# Implementation of Information System in Oil Palm Breeding Research: FGV's Experiences

Nurul Syafika Mohamad Fauzi, Muhamad Farid Abd Rahim, Mohd Nasruddin Hj Mohamad

*Breeding Department, FGV R&D Sdn Bhd, Malaysia*

**ABSTRACT**

*FGV is one of the largest plantation companies and crude palm oil producer in the world. Experienced in oil palm breeding research for fifty years since 1960s mainly to improve oil palm yield in term of FFB production and OER. Managing breeding research data is a challenges in any plant breeding research. As technology advancement for the past ten years witnessing the introducing GIS technology, remote sensing, precision agriculture and IoT, information technology plays main role to manage the information in form of application and relational database system to store and linking between information for the long years of research. Over the years, data collected will increase in number and become more difficult to manage. Therefore implementation of technology was used using database management system in order to tackle this problem and has been called FGV Integrated Breeding System (FIBS). The objective is to store collection of data and information of FGV oil palm breeding research such as breeding background, project information, crossing operation and field data. The database was developed using combination of powerful open source web development platform which are Linux as the operating system, Apache as the web server, MySQL as the relational database management system and PHP as the object oriented scripting language. Implementation of information system helps improve data integrity & traceability, reduce data redundancy and provide easy access to information for researcher. In conclusion, as technology rapid growth through the years ahead, towards the precision agriculture and advancement in breeding through biotechnology, advancement and improvement must be made concurrently so no technology are left behind.*

**Keywords:** *Database, Information System, Oil Palm.*

## I. INTRODUCTION

FGV is one of the largest plantation companies and crude palm oil (CPO) producer in the world. The biggest oil palm germinated seed producer in Malaysia cover about 40% market share. Experience in oil palm breeding research for fifty years since 1960s mainly to improve oil palm yield in term of fresh fruit bunch (FFB) production and oil extraction rate (OER). Enormous amount of breeding research data has been generated since then. A dynamic information system is needed to manage and store these data in one database which consequently helps generate meaningful information for research purpose. There are existing commercial research database packages available such as Agrobase® [1]. However since they focused more on annual agricultural crop, it is not suitable to be implemented for oil palm which is a perennial crop which have a different breeding strategy.

With a vast germplasm and oil palm breeding materials collection from different type of origin that have been collected and generated since five decades ago, information storage has been a crucial matter. These information is needed in order for us to continue our lineage conservation and materials improvement project. Improper information storage could lead to loss of valuable information. Hence, these information that come from a series of breeding research program starting from our early years of establishment really need to be stored safely and can be retrieve again easily.

With economical lifespan reach 25 years, oil palm starts yielding about three years after planting and consistently produce bunches since then. Research data each individual palms is being collected by operator since it's been planted. Over the years, data collected will grow exponentially and become a challenge to manage. Unstandardized format between each data collected and not being link between each other is becoming a problem. Therefore introduction of technology using database management system was implement for better data management in more organize way.

Furthermore, various type of data set such as trial information, crossing information and field data is produced in every breeding research. The process of compiling and processing of the data has been done manually in a conventional way is really time consuming and non-effective method. Data collected using conventional way also shows inconsistency and prone to have human error. Information system could speed up the whole entire process giving more extra time

for researcher to focus on other important parts as well as gave data integrity and consistency.

Before the implementation of information system in organizing breeding data is practices, AS/400 system is used. Over the years, this system started to be outdated and have limitations. Maintaining and repairing an old hardware is cost consuming. The interface also is not user friendly and the system used is not relevant anymore with the current amount of data produce this days.

Since FGV breeding research is highly private and confidential, data security is really important and our top priority. By using information technology we can control this issue by giving the permission to access the database system by different level of access depending on the user's role. Low level users can access limited set of information while higher level users can access to more important data on the system. Data leakage could be a disastrous matter if this precautionary act on emphasizing data security issue is not taken into consideration.

As technology advancement for the past ten years witnessing the introducing GIS technology, remote sensing, precision agriculture and the internet of things (IoT), information technology plays main role to manage the information in form of application and relational database system to store and linking between information for the long years of research. Therefore implementation of information system through the usage of our own fully design database called FGV Integrated Breeding System (FIBS) is used.

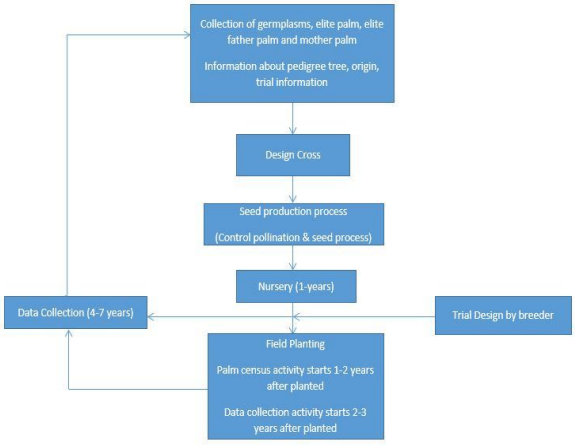


**Fig 1**. Summary of breeding research workflow

## II. METHODOLOGY

### *A. System Development Life Cycle*

System development life cycle (SDLC) is a detailed plan describing the process of developing a fully functioning information system. The process involved six stages which are planning, design, development, testing, deployment and maintenance. This mechanism gave clear view of the entire projects, workers involved, estimated cost and also timeline. By practicing SDLC, the end goal and standard of the desired information system are clearly defined

### *B. Designing database structure*

The main stage in developing a database is designing conceptual information model based on user's need preferences [2]. Thus, database structure is created and design specifically based on the oil palm breeding research workflow. Since breeding program is such a complex process that consist of multiple stages of work and involving large amount of information, the database structure is design by understanding breeding's system requirement first.

Database structure consist of each part of breeding stage of work for the full cycle of breeding research program starting from the process of breeder creating new cross from collection of germplasm and elite palm. Next, operation team will then continue to produce breeder's desirable seed and planted the new planting materials in field. For the next couple of years, each individual palm's performances will be assessed, recorded and analyzed to find the best palm to be commercialized. This entire process is shown in figure 1

### *C. Development of web database application*

Three tier system architecture is used instead of two tier architecture due to the present of database server. Classic two tier architecture only have client/web browser and web server. User request resources from web server and web server will responds with the resources. However, web server is not a sophisticated storage software. Instead another separate database is needed in order to process complicated operations. As in figure 2, three tier architecture consist of client/web browser, web server and database server. User request resources from web server then database server will generate the resources and respond to the request. Database server also is needed to support services such as data backup.

A powerful open source web development platform called LAMP stack was used to create our web database. LAMP stands for Linux, Apache, MySQL and PHP. The four software components used in web development. This platform use Linux as the operating system,

Apache as the web server, MySQL as the relational database management system and PHP as the object oriented scripting language.

Web server software called Apache HTTP server or commonly known as Apache was used. Apache serves webpage over the internet using HTTP protocol.

Web scripting was done using PHP: Hypertext Pre-processor, a scripting language used for web development. PHP is preferable compared to other scripting language due to some advantages such as it is an open source software, flexible for integration with HyperText Markup Language (HTML), suited to complex projects and fast at running script [4].

Software tool MySQL was used to manage relational database management system based on Structured Query Language (SQL). SQL is a widely used, standardized language for manipulating relational database [3]. SQL also used for analyzation and manipulation of relational data. MySQL which is an open source software was used instead of others due to the rapid connection handling and high performances. MySQL globally known for their secure and reliable database management system and this is prove when it is used in some popular web applications such as WordPress and Facebook

### *D. Hardware*

Both of our web server and database server is placed at FGV data centre. The installation of web server and database server on different machines is due to fast application and security issue. It is secure due to the fact that the database server is not link to the internet. Any improvement maintenance will be done only at web server. If there is failover, the database server is still safely function as it is separated from web server.

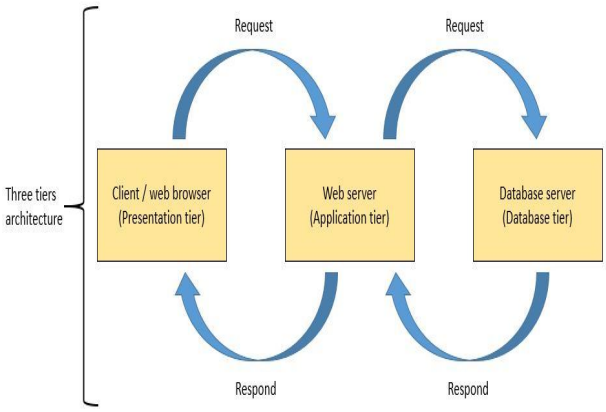


**Fig 2**. Three tier architecture model of web database application

## III. RESULTS

FGV Integrated Breeding System (FIBS) was build based on three main component which are information management, operation management and field data management (figure 3). Each component contain several modules and interrelates with each other. Information management holds core information in breeding research data such as germplasm origin and trial information. Two other component able to function smoothly based on this important information. Next, operation management holds complex workflow breeding process starting from seed production until field planting. Lastly, field data management holds data collection and data analysis.

This system was designed to integrate the breeding activities information carried by FGV and to store large amount of information such as progenies, planting materials, trial designs, seed production workflow, seed store, pollen store and various field data. Integration of breeding data provides comprehensive data management system. Since this system was design thoroughly looking at the breeding workflow, all of the breeding activities can be monitored closely. Breeding activities are automatically tracked in work flows.

A user friendly web interface was created to facilitate the process of retrieving data from the database (figure 4). Breeders could get information from specific trial they want such as progeny information or crossing matrix easily from FIBS. Quick search can be used to retrieve desired information by entering a keyword. If user enter a keyword of progeny code, the complete information involving the progeny code will be generated.

Lineage visualization of specific individual palm can be generated by using data set of progeny information. Progeny tree is construct and linked back to its parental progeny. A five decades of oil palm breeding program could span across three to four breeding cycles making complex relationship between progeny created. With this features, the lineage of oil palm's progeny can be visualized easily. Figure 5 shows progeny tree of TK221 that has been visualized using the database system. From the result, user can clearly see that progeny TK221 came from four breeding cycles. This helps user to see the relatedness between progenies.

Seed inventory is one of the useful features in this system database. The process of tracking seeds and pollen become much easier compared to the conventional way. Once breeder release their crossing matrix, field operation team will do bagging process to get the pollen and proceed with control pollination to produce seed. Produced seed will be kept in seed store until there is a request to be germinated. First-in, first-out method is practices here. Older seeds will sent

out from the store room first. This is to ensure the viability is still high. High viability needed in assuring success rate of the germinated seeds.

Various type of field data stored in FIBS database such as bunch analysis, yield recording, vegetative measurement, bunch census, flower census, gas liquid chromatography (GLC) result, leaf sampling analysis and palm census. Centralized data storage helps avoid data redundancy problem. Data collected can be monitored closely by yearly basis. Simple analytical tools are available to visualize the data and generate meaningful information. Auto data analytics for trials reduces manual operation in data analytics process and helps reduce human error problems. This result then can be easily exported into spreadsheet file format (.xls) or comma separated values file format (.csv). Figure 6 showing summary of field data generated by FIBS.

All of the uploaded data on specific trial or workflows required direct approval from the person in charge of the program. The approval is needed for the user to allow them continue entering information on to the next process. For example, in seed production workflows, a step of process that have been updated by inserting all required information will need to be approved by person in charge in order for the user continue updating on to the next processes. Step 1, palm and flower status must be approved first in order to allow user to move to the next steps which is step 2, pollen process. Approval system in stages helps avoiding mistakes in early stages.

Data standardization is the major advantage that comes from the usage of system database. This is because all of the collected data will be organized in one specific format. A common format or standard types of data is needed in order to do data analyzation accurately. Different types of data format is happen during our time using conventional way of collecting field data and this is really troublesome as the process of data analyzation is interrupted. Standardization helps create data set that are high in quality and consistency.

This system was design to be a multi-platform web based application that able to run on any operating system with a modern browser. A single session login which allow user to login from only one session was implemented. A role based access control was assigned making system users are control by roles with different permission sets in different project involved. Another good security practice implement in this system is logging history which can monitor the logging of system activities. All of this security precaution steps makes the stored information is more reliable.

This website database has a responsive web design which the display is adaptable to the size of screen. Users can view nicely either using desktop, tablet or mobile phone. It is design to be mobile friendly. This is really helpful since our research involving field trial, researcher could get direct information from FIBS at instant directly from the field.

System dashboard is used to show the overview of real time breeding activities. This helps user to get information in a glance as it is designed to be easy to read within a short range of time. The important figures in breeding activities such as the number of total projects has been done, the number of total trials being conducted, the number of total progenies has been produced and the number of total palms have been planted is shown in this graphical user interphase. All of this figures will automatically shows current value as it is synchronized with the database.

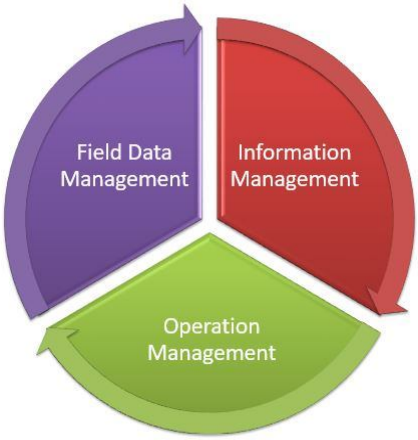


**Fig 3**. Three big parts in FIBS

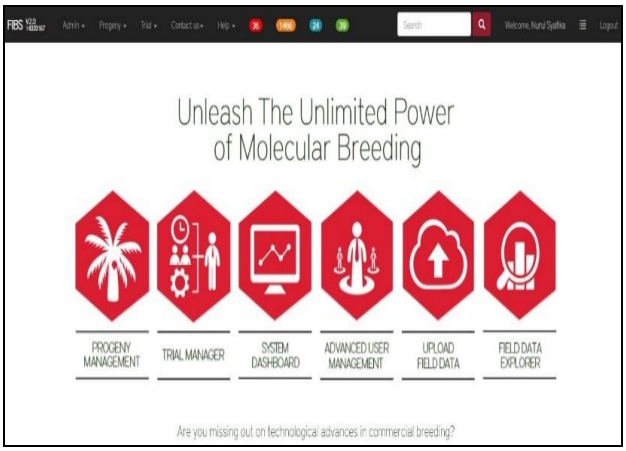


**Fig 4.** FGV Integrated Breeding System (FIBS) database user interface

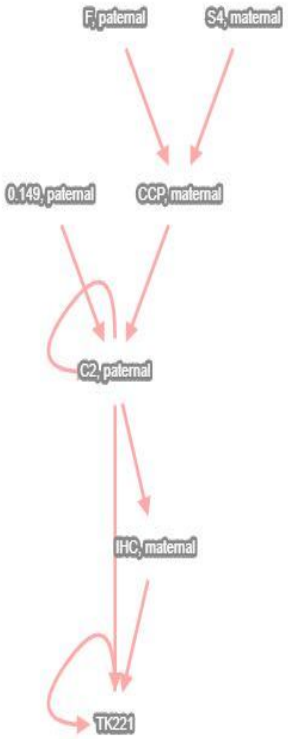


**Fig 5**. Progeny tree of TK221 constructed by FIBS database

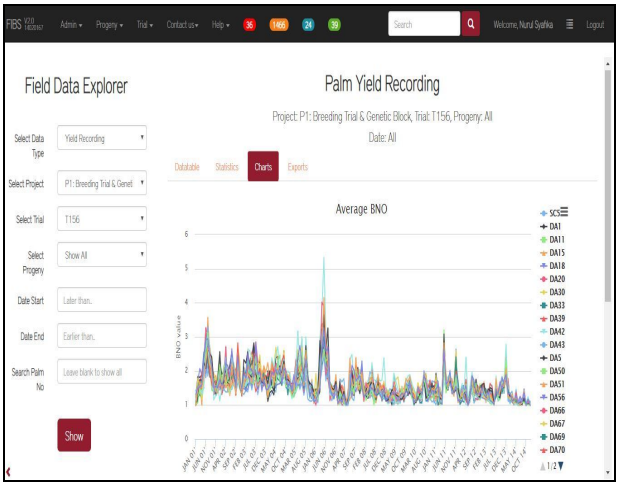


**Fig 6.** Summary yield data of bunch number over 14 years generated by FIBS

## V. CONCLUSION

FGV Integrated Breeding System (FIBS), a carefully design database system for breeding activities that able to store collection of data and information of FGV oil palm breeding research such as breeding background, project information, crossing information and field data. In moving toward digitalization, the implementation of information system through FIBS usage helps improve data integrity & traceability, reduce data redundancy and provide easy access to information for researcher. The application of information system in breeding research is seen can be further improve by integrating with other field of research such as genomic or upgrade with the latest technology to provide the best solution to oil palm breeding research.

## ACKNOWLEDGEMENT

This research was fully funded by FGV R&D Sdn Bhd. This writer would like to express special thanks of gratitude to all parties that have been helping in the making of this paper especially fellow researcher from breeding department of FGV R&D.

## REFERENCES

[1] Evans K., Sook J., Lee T. et al. "*Addition of a Breeding Database in the Genome Database for Rosaceae.*" Database, Vol. 2013

[2] L Chudinov et al. "*The methodology of database design in organization management systems*" IOP Conf. Series: Journal of Physics: Conf. Series 803 (2017)

[3] Maslakowski Mark and Butcher Tony. "*Sams Teach Yourself MySQL in 21 Days*" Indianapolis, IN: Sams Publishing. 2000

[4] David Lane, Hugh E. Williams (2004, May) "*Web Database Applications with PHP and MySql***"**, 2nd Edition. Chapter 1 Database Applications and the Web. Available: https://www.oreilly.com/library/view/web-database-applications/0596005431/ch01.html

[5] David T. Bourgeois (2014, February) "*Information Systems for Business and Beyond. Chapter 4: Data and Databases***"**. Available: https://bus206.pressbooks.com/chapter/chapter-4-data-and-databases/

[6] Ronald G. Wolak. (2001, April) "*System Development.*" Available: http://codecourse.sourceforge.net/materials/System-Development-Life-Cycle.html