different crop parts (i.e. pods and leaves) with distinctive characteristics of items are considered as harvested parts. To solve this problem apart from generic tables that are designed for all crops, specific tables have to be built per crop types.

Many tools are available for the relational data model to help in designing and developing the database and its user interface. For example, MYSQL Workbench, SQLMaestro PHP Generator and CakePHP are all open source tools that can be used for developing databases web pages and tools for data management and with less coding effort. The latter tools save a lot of time for interface development with limited resources.

The objectives were achieved when the data structure is able to hold the data across the value chain from seed to food products and present them in the desired format. The user can store crop germplasm details in ‘GERMPLASM’ table and store the details of seed availability in ‘SEED\_INVENTORY’ table. The latter is helpful as the seeds manager can make regular check-ups and updates on the availability of seeds in the store chambers. The details of plots and activities on the plot are presented in tables ‘PLOT’ and ‘EVENT\_TIMELINE’. In reality, many events or activities occur in one plot such as sowing, application of fertiliser, pests infection and in rare occasions for example floods. These events or activities should be recorded as the history of plots which are important for end-users. Foreign key ‘PLOT\_ID’ in table ‘EVENT\_TIMELINE’ allows multiple records of events or activities through many-to-one relationship. This design is seen to be more practical as it saves a lot of time for data entry. ICASA Version 2.0 standards developed by International Consortium for Agricultural

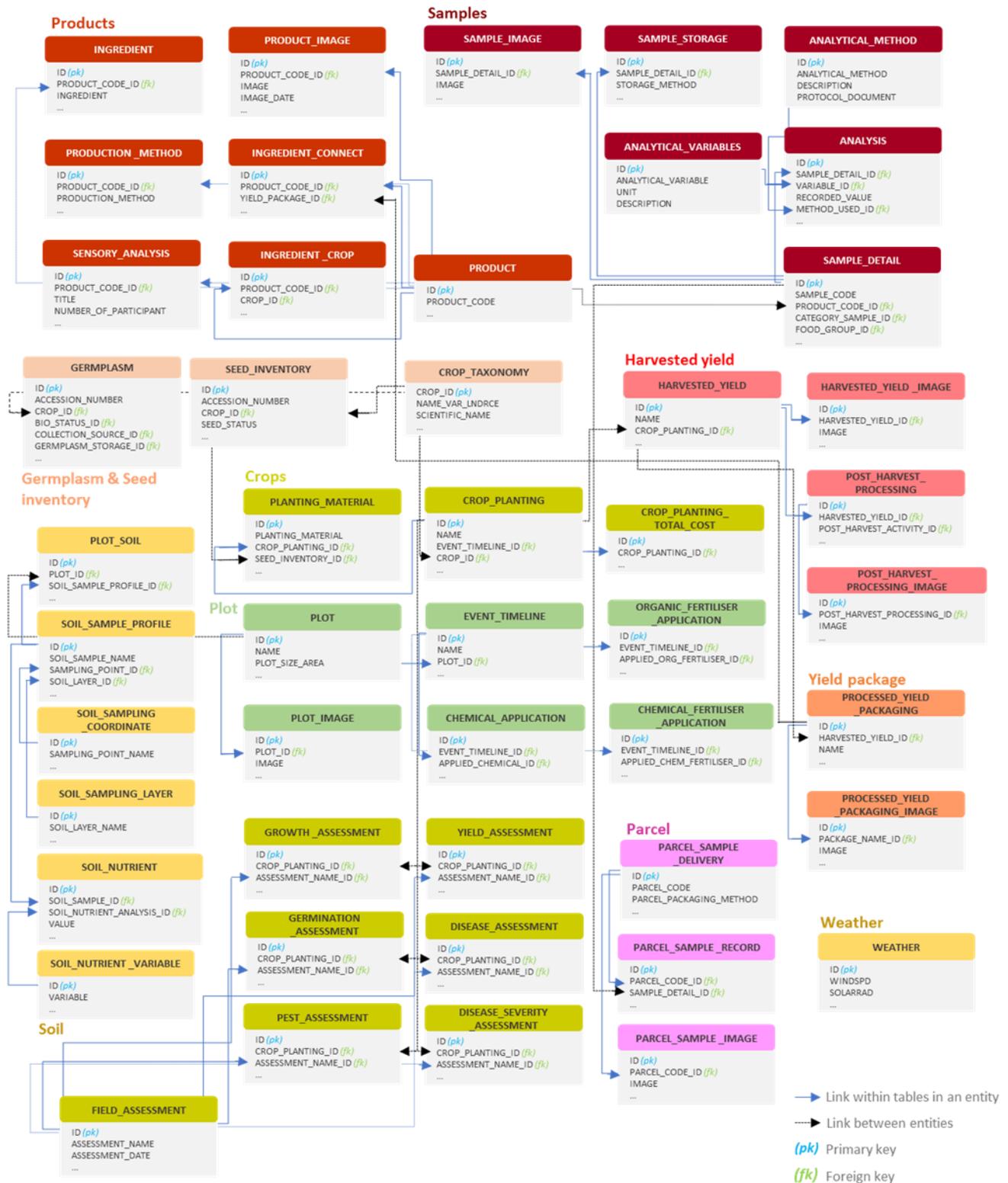


Figure 4. Illustration of linkages of tables and fields in database structure.

Systems Applications (ICASA) provides the closest data structure to be used in this project. However, the design is far too complicated to be used in this project because it records many details in field experiment and therefore time-consuming for data entry (White *et al.*, 2013). ICASA standards are designed to have fine details so that the data could be used in the crop modelling other than documenting experiment (White *et al.*, 2013). Our design is made to store plot historical data and it may not be useful in crop modelling. The data model structure,

however, will need to be extended to include more variables that are essential for both research and production. Different tables were designed for fertiliser and chemical applications so that it could hold an infinite number of records for fertiliser applications done on the plot. The tables consist of foreign key 'EVENT\_TIMELINE\_ID' which connects the record of fertiliser application details to the fertiliser application record in 'EVENT\_TIMELINE' table.

It is a common practice to have multiple crops

grown in one plot. Even though it did not happen in the project, the data structure design is able to hold multiple records of different crops planted in the same plot through foreign key 'EVENT\_TIMELINE\_ID' in 'CROP PLANTING' table. This table is not directly connected to table 'PLOT' instead to 'EVENT\_TIMELINE' so that the details on when the crop is grown is able to be traced. Since the grown crops might be harvested on a different date, table 'HARVESTED\_YIELD' was designed so that multiple yield batches records can be stored for the same crop.

An important achievement that was made in this research is a structure that could hold the evidence of the food products developed by underutilised crops. Table 'PRODUCT' stores the names and description of novel food products developed using Bambara groundnut and moringa. Important linkages to 'INGREDIENT\_CROP' and 'INGREDIENT\_CONNECT' tables allow the indication of crops use and the latter allows connection the whole section of entity "Products" to previous processes. The ingredient of Bambara groundnut and moringa powder can be traced through foreign key 'YIELD\_PACKAGE\_ID' in 'INGREDIENT\_CONNECT' table.

Another important objective of this research that was achieved was to build a structure that can store nutrients values of the food products made using underutilised crops. In reality, many samples of the same product can be tested on different dates and for different nutrient variables. 'SAMPLE\_DETAIL' table was designed to hold multiple records of samples for the same products and is linked by foreign key 'PRODUCT\_CODE\_ID'. Since there are many nutrient variables tested for one sample, 'ANALYSIS' table was designed to hold multiple values of nutrients for the same or different variables. As an example, in this project, Bambara groundnut powder (recognise as a product) was tested for proximate analysis which includes carbohydrate, protein, fat, ash and moisture content variables. Each of the samples from the product was tested three times. Therefore, there should be 15 values for five nutrient variables that should exist for the product. The structure is capable of storing these information and linkage is done through foreign key 'SAMPLE\_DETAIL\_ID' in 'ANALYSIS' table.

The design mentioned above is noticed to be similar to USDA National Nutrient Database for Standard Reference where NUT\_DATA table capable of storing an infinite number of records for the infinite number of nutrient variables similar to 'ANALYSIS' table (United States Department of Agriculture, 2018). The NUT\_DATA table is connected to FOOD\_DES table

which stores the food description through foreign key NDB\_No. 'ANALYSIS' table not only records the values of variables but the testing method, details of the person who recorded the value and details when and where the sample is tested are able to be recorded. This is unique compared to USDA design. Other databases might consist of "straightforward" design where columns of nutrients variables exist in one table. This allows an infinite number of records but a limited number of nutrient variables. AUSNUT 2011-13 food nutrient database is designed this way (Food Standards Australia New Zealand, 2016). As a matter of fact, authors believe this is more practical and easy to visualise, but it will be difficult to trace from where the values were obtained later. Especially for underutilised crops, these details are very important as they provide stronger evidence to

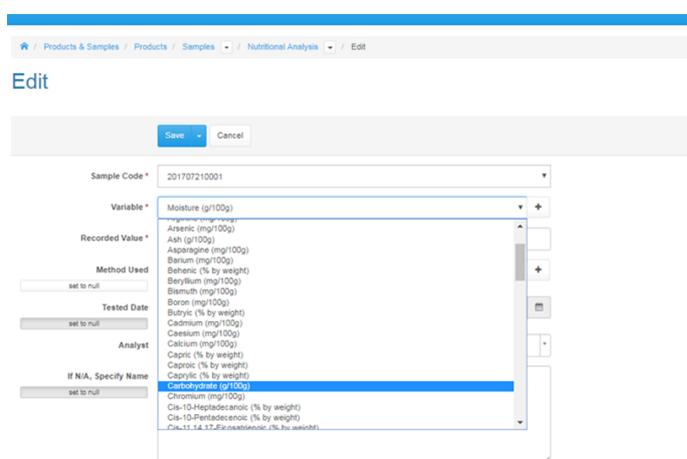


Figure 5. Drop-down menu on database interface which shows linkage between both 'ANALYSIS' and 'ANALYTICAL\_VARIABLES' tables.

promote new crops.

Since the collected data is from the whole value chain from seeds to food products, many departments or users from different fields of expertise were involved in this study. Because this involves many users who might have a lack of expertise or time for entering data, the database interface has to be simple and friendly. For example, the user does not need to type the same data every time, entering the same value or creating a unique ID for each record every time creating a new record. The latter is made easy with an auto-increment function which automatically creates new ID every time a record is saved. 'ANALYTICAL\_VARIABLES' table allows storage of a list of nutrient variables to be tested. The table is linked through foreign key of 'VARIABLE\_ID' in table 'ANALYSIS'. In the front interface, the nutrient variable list can be shown in the drop-down menu in the column called "Variable" as in Figure 5. This helps with the data entry where the user does not need to re-type the same variable again and again.

#### 4. Conclusion

Realising time limitation of project delivery, databases and data structures for the value chain of crops, we set out to build a relatively simple but practical data structure for storing such data. The structure is divided into 10 entities that relate germplasm and seed inventory, crops, plot, harvested yield, yield package, soil, weather, food products, samples and parcel numbers. Relational Database Model was used to design and link the data across the disciplines. The data model successfully linked data from seeds to food products in one accessible system that will not only preserve the data but also provides a one-stop information centre for scientists, farmers and other stakeholders to answer their own questions. This project involved many users from different background with no expertise in data models or with limited time to make data entry. Therefore, the design had to be user-friendly for data collection and management. The design of the data structure allows data in the agronomic practices section to be updated instantly as the activities happen in the plot. The structure holds evidence of product development made from underutilised crops. However, the current structure for nutrient values records is more complicated compared to other databases. This is because more details on the tested nutrient data were needed. Further works are needed to enhance data structure by removing unrelated tables and fields to reduce complexity. The current structure can accommodate value chain data for both underutilised and common crops. In the future, the structure will be extended to include crop specific traits, mass production and marketability of food products to store data on the market price of the product, the location of distribution, an organisation involved and other relevant subjects.

### Acknowledgement

The authors would like to thank Crops For the Future and all colleagues for support and assistance in the research. Especially to those who directly involved in the seeds to food product value chain data collection; Hilda Hussin, Hamdi Hamzah, Kalaivani Balakrishnan, Nur Izzaty Mohamad Jamal, Fadhil Razi, Hairil Hasyimi, Muhammad Zahrulakmal, Advina Julkifle, Tan Xin Lin, Gomathy Sethuraman and developers of applications; Thayanandan Samiappan, Malathi Loganathan and Murshid Omer.

### References

- Alercia, A., Diulgheroff, S. and Mackay, M. (2015). FAO/Bioversity Multi-Crop Passport Descriptors V.2.1 [MCPD V.2.1]. Rome, Italy: Food and Agriculture Organization of the United Nations (FAO) and Bioversity International.
- Batjes, N.H. (2008). ISRIC-WISE Harmonized Global Soil Profile Dataset (Ver. 3.1). Report of 2008/02. Wageningen: ISRIC – World Soil Information.
- Codd, E.F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, 13 (6), 377-387. <https://doi.org/10.1145/362384.362685>
- Codd, E.F. (1972). Further Normalization of the database relational model. In Rustin, R. (Ed). *Database Systems*, p. 33-64. New Jersey: Prentice Hall.
- Date, C.J. (1999). *An Introduction to Database Systems*. Boston, MA, United States: Addison-Wesley Longman.
- Food Standard Australia New Zealand. (2016). About AUSNUT 2011–13. Retrieved on March 7, 2018 from Food Standards Australia New Zealand Website: <http://www.foodstandards.gov.au/science/monitoringnutrients/ausnut/Pages/about.aspx>
- Fuhrer, J. (2003). Agroecosystem responses to combinations of elevated CO<sub>2</sub>, ozone, and global climate change. *Agriculture, Ecosystems and Environment*, 97(1-3), 1-20. [https://doi.org/10.1016/S0167-8809\(03\)00125-7](https://doi.org/10.1016/S0167-8809(03)00125-7)
- Homan, J.V. and Kovacs, P.J. (2009). A Comparison of The Relational Database Model and The Associative Database Model. *Issues in Information Systems*, 10 (1), 209-213.
- Jones, P.G. and Thornton, P.K. (2003). The potential impacts of climate change on maize production in Africa and Latin America in 2055. *Global Environmental Change*, 13(1), 51-59. [https://doi.org/10.1016/S0959-3780\(02\)00090-0](https://doi.org/10.1016/S0959-3780(02)00090-0)
- Leavitt, N. (2000). Whatever happened to object-oriented databases?. *Computer*, 33(8), 16-19.
- Lehnert, K., Su, Y., Langmuir, C.H., Sarbas, B. and Nohl, U. (2000). A global geochemical database structure for rocks. *Geochemistry, Geophysics, Geosystems*, 1(1), 1-14. <https://doi.org/10.1029/1999GC000026>
- Lin, B.B. (2011). Resilience in Agriculture through Crop Diversification: Adaptive Management for Environmental Change. *BioScience*, 61(3), 183-193. <https://doi.org/10.1525/bio.2011.61.3.4>
- Mabhaudhi, T., Chimonyo, V.G.P., Chibarabada, T.P. and Modi, A.T. (2017). Developing a Roadmap for Improving Neglected and Underutilized Crops: A Case Study of South Africa. *Frontiers in Plant Science*, 8, 2143. <https://doi.org/10.3389/fpls.2017.02143>
- Massawe, F.J., Mayes, S., Cheng, A., Chai, H.H., Cleasby, P., Symonds, R., Ho, W.K., Siise, A., Wong, Q.N., Kendabie, P., Yanusa, Y., Jamalluddin, N., Singh, A., Azman, R. and Azam-Ali, S.N.

- (2015). The Potential for Underutilised Crops to Improve Food Security in the Face of Climate Change. *Procedia Environmental Sciences*, 29 (2015), 140-141. <https://doi.org/10.1016/j.proenv.2015.07.228>
- Mayes, S., Massawe, F.J., Alderson, P.G., Roberts, J.A., Azam-Ali, S.N. and Hermann, M. (2011). The potential for underutilized crops to improve security of food production. *Journal of Experimental Botany*, 63(3), 1075-1079. <https://doi.org/10.1093/jxb/err396>
- Moore, F.C., Baldos, U., Hertel, T. and Diaz, D. (2017). New science of climate change impacts on agriculture implies higher social cost of carbon. *Nature Communications*, 8, 1607. <https://doi.org/10.1038/s41467-017-01792-x>
- Napolitano, E., Marchesini, I., Salvati, P., Donnini, M., Bianchi, C. and Guzzetti, F. (2018). LAND-deFeND – An innovative database structure for landslides and floods and their consequences. *Journal of Environmental Management*, 207(2018), 203-218. <https://doi.org/10.1016/j.jenvman.2017.11.022>
- O'Neil, E. and O'Neil, P. (2009). Entity-Relationship Concepts. In Teorey, T.J. (Ed). *Database Design: Know It All*, p. 11-84. Burlington, MA, United States: Elsevier.
- Paredaens, J., De Bra, P., Gyssens, M. and Van Gucht, D. (1989). Relational Database Model. In: *The Structure of the Relational Database Model*. EATCS Monographs on Theoretical Computer Science, Vol.17. Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-642-69956-6>
- Silberschatz, A., Korth, H.F. and Sudarshan, S. (2010). Appendix E: Hierarchical Model. In *Database System Concepts Sixth Edition*. New York, United States: McGraw-Hill. Retrieved from website: <http://codex.cs.yale.edu/avi/db-book/db6/appendices-dir/e.pdf>
- Truswell, A.S., Bateson, D.J., Madafiglio, K.C., Pennington, J.A.T., Rand, W.M. and Klensin, J.C. (1991). *INFOODS Guidelines for Describing Foods: A Systematics Approach to Describing Foods to Facilitate International Exchange of Food Composition Data*. New South Wales, Australia: Academic Press.
- United States Department of Agriculture. (2018). *USDA National Nutrient Database for Standard Reference: SR-Legacy*. Retrieved on September 6, 2018 from United States Department of Agriculture – Agriculture Research Service Website: <https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/nutrient-data-laboratory/docs/usda-national-nutrient-database-for-standard-reference/>
- White, J.W., Hunt, L.A., Boote, K.J., Jones, J.W., Koo, J., Kim, S., Porter, C.H., Wilkens, P.W. and Hoogenboom, G. (2013). Integrated description of agricultural field experiments and production: The ICASA Version 2.0 data standards. *Computers and Electronics in Agriculture*, 96(2013), 1-12. <https://doi.org/10.1016/j.compag.2013.04.003>
- Williams, J.T. and Haq, N. (2000). *Global Research on Underutilized Crops: Assessment of Current Activities and Proposals for Enhanced Cooperation*. Southampton, UK: International Centre for Underutilised Crops.