Indonesian Journal of Electrical Engineering and Computer Science
Vol. 13, No. 3, March 2019, pp. 1130~1135
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v13.i3.pp1130-1135

Interlaboratory data fusion repository system (InDFuRS) for tocotrienols-based treatment

Norhaslinda Kamaruddin[1], Abdul Wahab[2]
[1]Advanced Analytics Engineering Center, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Malaysia
[2]Kulliyah of Information and Communication Technology, International Islamic University Malaysia, Malaysia

Article Info

Article history:

Received Aug 25, 2018
Revised Nov 6, 2018
Accepted Nov 14, 2018

Keywords:

Big data analytics
Data fusion
Repository
Structured and semi-structured data

ABSTRACT

Tocotrienols and tocopherols are part of the vitamin E family and have shown to produce lots of benefits especially in health supplement product. Both tocotrienols and tocopherols exist in an edible oil but varies in their ratio. It is also observed that percentage of tocopherols is higher than tocotrienols in most of our diet. Recent researches have found that tocotrienols seems to have more benefit to health especially for delaying neuro-degeneration and this has led researchers to investigate tocotrienols rich fraction (TRF) from palm kernel oil. To date, the tocotrienols extraction process is still work in progress. Hence, it is imperative that all information and results from the various laboratories experiments to be made available thus data analysis can be optimized for optimal tocotrinols production. Data acquisition from inter-laboratory experiments are valuable for collaborative researches. Efforts from multiple sources need to be combined to make it accessible for data integration. The sources of fused data can be employed as secondary back up once the data is migrated to a central repository. Traditionally data has been residing in silos across organization. Such scenario posed as a major problem especially when there are insufficient human and computational resources to manage such data. In addition, longitudinal data collections always suffer from mismanagement of the data where the data are not labeled properly using mismatched data formatting resulting to poor data readability. Therefore, a repository to facilitate data fusion using a systematic cloud-based system is proposed to ensure the data are accessible with maintained data uniformity and format and yet the security of the data is ensured as well as cost effective and fault tolerant. It is envisaged a better solution can be identified to minimize repetition of experiments and looking towards at advancement of extraction processes.

Corresponding Author:

Norhaslinda Kamaruddin,
Advanced Analytics Engineering Center,
Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA,
Shah Alam, Selangor, Malaysia.
Email: norhaslinda@tmsk.uitm.edu.my

1. INTRODUCTION

Tocotrienols and tocopherols can normally be extracted from edible oils, which are their major natural dietary sources. Tocotrienols and tocopherols can also be extracted from plant foods with low lipid but with with very low quantities. Other sources of edible oil with reasonable amount of tocotrienols and tocopherols are from seeds and other plant food processing by-products. Tocotrienols seems to have special neuroprotective, anti-cancer and cholesterol lowering properties that are not are not found in tocopherols [1].

In addition, tocotrienols also have other functions that helps maintaining health and treating disease which tocopherols does not exhibit, especially in preventing brain cell degeneration through regulating specific mediators of cell death. The cholesterol-lowering properties and suppresses growth of human breast cancer cells in tocotrienols are not found in tocopherenols [2].

Although palm kernal oil consists of mixture from tocotrienols and tocopherols, tocotrienol-rich fraction (TRF) can extract up to an average of 70% tocotrienols. In both tocotrienols and tocopherols, 4 isomers are observed, namely; alpha, beta, gamma, and delta [3]. Alpha-tocopherols has been the focus of research in the early days [4]. Recent researchers found that tocotrienols differ from tocopherols by having an unsaturated side chain [5] that results in significantly different biological activities. In the enhanced TRF, it is envisaged that adjusting the extraction process parameters can help to improve the ratio of tocotrienols to tocopherols content. In fact, an improvement of tocotrienols content between 10% to 20% has been reported after the extraction process parameter adjustment, thus improve the effectiveness of tocotrienols in supplement. In some cases, enhanced formulation are used to further improved the effectiveness by using medium-chain triglycerides (MCT) as the carrier instead of conventional long-chain triglycerides. MCT is a class of lipids composed of glycerides with fatty acids consist of C6 to C10 in length and are normally found in coconut and palm kernel oil. MCT has been used for the dietary treatment of malabsorption syndrome and weight control [6] as well as absorption enhancers of a numbers of different drugs in lipid-based microemulsions.

One of the liver disease that may spread to be very severe liver diseases such as liver fibrosis, cirrhosis and cancer is the non-alcoholic fatty liver disease (NAFLD) [7]-[10]. Studies have shown the potential of hepato-protective effect of tocotrienols in patients with NAFLD seems to provide positive improvement. Increased complete remission of fatty liver can be achieved with various mixed tocotrienols (of at least 200 mg twice per day) for a year [11]. It is still unclear how hepato-protective effects and its biological mechanisms of tocotrienols on NAFLD works. It is assumed to be largely contributed due to the anti-oxidative, anti-inflammatory and cholesterol-lowering properties of tocotrienols [12, 13]. The redox and inflammation systems and lipid metabolism are complex pathways in our body and elucidating the mechanism of action of the tocotrienols in protecting against liver stiffness may provide better understanding in the pathogenesis of the disease and the mechanism of action of tocotrienols in protecting the progression of NAFLD.

From the example, it is imperative that a complex system is required to manage and an analyse huge amount of data and information from the body to administer such complex supplements of tocotrienols in order for it to be effective and function the way it is supposed to be. In addition, there is a need to understand how to improve the process parameters so that effective ways of extracting tocotrienols from various food products can be achieved.

There is a dire need to have a complex system to match the differring needs of dietry supplement for individuals such that researchers can share and complement data from the other researchers and collaborators. The complexity, variability of data and the location of the data is scattered among many organizational applications and systems makes it challenging to even access these data. Inter-laboratory collaboration often has difficulty in communicating between collaborators due to lack of personnel and computational resources required for managing a proper database of the experimental data. Therefore, data is often stored in external hard drives by graduate students working on their thesis and often the data is lost when these students graduated.

Typically, there is no standard data storage convention that is accepted by the researchers when the data are being acquired and stored. To complicate matters, even format of the data is different from one experiment to another. For instance, the date parameters can be stored using either DD-MM-yyyy or MM-DD-yyyy. Thus, data with the same information need to be stored in different computer memory location. As such system will not be able to recognize the date format since the MM could be more than 12 or duplication of data. Thus it is important that all data must be properly managed and maintained.

Traditionally data are collected using a single database that are stored physically in an organization. While this situation seems ideal because the data can be updated and easily accesible, such situation may pose a problem if the database is faulty, corrupted or experiencing the failure of computer system. A distributed database may be employed to overcome such challenge. However, the data must be regularly refreshed and synchronized to keep it up-to-date when it is in used. Hence, a repository that is secure, robust and fault tolerant is a must to ensure the up-to-date data is accessible for all collaborators and researchers.

Thus, we proposed an Inter-laboratory Data Fusion Repository System (InDFuRS) having state-of-the-art cloud-based data repository for data fusion from various laboratory data acquisition accross the country. Different data types of data yielded from various laboratories can be managed amd normalized using standard data storage conversion. The system proposed should be affordable, computationally acceptable and practical with ease of maintainance.

2. LITERATURE REVIEW

Big data analytics and computing is an emerging data science paradigm of multi-dimensional information mining for scientific discovery and business analytics [14]. The data collected/produced from scientific explorations often require tools to facilitate efficient data management, analysis, validation, visualization and dissemination, while preserving the intrinsic value of the data [15]. Scientists and researchers produce huge amounts of data per-day via experiments, however, extracting useful knowledge for decision-making purposes from these massive, large-scale data repositories is almost impossible for actual database management system-inspired analysis tools [16]. Therefore, an approach to handle these massive data is needed.

The advancement of technology in the areas of networking and cloud computing offer end users with seamless mechanisms for creating, storing, accessing and managing their massive databases on remote (data) servers. It is also known as Database as a Service (DaaS) [17]. Due to the naïve features of big data, DaaS is the most appropriate computational data framework to implement big data repositories [18]. Figure 1 illustrates a simplified enterprise cloud architecture for a big data and analytics environment. The architecture has three network zones: public network, provider cloud, and enterprise network [19]. Notice from the diagram of Figure 1, it also allows user from the public to access some data which is available for public access.

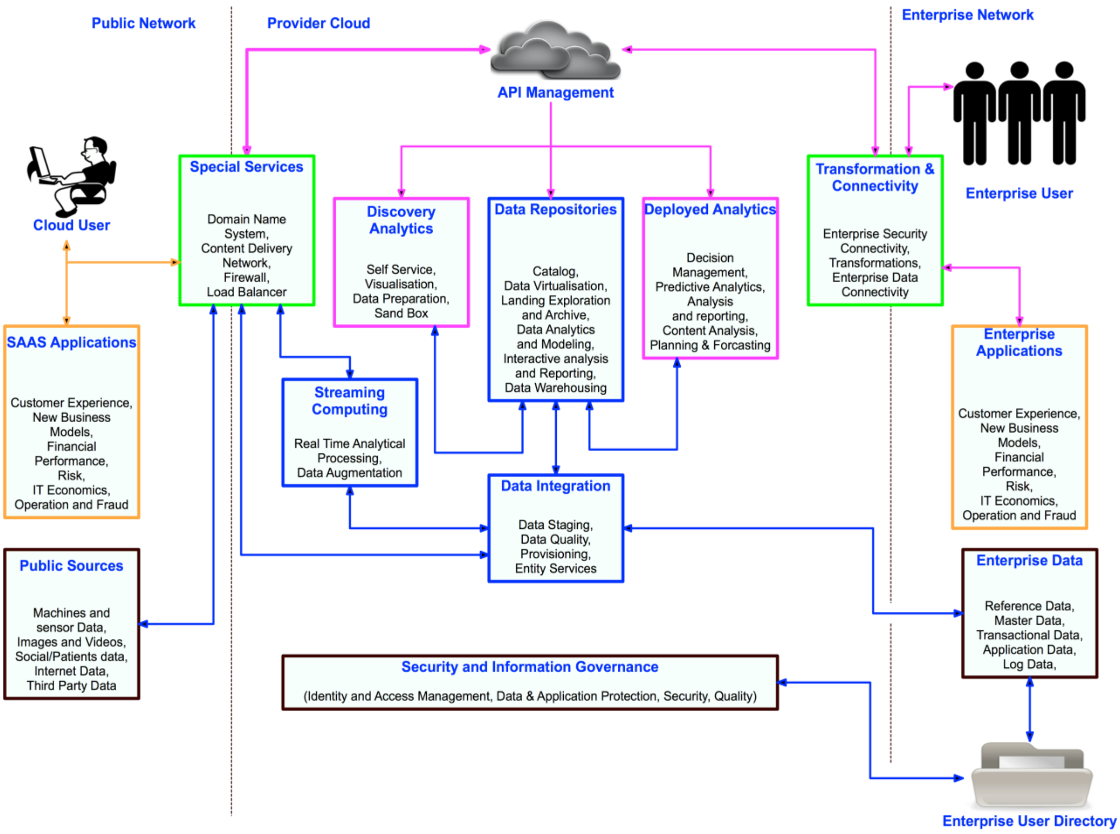

Figure 1. Simplified enterprise cloud architecture

Desktop as a Service is a desktop virtualization service that is hosted on the cloud, so users can access their virtual desktops and applications wherever they go, using whichever device they need. Because a virtual desktop is stored on a remote server, it is separated from the physical device that is used to access it. With Desktop as a Service, data gets saved automatically from the virtual desktop because it is synced with the Cloud. Customers generally manage their applications and desktop images, while the service provider handles all the back-end infrastructure and maintenance. This is particularly important so that researchers and collaborators can access data anywhere they want using any devices of their choice. In the laboratory environment data can be collected and stored into the cloud, while in the plant manufacturing the tocotrienols from the RTF can know the kind of tocotrienols needed for the patient supplement.

Hence, the implementation of the repository using cloud-based system can minimize the issues such as data is scattered among many organization applications and systems, data not properly managed and maintained and data not up-to-date for analysis. This is crucial in ensuring the latest information is available on the cloud so that all parties will have the latest information and are able to perform task accordingly. In addition, security of the data is also guarenteed since private cloud will ensure correct user will get the correct information and data.

3. RESEARCH METHODOLOGY

3.1. Data Acquisition and Analysis

This phase is carried out to identify and understand the goals and the needs of users. Both qualitative and quantitative research may need to be carried out to have a better understanding of the various requirements and needs for storage and analysing the data and the different types of data to be stored and analyse. This phase involves research and analysis of existing data including survey and interview with the collaborators and researchers. This phase can definitely help in defining the product vision, a shared understanding of the end products and services.

A table will be developed listing and comparing different target groups, their needs and various features including the data types to be stored and value used. In addition, vision board can take the form of post-its stuck on a wall but as we are based in different facilities, we will have to use a digital board such as Trello in this phase to analyze various forms of needs and different users. Common checklist will be generated to ensure all parties involed has the same the checklist and will be used to identify needs of all parties in terms of data storage and analytics.

3.2. Design

Concrete understanding of the users’ requirements have to be determined and agreed with all parties prior to sketching the wireframes. Iterating quickly from the users’ feedback, to reach a design that works can be carried out by sketching the wireframes. Once consensus on the basic design has been agreed, the work on high-fidelity versions of all the different content types will be carried out. This stage will undergo a few iterations as well, it will be incorporating other feedbacks to ensure there are no blind spots. Since this a quick turn around project (QTAP) software availability becomes critical and important rather then design from ground zero. The best choice will be to use open source available software that works with public cloud although private cloud will be ideal but costly.

At this stage it is aso important to consider data security and integrity as this will defined then type of available software to be incorporated with respect to cost and availability. Perhaps a more stringent security system with logging access through cloud is inevitable and can help bring down the cost and yet effective enough for data security and integrity.

3.3. Development

In this work, the approach of data reservoir repository by [20] will be adopted. It is imperative that whenever any researchers updated, or any new data is available the entire community need to be aware. This approach will make sure that all parties will be made aware of availaility of new or updated data through the use of catalog and advertise module.

The activities identified in Figure 2 are described as follows:

Advertise: Whenever there is a new source of data to add to the data reservoir or any updates of current available data, it is advertised in the data reservoir’s catalog. This is to ensure that all parties are made aware of availability of new data.

Catalog: The catalog described the data in the data reservoir indicating precise arrangement of how data is managed and governed. In this case all parties involved can then locate and manage the data they need. Thus having data catalog help organized classified data into various ways making it easy for all parties to find what they require.

Provision: To ensure all changes made to the original source of data are synchronized with the copies in the data reservoir, provision is incorporated into the data reservoir. Thus, flow of data into the data reservoir can be properly regulated.

Discover: The discovery will ensure that location of data can be made known through the data catalog. The whole idea of cataloging is smilar to that of the library cataloging.

Explore: The exploration of data is then carried out by verifying that the data values are correct and the data type also matched.

Access: Once explored and verified data can now be accessed directly or copied into a sandbox for use by an analysis tool.

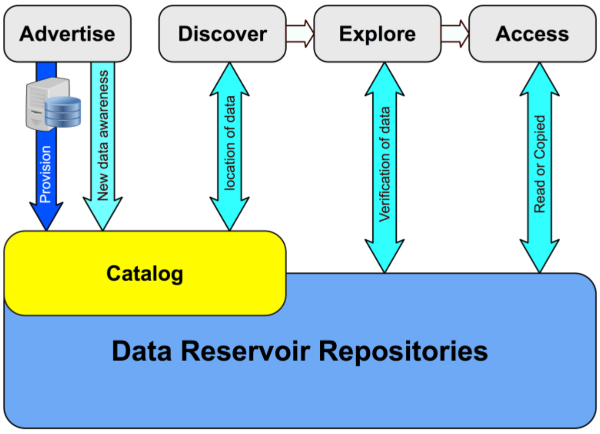

Figure 2. Data reservoir repository approach

3.4. Testing

Technological advancement and the fast-pace of churning out new software has made it almost impossible to test a software product rigorously to ensure quality of the software produce. Digitisation of the industries and the use of Internet of Things (IoT), has made it even worst that new software and product are produced almost instantly. Numerous testing automation software are also available but would require parameter tuning to optimize its usage. Nonetheless functionality and user acceptance testing must be conducted to measure usability and how well the researchers utilized the proposed system. Thus a concise and precise definition of the requirement must be well established to ensure the user acceptance test meet what ever have been set up earlier as the system requirements and this is critical to the success of the product and services.

4. CONCLUSION AND SIGNIFICANT OF THE RESERACH

The complexity of administering any supplement effectively has always been a big question. The huiman body is a complex plant and cannot be easily generalized as each of us react differently to dosage of drugs. In this case we proposed the use of big data analytics to collect, acquire, store and analyze data in the cloud system for ease on maintainance and managing the data complexity. A predefined and agreed data structure should provide an easier environment for analyzing data with standards open source applications. The use of Desktop as a service and Data as a service can allow the system to be affordable and available on time for the project. If data can be shared and made available to all parties concern easily then patients matching of various supplement to optimize intervention can be achieved with great possibilities, thus improving the supplement of drug with priori knowledge of the drug intervention and reaction.

ACKNOWLEDGEMENT

The authors would like to thank Universiti Teknologi MARA (UiTM), International Islamic University Malaysia (IIUM) and Ministry of Higher Education Malaysia (MOHE) for providing financial support through the MITRA grant (600-IRMI/PERDANA 5/3/MITRA (007/2018)-3) to conduct the work published in this paper.

REFERENCES

[1] T. R. Selvaraju, *et al.*, "The neuroprotective effects of tocotrienol rich fraction and alpha tocopherol against glutamate injury in astrocytes," *Bosnian Journal of Basic Medical Sciences*, vol. 14, pp. 195–204, 2014.

[2] K. Nesaretnam, *et al.*, "Tocotrienols and breast cancer: The evidence to date," *Genes and Nutrition*, vol/issue: 7(6), pp. 3–9, 2012.

[3] H. Ahsan, *et al.*, "Pharmacological potential of tocotrienols: a review," *Nutrition & Metabolism*, vol. 11, pp. 52, 2014.

[4] S. Fairus, *et al.*, "Alpha-tocotrienol is the most abundant tocotrienol isomer circulated in plasma and lipoproteins after postprandial tocotrienol-rich vitamin E supplementation," *Nutrition Journal*, vol. 11, pp. 5, 2012.

[5] V. K. Singh, *et al.*, "Vitamin E: Tocopherols and tocotrienols as potential radiation countermeasures," *Journal of Radiation Research*, vol. 54, pp. 973–988, 2013.

[6] N. D. Shah and B. N. Limketkai, "The Use of Medium-Chain Triglycerides in Gastrointestinal disorders," *Nutrition Issues in Gastroenterology*, pp. 20–28, 2017.
[7] L. A. Adams, *et al.*, "The Natural History of Nonalcoholic Fatty Liver Disease: A Population-Based Cohort Study," *Gastroenterology*, vol/issue: 129(1), pp. 113-121, 2005.
[8] G. Marchesini, *et al.*, "Nonalcoholic fatty liver disease and the metabolic syndrome," *Current Opinion in Lipidology*, vol/issue: 16(4), pp. 421-427, 2005.
[9] B. A. Neuschwander-Tetri, "Nonalcoholic Steatohepatitis and the Metabolic Syndrome," *The American Journal of the Medical Sciences*, vol/issue: 330(6), pp. 326-335, 2005.
[10] G. Kanuri and I. Bergheim, "In Vitro and in Vivo Models of Non-Alcoholic Fatty Liver Disease (NAFLD)," *International Journal of Molecular Sciences*, vol/issue: 14(6), pp. 11963, 2013.
[11] E. Magosso, *et al.*, "Tocotrienols for normalisation of hepatic echogenic response in nonalcoholic fatty liver: a randomised placebo-controlled clinical trial," *Nutrition Journal*, vol/issue: 12(1), pp. 1-8, 2013.
[12] P. Lettéron, *et al.*, "Acute and chronic hepatic steatosis lead to in vivo lipid peroxidation in mice," *Journal of Hepatology*, vol/issue: 24(2), pp. 200-208, 1996.
[13] E. Albano, *et al.*, "Review article: role of oxidative stress in the progression of non-alcoholic steatosis," *Alimentary Pharmacology & Therapeutics*, vol. 22, pp. 71-73, 2005.
[14] R. Kune, *et al.*, "The Anatomy of Big Data Computing," *Software: Practice and Experience*, vol/issue: 46(1), pp. 79-105, 2016.
[15] P. B. Jensen, *et al.*, "Mining Electronic Health Records: Towards Better Research Applications and Clinical Care," *Nature Reviews Genetics*, vol/issue: 13(6), pp. 395-405, 2012.
[16] A. Cuzzocrea, *et al.*, "Analytics over Large-scale Multidimensional Data: The Big Data Revolution!" *Proceedings of the ACM 14th International Workshop on Data Warehousing and OLAP (DOLAP '11). ACM, New York, NY, USA*, pp. 101-104, 2011.
[17] Cloud Database, "Definition from WhatIs.com," http://searchcloudapplications.techtarget.com/definition/cloud-database, Retrieved 9 Oct 2017.
[18] D. Agrawal, *et al.*, "Big Data and Cloud Computing: Current State and Future Opportunities," *Proceedings of the 14th International Conference on Extending Database Technology*, pp. 530-533, 2011.
[19] M. Gupta, *et al.*, "How IBM leads in Building Big Data Analytics Solutions in the Cloud: Implementing the CSCC Customer Cloud Architecture for Big Data and Analytics," 1996. https://www.ibm.com/developerworks/cloud/library/cl-ibm-leads-building-big-data-analytics-solutions-cloud-trs/index.html, Retrieved 9 Oct 2017.
[20] M. Chessell, *et al.*, "Governing and Managing Big Data for Analytics and Decision Makers," *IBM Redguides for Business Leaders*, 2014.