Economic and Technology Management Review, Vol. 13(2018): 111 – 120

Development of dashboard to identify rice germplasm in MARDI Genebank

(Pembangunan papan pemuka bagi mengenal pasti germplasma padi di Bank Gen MARDI)

Muhammad Izzat Farid Musaddin*, Faizah Patahol Rahman*, Site Noorzuraini Abd Rahman**, Mohd Shukri Mat Ali @ Ibrahim***, Azuan Amron**** and Mohd Shafiq Azizan*

Keywords: dashboard, Extract, Transform and Load (ETL), database, rice

Abstract

Information on local rice accessions is stored in a system called Agrobiodiversity Information System (AgrobIS) to solve issues on conservation. However, due to the in-comprehensiveness of the search function embedded in the system, it is lack of data filtering feature to obtain specific accessions. This becomes a cumbersome task for researchers who are keen to acquire rice accessions, which are filtered based on selected trait values as they are required to go to the genebank and use a separate system named Rice Genebank Information System (RGBIS) that contains data filtering features. However, the RGBIS is a desktop application that has its limitation because it can only be accessed through a computer after it has been installed rather than being able to use it online. This paper describes the development of a dashboard that is able to provide information on rice germplasm focusing on high-yield, pest and diseases, and quality and specialty. The approaches taken were gathering requirements, dashboard designing, Extract, Transform and Load (ETL) process and the utilization of visualisation tools. The developed MARDI Rice Genebank dashboard is accessible online which allows users to search and retrieve information on specific accessions based on selected traits.

Introduction

The Malaysian Agricultural Research and Development Institute which was established in 1969 was mandated to carry out R&D on rice as one of its main tasks to generate new technologies that can enhance the rice industry (Rosiah 2014). In 1989, a rice genebank was set up in MARDI Seberang Perai, Penang. Local rice seeds were gathered through expeditions around the country while foreign seeds were accessed through collaborative work or exchange programmes with other external rice research centres or institutes. During the early stage, all the rice data collected were kept either on paper or stored on researcher's individual computer. Thus, data were scattered across multiple research sites in MARDI. This situation led to serious problems in accessing the rice research

*Information Management Centre, MARDI Headquarters, Persiaran MARDI-UPM, 43400, Serdang, Selangor
**Genebank and Seed Research Centre, MARDI Seberang Perai, Kepala Batas, Penang
***Genebank and Seed Research Centre, MARDI Headquarters, Persiaran MARDI-UPM, 43400, Serdang, Selango
****Corporate Communication and Quality Centre, MARDI Headquarters, Persiaran MARDI-UPM,
43400, Serdang, Selangor
E-mail: izzatf@mardi.gov.my

data when the hard copies of records were missing or when the researchers were no longer working in MARDI. The lack of mechanism to gain access to these data hinders or prevent users from rationalising, identifying specific accessions and understanding the general characteristics of the seeds collected.

With a huge number of seeds collected over the years, the amount of data available has increased considerably. Therefore, MARDI anticipates the need to come out with a data management system because managing information related to accessions in germplasm collections is an integral part of our genetic resources conservation efforts as suggested by (Blixt 1984). The development of an efficient information system would help rice genebank curators to better manage germplasm activities such as collection, preservation, regeneration, distribution and their exchange. In view of this, MARDI has established the Rice GeneBank Information System (RGBIS) in 2002. However, the database has limited access with a stand-alone feature. With the advancement of Information and Communications Technology (ICT), AgroBiodiversity Information System (AgrobIS), a web-based system was developed in 2006 (Azuan et al. 2016). All the data from RGBIS system were migrated to this new system. Apart from managing rice collection, AgrobIS also manages data for other Plant Genetic Resources for Food and Agriculture (PGFRA) collections namely fruits, vegetables and medicinal plants. In the year 2013, under the 10th Malaysian Plan, an upgraded version of AgrobIS was introduced to cater for livestock and biotechnology components. Besides, data management for PGFRA has also been expanded to include new components namely floriculture, palms and tubers.

The use of data management systems in rice germplasm collections have increased over the years but much still needs to be done to improve their potential use. To date, there are more than 13,000 accessions collected on rice. The collections maintain valuable materials but efficient ways of accessing information and the materials are often lacking. As for the case of AgrobIS, much emphasis was given on developing the system for data capturing but lacked-comprehensive searching capabilities and analysis to be utilised by users particularly those interested in rice research. For instance, the search feature available in the old system only covers filtering of accessions based on certain criteria such as accession numbers and the typing of keywords in the search box provided. This basic search criteria and filtering of accessions is certainly insufficient. Researchers tend to require more significant information that contains specific values for traits like Panicle Length and 1,000 grain weight of rice accessions. Furthermore, in normal situations, users often relied on the curator to obtain more information on the accession they wanted in order to carry out their research. The genebank curator will browse through the rice information from the list of accessions available in AgrobIS, download relevant information and subsequently pass it on to the person who requests the information, thus creating additional work to the curator. Besides, researchers may drop by at the Genebank Centre to use Rice Genebank Information System (RGBIS) instead of AgrobIS system due to its additional functionality which is inclusive of data filtering based on traits and characteristics selection. However, since RGBIS is a desktop application and not a web-based application, it requires the researchers to always visit the Genebank when there is a need to get the accessions. Therefore, a more user-friendly online application should be developed and made available to Genebank users. MARDI researchers will no longer be required to visit the Genebank when collecting rice accessions because they can easily access the application via online. This enables users to mitigate the tedious process experienced by the earlier researchers. Thus, this situation led to the idea of creating the MARDI Rice Genebank Dashboard which acts as a self-service site that provides improved accession searching criteria which is user friendly, easily accessible and high usability.

Dashboard

Technically, a dashboard is a data visualisation tool which consolidate and arrange numbers, metrics and sometimes performance scorecards on a single screen (Rouse 2010). A dashboard comes with an interactive interface that depicts information that is garnered in databases. The basis of interactive interface is to incorporate information visualisation in the context of providing information to the end users which assists them in making better decisions in a more rapid manner (UKEssays 2013). In dashboards, data in the form of attributes and dimensions are plotted against data graphical representation symbols like charts and graphs. These graphical representations of data will easily help an organisation or researchers to have a general view on current performances as the charts and graphs are interactive with each other. This is due to the presence of selectors and filter functions available in dashboard development tools. In addition, a user friendly dashboard will assist in decision making for future accomplishments.

There are two types of dashboards that can be used namely, the analytical and operational dashboards (Bustos 2017). Analytical dashboard focuses on large volume data which consists of historical data for specific area of interest. The advantage of having a big amount of historical data is that users will be able to derive predictions of future data trends when implementing time series models on the data. On the other hand, operational dashboard is a visualization dashboard that provides users a-glance view of an organisation's core business area as suggested by Tröster (2016). In the context of AgrobIS data, the type of dashboard suitable for development is the operational dashboard as it prioritises the focus on the performance of rice data. Precisely, the amount of rice data is an important indication of how vastly MARDI's research are being conducted.

A dashboard with friendly interface and search features was developed using DevExpress, a dashboard development tool by Microsoft. Since rice germplasm information system was first established as compared to other collections, the rice germplasm is the most suitable crop to start the dashboard facility with. This facility will help rice researchers to look for information faster to carry out their research for new rice varieties.

Methodology

The process of developing the MARDI Rice Genebank Dashboard involves various steps namely, gathering requirements, dashboard designing, Extract, Transform and Load (ETL) process and the Utilisation of visualisation tools.

Requirements gathering

In any system development, requirements gathering is the most important aspect that should be taken seriously by project members and subject matter experts (SME). Requirements gathering activity is one of the components of the Software Development Life Cycle (SDLC) as shown below in *Figure 1*.

The SDLC is a process of developing or enhancing a system that solves problems in the form of system development. System development can occur in either two scenarios. The first scenario is to develop a totally new system according to requirements of end users. While the second scenario is to alter or enhance an existing system for better performance. Altering or enhancing a system occurs when there is a need to add new modules in the existing system or even applying changes towards the process flow of a system. Both the situations above will follow the flow of SDLC starting from requirements gathering until the maintenance activities of a particular system.

The SDLC framework also applies to dashboard development, as it is a type of system development. The activity of collecting requirements with the process owners is a very important phase in creating a dashboard as developers will be able to get a clear picture on the user's needs. By implementing this step, it will reduce the risk of having a failed project. Failure of a project is bound to happen if the requirements of users are not gathered and understood accordingly (Whitehorn 2012). In order to understand the requirements of the dashboard, several workshops were conducted with the genebank curator or normally addressed as the owner of the dashboard. The genebank curator explained the process of delivering information to the requester. Dashboard screens and the underlying data structures were designed by the development team upon understanding the process. All requirements were gathered and designs were then documented for future reference.

Dashboard design

After the requirements were successfully gathered, the next activity to be carried out was designing the dashboard. It specified which data to be shown and which charts or graphs it is accommodated to. The design of dashboards are usually documented in the form of storyboards. Developers create storyboards to indicate the flow of the dashboards which also includes other features such as selectors and searching capabilities. MARDI Rice Germplasm Dashboard comprises of three focus areas which are high yield-related traits, pest and diseases and quality and special traits.

High yield-related traits dashboard concentrates on qualitative and quantitative descriptors that illustrates the state of rice yield whether being high yield or low yield (*Figure 2*). On the other hand, pest and diseases dashboard as shown in *Figure 3* demonstrates the top three occurring pest and diseases in MARDI rice fields namely bacterial leaf blight, blast and brown plant hopper 1. By having this dashboard, user will be able to identify accessions that are susceptible or resistant towards the pest and

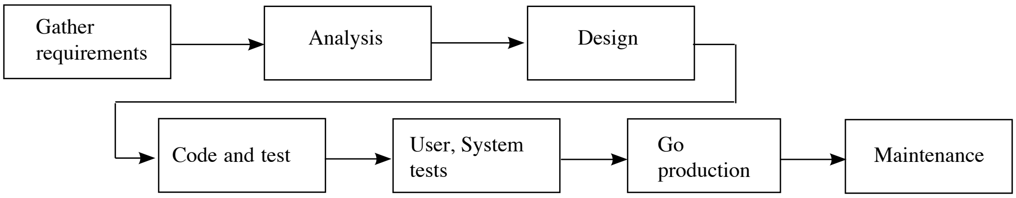
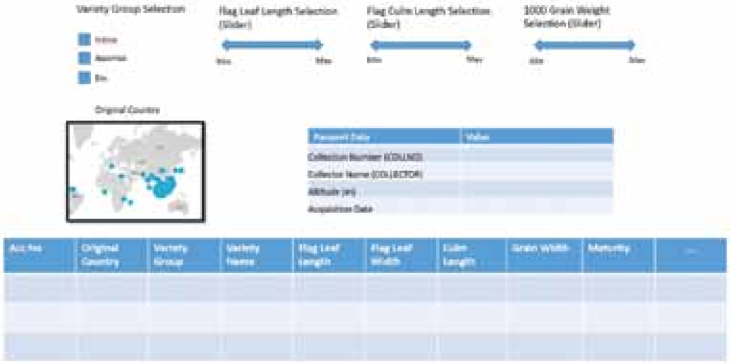


Figure 1. The stages and flow of the software Development Life Cycle


Variety Group Selection
Flag Leaf Length Selection (Slider)
Min
Max
Flag Culm Length Selection (Slider)
Min
Max
1000 Grain Weight Selection (Slider)
Min
Max
Original Country

Passport Data	Value
Collection Number (COLLNO)	
Collector Name (COLLECTOR)	
Altitude (m)	
Acquisition Date	

Acc No	Original Country	Variety Group	Variety Name	Flag Leaf Length	Flag Leaf Width	Culm Length	Grain Width	Maturity	...

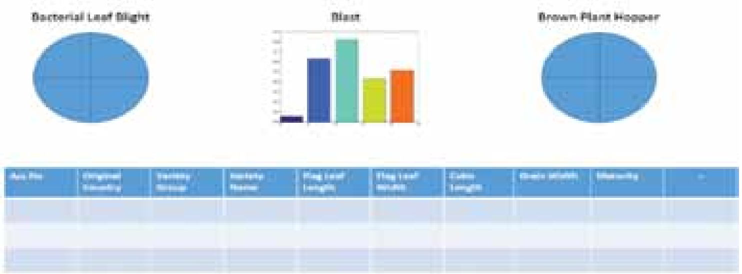
Figure 2. High yield-related traits dashboard design

diseases. Meanwhile, quality and specialty traits dashboard gives more emphasis on the quality characteristics as shown in *Figure 4*.

Extract, Transform and Load

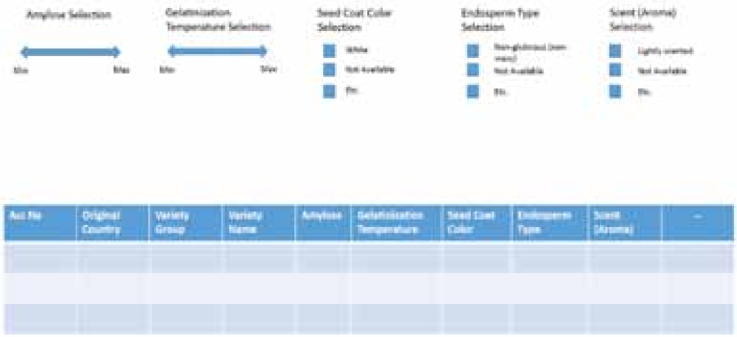
Data is the key to every working dashboard. There are two types of data taken into consideration when developing a dashboard namely structured data and unstructured data (Taylor 2017). Structured data is easily obtainable and searchable as the data are stored in a well-organised database. Although unstructured data is important, it is unorganised and difficult to retrieve. For example, unstructured data can be found in video files, audio files and Microsoft word files (Beal 2017). However, to get a better dashboard performance, data transformation is required in order to make full use of unstructured data. Data in AgrobIS system is labelled as unstructured data since the data cannot be used directly but requires transformation.

The phass of Extract, Transform and Load (ETL) is a data transformation technique commonly applied in dashboard development. ETL is a process where data is retrieved from data sources, transformed and then transferred into a data warehouse. The Extract phase is a stage where data extraction occurs which could be taken from a variety of data sources. Examples include flat files and database of existing systems. Once all data has been extracted, the transformation activity can be executed. Transformation is ideal in configuring a standard database structure which is important to subdue multiple instances of the same fields which in the end helps smoothes up the loading process. The Talend Open Studio data integration tool was used to perform ETL process. This tool covers all phases of ETL which includes, extracting, transforming and loading of data into data warehouse (Bowen 2012). Besides, it can generate batch files which is used in schedulings software to automatically load data into the data warehouse. By this, users will always view an up-to-date dashboard. Lastly, the transformed data is then transmitted into a database known as data warehouse.



Acc No	Original Country	Variety Group	Variety Name	Flag Leaf Length	Flag Leaf Width	Culm Length	Grain Width	Maturity	...

Figure 3. Pest and diseases dashboard design



Acc No	Original Country	Variety Group	Variety Name	Amylose	Gelatinization Temperature	Seed Coat Color	Endosperm Type	Scent (Aroma)	...

Figure 4. Quality and specialty traits dashboard design

Visualisation tool

Once the data warehouse has been established and the transformed data has been loaded, the dashboard creation comes into place. DevExpress, a dashboard visualisation tool, is a third party library embedded in Microsoft Visual Studio to design and develop dashboards. Normally, data from the data warehouse is uploaded into the visualization tool based on query statements that request for specific data from the database. *Figure 5* shows a print screen image of the microsoft visual studio tool code design view.

Issues and challenges

The main issue faced during the development of the MARDI Rice Genebank dashboard was the complexity of the database design in AgrobIS.

Database design

The current AgrobIS database contains 62 tables whereby the architecture is very complex. Thus, it is very difficult to use the current data for data visualisation as it incorporated different genetic resource germplasms. Technically, it is difficult to acquire the preferred data that has undergone a filtering process with just a mere query statement. It requires effort in using the external method~~s~~ namely, ETL process.

Each germplasm has their own tree design which represents a parent child relationships between categories. For instance, the rice structure encompasses categories which are Agro Crops, Crop Grouping, Crops, Crop Genus, Crop Species and Crop Accessions. Agro Crops is the highest level parent whereas Crop Accessions is the lowest level child of the rice tree structure. Consequently, manipulating the data in AgrobIS database for dashboard development is not a straight forward process and thus, requires data transformation.

Results and discussions

The Rice Genebank Dashboard is accessible only by MARDI's researchers which resides in Anjungnet portal, MARDI's main portal. The dashboard is designed to display information based on rice breeding research objectives which includes development of pest & diseases resistance, high yielding and special varieties. Each breeding research objective will be represented by a different tab in the dashboard. The screenshot of High yield-related traits dashboard is shown in *Figure 6* which is the outcome of the dashboard design in *Figure 2*.

Figure 7 depicts the dashboard version of the design illustrated in *Figure 3* whereas, *Figure 8* depicts the screenshot of quality and specialty traits dashboard, which is

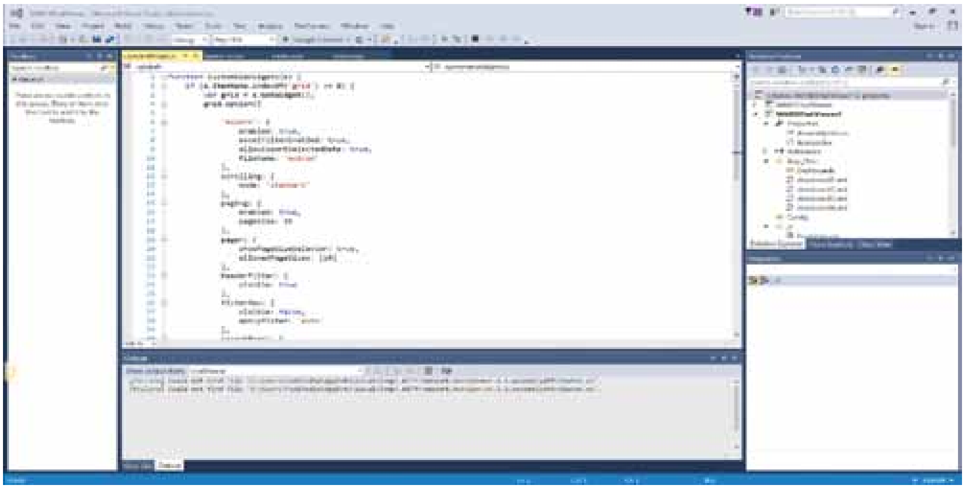


Figure 5. Snapshot of microsoft visual studio code design view

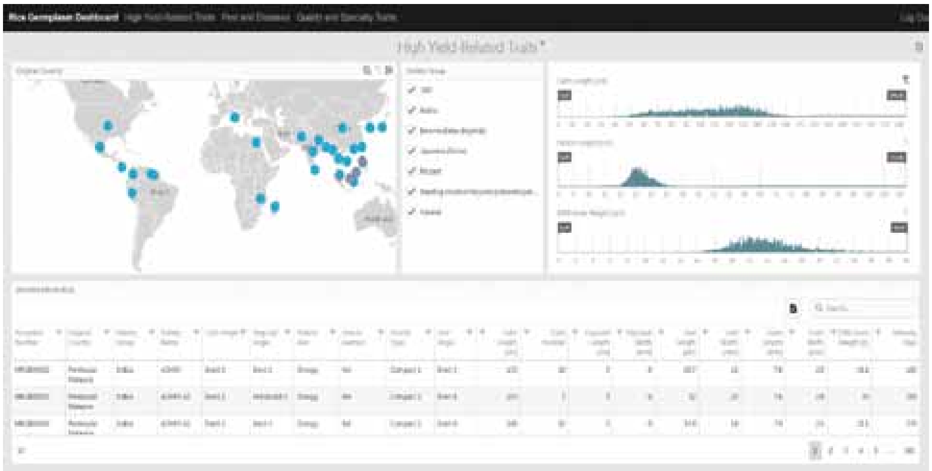

Rice Germplasm Dashboard
High Yield-Related Traits
Pest and Diseases
Quality and Specialty Traits
High Yield-Related Traits

Figure 6. *Screenshot of high yield-related traits dashboard*

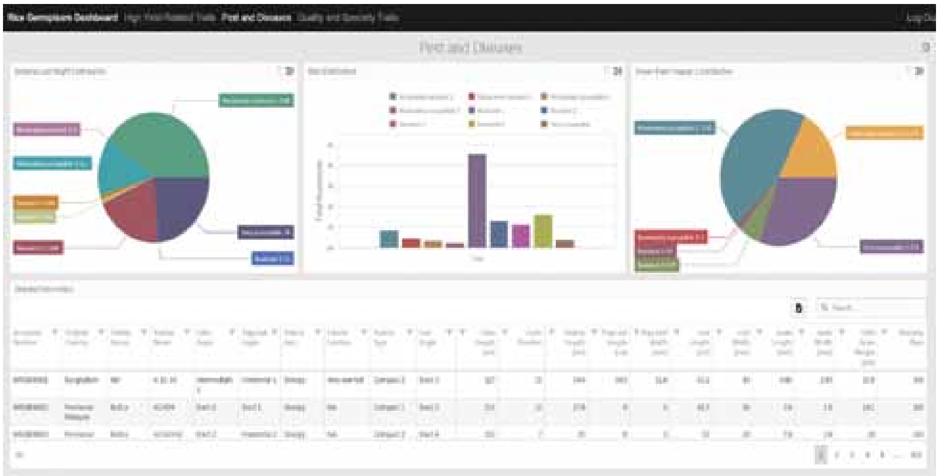


Figure 7. *Screenshot of pest and diseases traits dashboard*

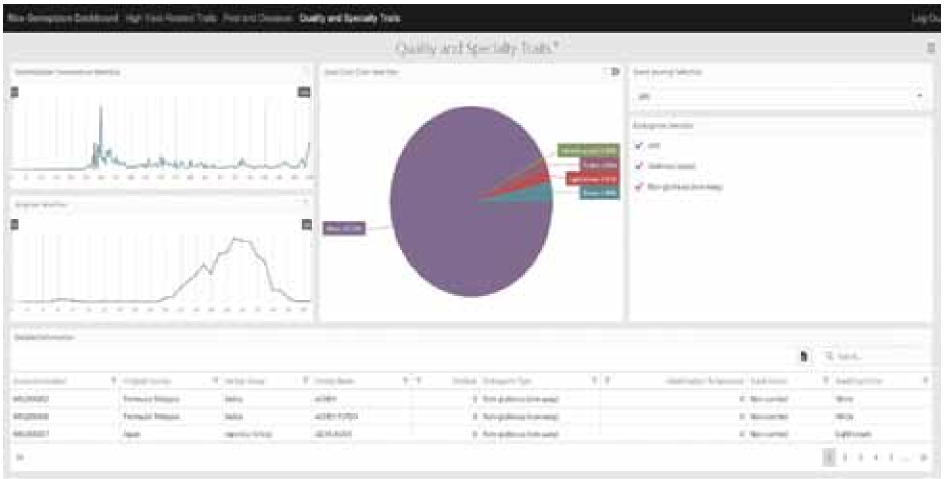

Rice Germplasm Dashboard
High Yield-Related Traits
Pest and Diseases
Quality and Specialty Traits
Quality and Specialty Traits

Figure 8. *Screenshot of quality and specialty traits dashboard* the end product of the design in *Figure 4*. The system has an extensive searching capability either as a wild search or specific search at every descriptor column. Besides, the presence of selectors of different characteristics allow users to filter the table with detailed information in order to obtain specific accession data. The data can also be downloaded as a pdf or excel file.

Some of the available fields used for data filtering and fields that represent columns in the output table in each dashboard scenario are shown in *Table 1*. In the high yield-related traits dashboard, the data field for selectors may include original country, variety group, culm length, field length and 1,000 grain weight. In the

Table 1. The selectors and and fields in high yield-related, pest and diseases and quality and specialty dashboards

Name of dashboard	Data fields
High yield-related traits	**Selectors**
	Original country
	Variety group
	Culm length
	Field length
	1,000 grain weight
	Columns in table
	Accession number
	Original country
	Variety group
	Variety name
	Culm angle
	Flag leaf angle
	Panicle axis
	Panicle exertion
	Panicle type
	Leaf angle
	Culm length
	Culm number
	Flag leaf length
	Flag leaf width
	Leaf length
	Leaf width
	Grain length
	Grain width
	1,000 grain weight
	Maturity days
Pest and diseases	**Selectors**
	Bacterial leaf blight
	Blast
	Brown plant hopper 1
	Columns in table
	Accession number
	Original country
	Variety group
	Variety name
	Culm angle
	Flag leaf angle
	Panicle axis
	Panicle exertion
	Panicle type
	Leaf angle
	Culm length
	Culm number
	Flag leaf length
	Flag leaf width
	Leaf length
	Leaf width
	Grain length
	Grain width
	1,000 grain weight
	Maturity days
Quality and specialty	**Selectors**
	Gelatinisation temperature
	Amylose
	Seed coat colour
	Scent (aroma)
	Endosperm
	Columns in table
	Original country
	Variety group
	Variety name
	Amylose
	Endosperm type
	Gelatinization temperature
	Scent aroma
	Seed coat colour

(*cont.*)

Table 1. (*cont.*)

pest and diseases dashboard, the selectors include bacterial leaf blight, blast and brown plant hopper 1. Finally, in the quality and specialty dashboard, the selectors include gelatinization temperature, amylose, seed coat colour, scent (aroma) and endosperm. The fields that act as selectors in the dashboards will filter the table outcome that results in obtaining specific rice accessions.

Conclusion

Agrobiodiversity conservation is very important for future generation. Having an ICT application in place to capture these data is just as important. In conclusion, the developed online MARDI Rice Genebank dashboard is an important tool that enable users especially rice breeders to further search specific accessions based on selected traits as it will assist them in breeding accessions of good traits.

Acknowledgement

Special thanks to Madam Rosna Amir for her guidance and valuable comments. I am grateful to staffs of Genebank and Seed Centre for their cooperation in the workshops held for the requirements gathering activity.

References

Azuan, A., Muhammad Izzat Farid, M., Mohd Shukri, M.A., Faizah, P.R., Muhammad Luqman Hakim, M.F., Mohd Saifuddin, Z. and Rusli, A. (2016). *Sistem Pangkalan Data Agrobiodiversiti (AgrobIS) Ver. 2: Sejarah dan Sumbangan National Agrobiodiversity Conference 2016, Kuala Terengganu*

Beal, V. (2017). Unstructured data. Retrieved on 7 Dec. 2017 from https://www.webopedia.com/TERM/U/unstructured_data.html

Bilxt, S. (1984). *Application of computers to genebanks and breeding programmes. In "Crop Breeding, a Contemporary Basis"* (P. V. Vose and S. G. *Blixt*, Eds.). Pergamon Press, Oxford. Cited by Committee on Managing Global Genetic Resources. Agriculture Crops Issues and Policies National Academy Press

Bowen J. (2012) *Getting started with talend open studio for data integration*. Packt Publishing Ltd.

Bustos, A. (2017). Operational and analytical dashboards – key differences. Retrieved on 26 Dec. 2017 from https://www.business2community.com/business-intelligence/operational-analytical-dashboards-key-differences-01757479

Rosiah, H. (2014*)*. Sharing of knowledge on rice technology through publication. *Economic and Technology Management Review* 9b: 181 – 192

Rouse M. (2010). What is business intelligence dashboard? Retrieved on 7 Dec. 2017 from http://searchbusinessanalytics.techtarget.com/definition/business-intelligence-dashboard

Taylor, C. (2017) Structured vs. unstructured data. Retrieved on 7 Dec. 2017 from https://www.datamation.com/big-data/structured-vs-unstructured-data.html

Tröster, H. (2016). Make sure you know the difference between strategic, analytical, operational and tactical dashboards*?* Retrieved on 7 Dec. 2017 from https://www.datapine.com/blog/strategic-operational-analytical-tactical-dashboards/

UKEssays. (2013). The advantages of information visualization information technology essay. Retrieved on 4 Oct. 2018 from https://www.ukessays.com/essays/information-technology/the-advantages-of-information-visualization-information-technology-essay.php?vref = 1

Whitehorn, M. (2012). *How to gather BI dashboard user requirements to nail business strategy alignment*? Retrieved on 7 Dec. 2017 from http://www.computerweekly.com/feature/How-to-gather-BI-dashboard-user-requirements-to-nail-business-strategy-alignment

Abstrak

Maklumat berkaitan aksesi padi disimpan dalam satu sistem yang dikenali sebagai *Agrobiodiversity Information System* (AgrobIS) bagi tujuan menyelesaikan isu berkaitan pemuliharaan padi. Namun, oleh kerana fungsi pencariannya yang tidak komprehensif, sistem ini tidak mempunyai fungsi penapisan data bagi mendapatkan maklumat berkaitan aksesi yang spesifik. Keadaan ini akan merumitkan para penyelidik padi yang berminat untuk mendapatkan maklumat berkaitan aksesi padi yang ditapis berdasarkan ciri terpilih di mana mereka perlu berkunjung ke bank gen dan menggunakan sistem berasingan iaitu *Rice Genebank Information System* (RGBIS) yang mempunyai fungsi untuk menapis data. Aplikasi RGBIS ini tidak boleh diakses secara atas talian dan hanya terdapat di Pusat Genebank MARDI. Namun, mempunyai fungsi penapisan data berasaskan nilai ciri terpilih. Selain itu, pengguna cuma boleh mengakses aplikasi ini setelah dipasang dalam komputer. Artikel ini memperihalkan pembangunan satu papan pemuka yang boleh memberi maklumat berkaitan dengan germplasma padi beasaskan Ketinggian hasil, perosak dan penyakit, kualiti dan keutamaan. Pendekatan yang telah diambil adalah mengumpul keperluan, mereka bentuk papan pemuka, proses *Extract, Transform and Load* (ETL) dan penggunaan aplikasi penglihatan. Papan pemuka *MARDI Rice Genebank* yang dibangunkan boleh diakses secara atas talian yang mana pengguna boleh membuat pencarian serta memperoleh maklumat berkenaan aksesi yang khusus berdasarkan ciri-ciri yang dipilih.